Published in final edited form as:

Nat Struct Mol Biol. 2016 December ; 23(12): 1117–1123. doi:10.1038/nsmb.3317.

The ribosome-engaged landscape of alternative splicing

Robert J. Weatheritt^{1,2,4}, Tim Sterne-Weiler¹, and Benjamin J. Blencowe^{1,3,4}

¹Donnelly Centre, University of Toronto, Toronto, Ontario M5S 3E1, Canada

²MRC Laboratory of Molecular Biology, Francis Crick Avenue, Cambridge CB2 0QH, UK

³Department of Molecular Genetics, University of Toronto, Toronto, Ontario M5S 1A8, Canada

Abstract

High-throughput RNA-sequencing (RNA-Seq) has revealed an enormous complexity of alternative splicing (AS) across diverse cell and tissue types. However, it is currently not known to what extent repertoires of splice variant transcripts are translated into protein products. Here, we survey AS events engaged by the ribosome. Remarkably, at least 75% of human exon skipping events detected in medium to high abundance transcripts using RNA-Seq data are also detected in ribosome profiling data. Furthermore, relatively small subsets of functionally related splice variants are engaged by ribosomes at levels that do not reflect their absolute abundance, indicating an important role for AS in modulating translational output. We show that this mode of regulation is associated with control of the mammalian cell cycle. Our results thus suggest that a major fraction of splice variants is translated, and that specific cellular functions including cell cycle control are subject to AS-dependent modulation of translation output.

Species with vastly differing biological complexity have comparable numbers of protein coding genes. In contrast, transcriptome profiling studies have revealed that the complexity of AS parallels both organ and species complexity^{1,2}. More than 95% of human multi-exon genes are subject to AS events that affect exonic sequence^{3,4}, and more than two-thirds produce transcripts that contain one or more retained intron⁵. Moreover, tens of thousands of AS events are differentially regulated between cell/tissue types or conditions^{6–9}. Regulated AS events typically are the most highly conserved and enriched for frame preservation, implying that they often have important biological functions^{7,9}. Moreover, large programs of co-regulated AS events are significantly enriched in genes that operate in specific biological processes and pathways, and, in many cases, alternative exons in these networks have been shown to contribute critical biological functions^{2,7,10–13}.

Despite these observations, the extent to which AS events detected at the transcriptomic level are translated and functional as protein products is not known^{10,14–16}. A major

⁴**Corresponding authors** Robert J. Weatheritt and Benjamin J. Blencowe, Address for Correspondence: Donnelly Centre, University of Toronto, 160 College Street, Toronto, ON, M5T 2R5, Canada, b.blencowe@utoronto.ca; robert.weatheritt@utoronto.ca.

Author Contributions

R.J.W. conceived the study, designed and performed analyses, with input from B.J.B. T.S.W. contributed to methods for analysing ribosome profiling data. R.J.W. and B.J.B. wrote the manuscript with input from T.S.W.

Competing financial interests

The authors declare no competing financial interests

challenge in addressing this question is that current mass spectrometry (MS) datasets do not afford sufficient coverage and sensitivity to comprehensively detect peptides corresponding to splice variant sequences¹⁷, and are especially limited in terms of detection of cell and tissue-specific splice variants¹⁴. However, recent studies have attempted direct comparisons of AS events detected at the level of RNA and MS data and have led to the surprising proposal that very small numbers of splice variants are translated^{14,18}.

Ribosome profiling affords a sensitive method for the detection of translated RNAs. Moreover, extensive analyses have shown that, in general, the degree of engagement of an mRNA by the ribosome correlates well with its level of translation^{19–21}. Accordingly, we employed ribosome-profiling data to assess the extent to which AS events detected in RNA-Seq data are engaged by the ribosome and thus likely to be translated. We further used these data to discover genes whose translational output is regulated by AS in a cell cycle dependent manner.

Results

To investigate the extent to which alternatively spliced sequences are engaged by ribosomes and potentially translated, we developed an analysis pipeline enabling the examination of matched RNA-Seq and ribosome-profiling datasets across several cell types, conditions and species (Online Methods). Initially, we compared the frequency of cassette exon AS using RNA-Seq and ribosome-profiling data generated from the same samples of synchronized human cells²². The percentage of transcripts with an exon spliced in ('Percent Spliced In' [PSI]) was estimated for all exons for which there was sufficient read coverage to derive an estimate. Importantly, the ribosome profiling datasets employed in our analysis were of high sequence coverage (i.e. >130 million reads; Supplementary Table 1) with minimal ribosomal RNA contamination. As such, we were able to reliably quantify AS events over a wide range of transcript abundance. In the comparisons described below, cassette exon AS frequency was calculated by comparing the frequency of exon skipping events per detected exon at different levels of read coverage, as determined by the number of reads mapping to exon-exon junctions.

Remarkably, for transcripts expressed at medium to high abundance levels, the frequency of cassette exon AS detected in ribosome-profiling data is comparable to that detected in corresponding RNA-Seq datasets, after normalization using detection of proximal constitutive exon-exon junction sequences (Fig. 1a). Similar results were obtained when comparing matched RNA-Seq and ribosome profiling data from other human cell lines (HEK293 and BJ cells), and from mouse stem cells (see Supplementary Fig. 1 and Supplementary Table 1). These results reveal that the absolute frequencies of exon skipping in transcripts engaged by the ribosome and in whole-cell RNA-Seq data are comparable over a broad range of transcript abundance levels.

In contrast, at relatively low abundance levels, there is a reduced frequency of exon skipping detected in ribosome-engaged transcripts relative to transcripts detected in RNA-Seq data. To investigate whether this observation is related to low levels of mRNA expression and is not simply a consequence of reduced detection sensitivity, we compared the relative ratios of

constitutive versus alternative exons detected in transcripts expressed at different levels of expression. Consistent with the results in Figure 1a, at low levels of abundance, there is a significantly reduced proportion of alternative versus constitutive exons detected in ribosome-engaged transcripts, relative to transcripts detected in RNA-Seq data ($p < 9.10 \times 10^{-5}$, Wilcoxon-Test), whereas for genes expressed at higher levels, the relative proportions of alternative versus constitutive exons are not significantly different between these datasets (Fig. 1b). To verify these findings, we subsampled ribosome-protected reads from highly expressed, high read-coverage transcripts to assess the recall precision of cassette alternative exons when read coverage from subsampling was comparable to that for low abundance transcripts (Online Methods). Although subsampling progressively smaller numbers of reads resulted in a small (i.e. 10-20%) decrease in detection of alternative events, this decrease is significantly less than that observed when comparing the ratios of junction sequences between low and medium-high abundance transcripts (Supplementary Fig. 2d, $p < 8.20 \times 10^{-8}$, Chi-square test). These results thus confirm that alternative exons are significantly underrepresented in low abundance transcripts.

We next assessed the proportions of total, distinct cassette AS events detected in RNA-Seq data, and overlapping coding sequence, that are engaged by ribosomes at different levels of transcript abundance (Fig. 1c). In order to control for technical differences between RNA-Seq and ribosome profiling data, we restricted these comparisons to regions of transcripts in which at least both constitutive exons flanking an AS event detected in RNA-Seq data are also confidently detected as engaged by the ribosome. Consistent with the results described above, for medium to high abundance transcripts we observe that approximately 75-85% of cassette AS events detected in RNA-Seq data are also engaged by ribosomes, whereas a significantly smaller proportion is engaged by ribosomes at lower levels of mRNA expression (Fig. 1c and Supplementary Fig. 2; $p < 1.94 \times 10^{-6}$, Fisher's exact Test). These results thus suggest that for medium to highly-expressed transcripts the majority of cassette AS events overlapping coding sequences are translated.

We next investigated differences in the sequence properties of exon skipping events detected in transcripts that are ribosome-engaged versus transcripts detected in RNA-Seq data. Importantly, as expected, there is a significant enrichment for cassette exons predicted to contribute to open reading frames in ribosome-engaged versus RNA-Seq transcripts (Fig. 1d; $p < 2.28 \times 10^{-22}$, Fisher's Exact Test; Online Methods), and a corresponding reduction in the inclusion of cassette exons predicted to disrupt open reading frames (Fig. 1e; $p < 3.17 \times 10^{-18}$, Fisher's Exact Test; Online Methods). This trend however is consistent with a significant enrichment of frame-preserving events in medium to high abundance transcripts in RNA-Seq data (Supplementary Fig. 2; $p < 2.72 \times 10^{-05}$, Fisher's Exact Test).

We next compared the frequency of detection of intron retention (IR) events in RNA-Seq and ribosome profiling data. IR was measured as the percentage of total transcripts from a gene with intron retention (PIR). Previous studies have shown that the majority of IR events lead to nuclear retention or, if the retained intron transcripts are exported, to cytoplasmic NMD^{5,13,23}. A smaller fraction of IR events are, however, frame-preserving and may be translated^{5,24}. Consistent with these findings, a comparison of ribosomal profiling and RNA-Seq data from the whole cell fraction reveals that, overall, there is a significant

reduction in the frequency of detected IR events engaged by the ribosome (Fig. 2a; $p < 5.69 \times 10^{-5}$; Wilcoxon Test). Surprisingly, however, there is also a comparable frequency of IR events detected in ribosome profiling and cytosolic RNA-Seq data (Fig. 2a) suggesting that transcripts containing IR events that are exported to the cytoplasm are often engaged by the ribosome.

Given these findings, we investigated the fraction of total IR events detected in RNA-Seq data that are also ribosome-engaged, using a similar approach as in Figure 1c for cassette exons (Online Methods). Consistent with the results in Figure 2a and our previous observations⁵, there is a significantly higher fraction of IR events in transcripts with relatively low mRNA expression (Fig. 2b; $p < 7.88 \times 10^{-7}$, Fisher's exact Test). However, we also observe an increase in ribosome-engaged IR events associated with relatively highly expressed transcripts (Fig. 2b and Supplementary Fig. 3; $p < 1.91 \times 10^{-3}$, Fisher's exact Test). Consistent with this observation, IR events engaged by the ribosome are significantly enriched in 5'-UTR sequences (Supplementary Fig. 3; $p < 0.015$, Fisher's exact Test) and a significant fraction of these sequences include predicted upstream opening reading frames (uORFs), as compared to total IR events detected in RNA-Seq (Supplementary Fig. 3; $p < 3.58 \times 10^{-5}$, Fisher's exact Test). Moreover, further consistent with the ribosome-engagement of these retained intron sequences, they are significantly enriched in cytoplasmic versus nuclear RNA fractions (Fig. 2d; $p < 2.0 \times 10^{-3}$, Fisher's exact Test). Further supporting these observations, the 5'-UTR ribosome-engaged IR events are markedly shorter than the average length of retained introns (i.e. 289 nt vs. 1583 nt; $p < 5.96 \times 10^{-16}$, Wilcoxon-Test). In fact, their length is similar to the mean length of exons (Fig. 2e). In line with this observation, cross-species sequence comparisons reveal that many of these IR events appear to have evolved through intronization (Fig. 2f), the evolutionary process by which an intron has arisen from within an ancestral exon²⁵.

Genes producing ribosome-engaged transcripts that contain 5'-UTR IR events are significantly enriched in essential functions and include many housekeeping roles^{26,27} (Fig. 3a). Among the functional categories enriched in this set of genes are cell cycle control, translation and DNA repair (Fig. 3b; examples shown in Fig. 3c). Given the link with cell cycle control, we next investigated whether IR, cassette, and alternative 5' and 3' AS events, potentially modulate the degree to which transcripts are ribosome engaged in a cell-cycle dependent manner. To this end, we asked whether genes with cell cycle stage-dependent periodic regulation of AS are significantly more often differentially engaged by ribosomes, as compared to genes that do not contain periodically regulated AS events (Supplementary Fig. 4; Online Methods).

Remarkably, transcripts with all types of periodically-regulated AS events have a significantly higher than expected (i.e. when comparing to total transcript abundance) degree of ribosome engagement between cell cycle stages, as compared to transcripts that lack these events (Fig. 4a; $p < 1.04 \times 10^{-15}$, Wilcoxon-Test). Analysis of quantitative MS data²⁸ from individual stages of the cell cycle confirms that there is a significantly higher degree of detected changes in the relative abundance of peptides corresponding to transcripts that contain periodic AS events versus transcripts that lack these events (Fig. 4b; $p < 1.49 \times 10^{-04}$, Wilcoxon-Test). Moreover, we observe a significantly higher degree of differential

ribosome engagement of individual transcripts that have cell cycle-periodic AS events, as compared to transcripts that contain cell type-specific AS events (Fig. 4c; $p < 6.49 \times 10^{-03}$, Wilcoxon-Test), or transcripts that are differentially regulated during the cell cycle at the transcriptional level (Fig. 4c; $p < 0.02$, Wilcoxon-Test). Collectively, these results suggest that periodically-regulated AS events likely significantly impact the translation of a specific subset of transcripts during the cell cycle.

To investigate possible mechanisms and functions associated with cell-cycle specific AS-dependent modulation of ribosome engagement, we asked whether there are specific positional or structural features that preferentially overlap cell cycle-regulated AS events versus cell type-specific events (see Supplementary Table 1). Consistent with the results in Figure 3, when comparing between these ribosome profiling datasets, cell cycle-regulated AS events are significantly more often located within 5'-UTRs and/or overlapping the translational start site (Supplementary Fig. 4; $p < 1.80 \times 10^{-04}$, Fisher's exact Test). They are also predicted to significantly more often impact putative uORFs (Supplementary Fig. 4; $p < 4.31 \times 10^{-04}$, Fisher's exact test). Moreover, transcripts whose 5'-UTRs contain terminal oligopyrimidine tract (TOP) motifs, which previously have been shown to influence translational efficiency during the G2/M phase^{29,30}, are significantly enriched in ribosome engaged, periodically regulated AS events, including those that overlap uORFs (Supplementary Fig. 4; $p < 3.80 \times 10^{-03}$, Fisher's exact test). Furthermore, there is also a higher degree of differential secondary structure overlapping cell cycle-dependent AS events versus cell type-specific AS events (Fig. 4d; $p < 7.24 \times 10^{-7}$ compared to cell type-specific, Wilcoxon-Test). Collectively, these findings suggest that an important role for AS is to regulate cell-cycle genes by altering transcript sequence and structural features that in turn control translation initiation^{31,32}.

Finally, genes producing transcripts that contain periodic AS events that increase ribosome engagement encode numerous important cell cycle regulators, including multiple components of the anaphase-promoting complex (APC) (Fig. 4e and Supplementary Fig. 4). An example is the cell division cycle protein 20 (CDC20), an APC protein that functions as a key cell cycle checkpoint control factor during mitosis³³. A 5'-UTR IR event in CDC20 peaks during the M/G1-phase, coincident with increased ribosomal loading of CDC20 transcripts. Another example is Aurora Kinase A (AURKA), a key cell cycle regulator (Supplementary Fig. 4), which contains a periodically regulated 5' UTR cassette exon event, whose inclusion similarly peaks during M phase. Taken together, these data provide evidence for an extensive role for periodic AS in the translational control of important cell cycle regulators.

Discussion

The results of this study support the conclusion that a major fraction of cassette exon AS events in medium to high abundance transcripts are engaged by ribosomes and therefore likely translated. In contrast, there is a significantly reduced frequency of ribosome engagement of cassette AS events in low abundance transcripts. We provide evidence that this reduced frequency of ribosome engagement is unlikely a consequence of limited detection sensitivity of splice junctions in low abundance transcripts, but rather because

these transcripts are more often subject to IR events which prevent or reduce ribosome engagement⁵. This observation is consistent with recent results demonstrating that IR events are detected more frequently in low abundance transcripts⁵, and that a substantial fraction of these events result in nuclear retention^{11,34}. However, it is also possible that other forms of inefficient or incomplete RNA processing may also limit ribosome engagement in low abundance transcripts.

The conclusions of our study contrast sharply with those of several recent reports comparing AS events detected in cDNA databases, RNA-Seq and MS data^{14,18}. In these studies, very small numbers of peptides were detected that map uniquely to splice variant sequences represented in transcript sequencing data, leading the authors to propose that only a minor fraction of splice variants are translated. However, a more parsimonious explanation is that current MS datasets lack sufficient coverage and sensitivity to detect the vast majority of expressed AS events. The technical limitations of current MS datasets are further compounded by biases that likely limit detection of a significant fraction of AS events at the protein level. For example, up to one third of cassette exons comprise cell/tissue-differential cassette AS events, of which significant fractions are conserved and frame preserving⁹. However, because of the sparseness of high confidence spectral calls within individual MS datasets, currently available MS data afford the detection of only a very minor number of cell and tissue specific splice isoforms^{14,35} and therefore are further under-represented for AS events that impact coding sequence. The results of our study are however consistent with recent findings indicating that large numbers of splice variant transcripts detected in RNA-Seq data associate with polysomes. For example, ~2,000 and ~60,000 splice variant transcripts were detected in polysome fractions using 'FRAC-Seq' and 'TriP-Seq' methods, respectively^{30,31}. However, these studies did not assess the proportion of AS events detected in RNA-Seq data that are also detected in polysome fractions, nor which AS events are directly bound by ribosomes.

In the present study we also detect a class of AS events overlapping 5'-UTR sequences that correlate with significant changes in ribosome occupancy. These observations extend previous reports of AS events that modulate translation output^{36–39}, and are further consistent with the aforementioned polysome analyses in which it was observed that specific 5'-UTR sequences correlate with differential polysome association and can alter translation output over a 100-fold range in reporter assays^{30,31}. Thus, while some studies have supported an overall concordance between mRNA abundance and protein abundance⁴⁰, specific features of transcripts, including the 5'-UTR intron retention events defined in the present study, can significantly impact this relationship. We further show that AS-dependent regulation of ribosome engagement occurs frequently during the cell cycle and likely contributes to cell cycle progression by controlling the timing of translation of key cell cycle regulators, such as CDC20 and AURKA. Future studies will undoubtedly uncover additional programs of AS and associated mechanisms that function in the control of translation.

Methods

Sequence Alignments and Normalization

Prior to alignment, linker and poly(A) sequences were removed from 3'-ends of reads. Reads were first aligned using Bowtie49 to human/mouse rRNA sequences and matches discarded (-v 3). All remaining reads were aligned to human (hg19) genome or mouse (mm9) genome using an adapted version of VAST-TOOLS allowing only unique reads and no more than two mismatches. Ribosomal profiling densities and mRNA expression levels were calculated as described in50. Briefly, an initial correction was undertaken for each position in each transcript using 50-basepair windows (30nt for ribosomal profiling data), which were mapped back against the whole transcriptome using Bowtie with the same parameters. If this sequence did not map to a unique position, it was discarded and discounted from the transcript effective length (length-50/30). The "effective length" was then used to divide the raw read counts per million mapped reads for each gene to obtain corrected-RPKM values (cRPKM)50. As well as a gene-level cRPKM value for expression, a transcript-based expression analysis using Kallisto51 (<https://pachterlab.github.io/kallisto/>) was used. For constructing indices, the fragment length was adjusted for ribosome profiling data and Ensembl42 (Hg19) indices produced. All transcripts with a median of under a 5 transcripts per million (TPM) across samples were not included in analysis. To ensure gene expression values could be compared across samples, a Between Sample Normalization (BSN) approach was undertaken using the DESeq normalization algorithm using the median-of-ratios method.

Detection of Alternative Splicing

To comprehensively detect and quantify all types of AS events involving alternative exons (AltEx), alternative 5' and 3' splice site selections (Alt5/Alt3), and intron retention (IR) for use in ribosome profiling analysis, we adapted the VAST-TOOLS multi-module analysis pipeline (described in detail in5,7, as well as in Irimia et al. (in preparation)). Briefly, reads were initially mapped to genome assemblies using Bowtie, using -m 1 -c 2 parameters with reads that mapped to the genome discarded for AS quantifications. Unique EEJ (exon-exon junction) libraries were generated to derive measurements of exon inclusion levels using the metric "Percent Spliced In" (PSI). This utilized all hypothetically possible EEJ combinations from annotated and *de novo* splice sites, including both cassette, mutually exclusive and microexon events5,7. In addition, a "Percent Splice-site usage" (PSU) based metric was used to detect and quantify tandem splice site acceptor or donor sites (Alt3 (Alternative 3') and Alt5 end events). An intron retention (IR) analysis pipeline was used to detect and quantify IR as the percentage of transcripts with an intron retained (PIR). This pipeline employs a comprehensive set of reference sequences comprising for each IR event: two exon-intron junctions (EIJs), intron mid-point sequences, and EEJs formed by intron removal31. Each IR event requires multiple reads mapping to both the EIJ and the intron mid-point sequence, as described previously31. For all modules and AS types, quantifications were done based on read counts corrected for the number mappable positions in each EEJ or EIJ following the formula :

$$\text{Corrected}_{\text{EEJ}_{\text{count}}} = \text{EEJ}_{\text{count}} \frac{J_{\text{count}} * \text{Maximum}_{\text{mappability}}}{\text{EEJ}_{\text{mappability}}}$$

where $\text{EEJ}_{\text{count}}$ is the number of read groups mapped to the EEJ, $\text{Maximum}_{\text{mappability}}$ the maximum number of mapping positions that any EEJ can have for reads of length 50 nt (i.e. 35 positions) for RNA-Seq and 30nt (i.e. 15 positions) for ribosomal profiling data while ensuring a minimal overlap of 8nt either side of EEJ, and $\text{EEJ}_{\text{mappability}}$ the number of positions that can be mapped uniquely to the EEJ using specific bowtie parameters ($-m 1 -v 2$), and thus $\text{EEJ}_{\text{mappability}} \times \text{Maximum}_{\text{mappability}}$ (see 5,7 for details).

Differential Splicing Analysis

Differential identification of percentage splicing in (PSI) for AS events or percentage intron retained (PIR) for IR events were calculated using the Diff module (available at <https://github.com/vastgroup/vast-tools>). Briefly, diff is a differential splicing analysis module that uses Bayesian inference, where the prior distribution is a uniform Beta($\alpha=1$, $\beta=1$), and the likelihood function follows a Binomial distribution where the number of inclusion reads $K \sim \text{Binomial}(\Psi, N)$, where Ψ represents PSI or PIR, and N the total number of junction reads per-event. Following Bayes theorem we are left with a conjugate posterior distribution over $\Psi \sim \text{Beta}(K + \alpha, (N-K) + \beta)$. When replicates are available joint posterior distributions for a sample are estimated from sampling empirical posterior distributions of the replicates and fitting a new posterior Beta using maximum-likelihood (MLE) estimation in R. Statistical significance between two posterior distributions $X \sim \text{Beta}$, and $Y \sim \text{Beta}$, is calculated as $P(X-Y > 0)$ and estimated using empirical distributions sampled from X and Y .

Frequency analysis

For initial analysis of AS frequency, only cassette events supported by multiple reads at each EEJ and a PSI value of between 10 and 90 were considered. Only genes with at least one AS events were evaluated. To calculate normalized read depths for the comparison between RNA-Seq and ribosomal profiling, all reads (corrected for mappability) that uniquely mapped to an exon-exon junction (EEJs) were considered. This read count was then divided by the sum of the exons from respective transcript. The fraction of AS exons per known exon was then calculated. A similar approach was undertaken for IR, with events only considered when there were multiple reads supporting each exon-intron junction, as well as multiple internal intron reads. Only genes with at least one IR event were evaluated. In contrast to cassette events all IR events with a percentage intron retention (PIR) values above 10 were considered.

The constitutive versus alternative analysis used the same analysis pipeline but only includes genes when at least 3 (constitutive or alternative) exons were confidently detected. The same criteria for both cassette events and intron retention events were used and a constitutive exon was defined as having a consistent PSI value above 95. Additional analysis (Supplementary Fig. 2d) subsampled mapped reads from transcripts from highly expressed genes (i.e. >100 cRPKM) at intervals of 10%, using the same criteria for alternative/constitutive definitions as above.

Ribosomal engagement calculation

Ribosome engagement is defined as the log₂ ratio between the corrected ribosomal fraction densities and the corrected mRNA expression levels. To decrease the likelihood of low expression transcripts/genes spuriously creating large fold changes, only genes with a read count of at least 50 in both the RNA-seq and ribosomal profiling datasets were included, as recommended in⁵². To compare global levels of translational efficiency, a cumulative frequency distribution is used and non-parametric tests used to compare distributions of efficiency. The initial analysis (Fig. 4a) was performed using VAST-TOOLS with additional transcript-level expression comparisons done using Kallisto⁵¹ (Fig. 4c). Genes affected by intron retention events were removed to control for potential influence of NMD but results from analysis did not change (data not shown).

Identification of regulatory datasets

The tissue-specific data was extracted for each cell type using criteria from a previous transcriptomics analysis we undertook⁷ and only included if events were also identified as alternative in ribosomal profiling data within same cell type. For periodic AS data, differential splicing analysis between cell cycle stages was followed by Fourier transform analysis⁵³ to identify repeating periodic changes in the cell cycle for particular events over 2 cycles of the cell cycle. Transcriptionally-regulated genes were download from Cyclebase⁵⁴. Protein annotation was parsed from UniProt annotation

Annotated terminal oligopyrimidine tract (TOP) repeat motifs were extracted from a previous study by Yamashita et al. (2008)⁵³ and uORFs were mapped using the Global Translation Initiation sequencing (GTI-seq) method⁵⁵. The occurrences of TOP motifs in the 5'-UTRs of genes containing periodic- and cell type-differential AS were compared.

Transcript features

For the identification of translation start sites only canonical AUG sites were considered based on read build up after subtraction of background, as described in⁵⁶. Positions of exons within transcripts (i.e. CDS or UTR etc) were mapped using Ensembl GTF files of protein-coding genes with an occurrence within a CDS region ceding to the occurrence of an exon within a UTR region. An event within the CDS was predicted to create a functional proteins if the exon is either divisible by 3 and does not introduce an in-frame stop codon or if it partially overlaps the body the 5'-UTR. Potentially disrupting exons are events that will introduce a frameshift (and therefore NMD) upon exclusion or inclusion only. ORF-preserving events are therefore events that change in PSI between RNA-Seq and ribosome profiling datasets, and promote the stable (non-frame-shifted) transcript in the ribosome profiling data whereas ORF-disrupting events show changes in PSI between RNA-Seq and ribosome profiling datasets, and promote the frame-shifting variant in ribosome profiling data. Annotated ATG sites were from Ensembl v71⁴². For IR analysis, 5'-UTR IR events were compared to "All IR events", which comprise human IR events assembled using diverse cell and tissue polyA+ RNA-Seq data in Braunschweig et al. (2014)⁵. Differences in RNA secondary structure were determined using the *in vivo* data from Rouskin et al. (2014)⁴⁶ with events only included if both the AS exon and at least the first adjacent constitutive exon were mapped with secondary structure features from Rouskin et al.⁴⁶. The

protein abundance difference values were normalized to account for changes in mRNA expression by dividing the label free quantification (LFQ) intensity value produced by label free mass spectrometry by the fragments per Kilobase of transcript per million (FPKM) values identified for mRNA expression within the same samples. The change was then calculated between cell cycle stages (G1, S or G2). Data extracted from Ly et al. (2014)²⁸. Intronic information was extracted from Braunschweig et al. (2014)⁵

Statistics

Wilcoxon-rank sum test was used for comparing distributions and Fisher's exact test for comparing enrichments. All statistical tests are two-sided. Error bars show standard error of the mean with assumption of Poisson distribution of counts. Functional analysis was performed using GProfiler with a FDR multiple testing correction, a background of expressed genes, and only including genes sets with a maximum number of 1000. Enrichment Map⁴⁷ was used to create functional networks in Cytoscape.

Code Availability

Software for analysis of RNA-Seq data is available at Github (<https://github.com/vastgroup>). All other scripts were written in Python and are available upon request from RJW. Analyses were performed in R statistical package

Data Availability

Previously published data sets reanalyzed here are listed with accession numbers in Supplementary Table 1. Source Data for Figures 1-4 are available with the paper online. Other data supporting the findings of this study are available from the corresponding authors upon request.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

Acknowledgements

We thank U Braunschweig, J Ellis, T Gonatopoulos-Pournatzis, S Gueroussov, K Ha, M Irimia, and J Roth for helpful comments on the manuscript and technical assistance. We thank M Irimia for providing annotations for ORF disrupting/preserving AS events. This work is supported by grants from the Canadian Institutes of Health Research (CIHR) to B.J.B, by CIHR postdoctoral and Marie Curie IOF fellowships to R.J.W and by CIHR and Charles H. Best postdoctoral fellowships to T.S.W. B.J.B. holds the Banbury Chair in Medical Research at the University of Toronto.

References

1. Barbosa-Morais NL, et al. The evolutionary landscape of alternative splicing in vertebrate species. *Science*. 2012; 338:1587–1593. [PubMed: 23258890]
2. Merkin J, Russell C, Chen P, Burge CB. Evolutionary dynamics of gene and isoform regulation in Mammalian tissues. *Science*. 2012; 338:1593–1599. [PubMed: 23258891]
3. Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet*. 2008; 40:1413–1415. [PubMed: 18978789]

4. Wang ET, et al. Alternative isoform regulation in human tissue transcriptomes. *Nature*. 2008; 456:470–476. [PubMed: 18978772]
5. Braunschweig U, et al. Widespread intron retention in mammals functionally tunes transcriptomes. *Genome Res*. 2014; 24:1774–1786. [PubMed: 25258385]
6. Dominguez D, et al. An extensive program of periodic alternative splicing linked to cell cycle progression. *Elife*. 2016; 5
7. Irimia M, et al. A highly conserved program of neuronal microexons is misregulated in autistic brains. *Cell*. 2014; 159:1511–1523. [PubMed: 25525873]
8. Li Q, Lee JA, Black DL. Neuronal regulation of alternative pre-mRNA splicing. *Nat Rev Neurosci*. 2007; 8:819–831. [PubMed: 17895907]
9. Xing Y, Lee CJ. Protein modularity of alternatively spliced exons is associated with tissue-specific regulation of alternative splicing. *PLoS Genet*. 2005; 1:e34. [PubMed: 16170410]
10. Buljan M, et al. Tissue-specific splicing of disordered segments that embed binding motifs rewires protein interaction networks. *Mol Cell*. 2012; 46:871–883. [PubMed: 22749400]
11. Jangi M, Sharp PA. Building robust transcriptomes with master splicing factors. *Cell*. 2014; 159:487–498. [PubMed: 25417102]
12. Weatheritt RJ, Davey NE, Gibson TJ. Linear motifs confer functional diversity onto splice variants. *Nucleic Acids Res*. 2012; 40:7123–7131. [PubMed: 22638587]
13. Wong JJ, et al. Orchestrated intron retention regulates normal granulocyte differentiation. *Cell*. 2013; 154:583–595. [PubMed: 23911323]
14. Abascal F, et al. Alternatively Spliced Homologous Exons Have Ancient Origins and Are Highly Expressed at the Protein Level. *PLoS Comput Biol*. 2015; 11:e1004325. [PubMed: 26061177]
15. Abascal F, Tress ML, Valencia A. The evolutionary fate of alternatively spliced homologous exons after gene duplication. *Genome Biol Evol*. 2015; 7:1392–1403. [PubMed: 25931610]
16. Weatheritt RJ, Gibson TJ. Linear motifs: lost in (pre)translation. *Trends Biochem Sci*. 2012; 37:333–341. [PubMed: 22705166]
17. Bensimon A, Heck AJ, Aebersold R. Mass spectrometry-based proteomics and network biology. *Annu Rev Biochem*. 2012; 81:379–405. [PubMed: 22439968]
18. Ezkurdia I, et al. Most highly expressed protein-coding genes have a single dominant isoform. *J Proteome Res*. 2015; 14:1880–1887. [PubMed: 25732134]
19. Battle A, et al. Genomic variation. Impact of regulatory variation from RNA to protein. *Science*. 2015; 347:664–667. [PubMed: 25657249]
20. Ingolia NT, Ghaemmighami S, Newman JR, Weissman JS. Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. *Science*. 2009; 324:218–223. [PubMed: 19213877]
21. Ingolia NT. Ribosome Footprint Profiling of Translation throughout the Genome. *Cell*. 2016; 165:22–33. [PubMed: 27015305]
22. Tanenbaum ME, Stern-Ginossar N, Weissman JS, Vale RD. Regulation of mRNA translation during mitosis. *Elife*. 2015; 4
23. Yap K, Lim ZQ, Khandelia P, Friedman B, Makeyev EV. Coordinated regulation of neuronal mRNA steady-state levels through developmentally controlled intron retention. *Genes Dev*. 2012; 26:1209–1223. [PubMed: 22661231]
24. Marquez Y, Hopfler M, Ayatollahi Z, Barta A, Kalyna M. Unmasking alternative splicing inside protein-coding exons defines exitrons and their role in proteome plasticity. *Genome Res*. 2015; 25:995–1007. [PubMed: 25934563]
25. Irimia M, et al. Origin of introns by ‘intronization’ of exonic sequences. *Trends Genet*. 2008; 24:378–381. [PubMed: 18597887]
26. Eisenberg E, Levanon EY. Human housekeeping genes, revisited. *Trends Genet*. 2013; 29:569–574. [PubMed: 23810203]
27. Wang T, et al. Identification and characterization of essential genes in the human genome. *Science*. 2015; 350:1096–1101. [PubMed: 26472758]
28. Ly T, et al. A proteomic chronology of gene expression through the cell cycle in human myeloid leukemia cells. *Elife*. 2014; 3:e01630. [PubMed: 24596151]

29. Park JE, Yi H, Kim Y, Chang H, Kim VN. Regulation of Poly(A) Tail and Translation during the Somatic Cell Cycle. *Mol Cell*. 2016; 62:462–471. [PubMed: 27153541]
30. Sterne-Weiler T, et al. Frac-seq reveals isoform-specific recruitment to polyribosomes. *Genome Res*. 2013; 23:1615–1623. [PubMed: 23783272]
31. Floor SN, Doudna JA. Tunable protein synthesis by transcript isoforms in human cells. *Elife*. 2016; 5
32. Kutchko KM, et al. Multiple conformations are a conserved and regulatory feature of the RB1 5' UTR. *RNA*. 2015; 21:1274–1285. [PubMed: 25999316]
33. Chang L, Barford D. Insights into the anaphase-promoting complex: a molecular machine that regulates mitosis. *Curr Opin Struct Biol*. 2014; 29:1–9. [PubMed: 25174288]
34. Boutz PL, Bhutkar A, Sharp PA. Detained introns are a novel, widespread class of post-transcriptionally spliced introns. *Genes Dev*. 2015; 29:63–80. [PubMed: 25561496]
35. Olsen JV, Mann M. Status of large-scale analysis of post-translational modifications by mass spectrometry. *Mol Cell Proteomics*. 2013; 12:3444–3452. [PubMed: 24187339]
36. Juntawong P, Girke T, Bazin J, Bailey-Serres J. Translational dynamics revealed by genome-wide profiling of ribosome footprints in Arabidopsis. *Proc Natl Acad Sci U S A*. 2014; 111:E203–E212. [PubMed: 24367078]
37. Newton DC, et al. Translational regulation of human neuronal nitric-oxide synthase by an alternatively spliced 5'-untranslated region leader exon. *J Biol Chem*. 2003; 278:636–644. [PubMed: 12403769]
38. Remy E, et al. Intron retention in the 5'UTR of the novel ZIF2 transporter enhances translation to promote zinc tolerance in arabidopsis. *PLoS Genet*. 2014; 10:e1004375. [PubMed: 24832541]
39. Zhang Y, et al. Translational control of the rat angiotensin type 1a receptor by alternative splicing. *Gene*. 2004; 341:93–100. [PubMed: 15474292]
40. Vogel C, Marcotte EM. Insights into the regulation of protein abundance from proteomic and transcriptomic analyses. *Nat Rev Genet*. 2012; 13:227–232. [PubMed: 22411467]
41. Anders S, Huber W. Differential expression analysis for sequence count data. *Genome Biol*. 2010; 11:R106. [PubMed: 20979621]
42. Cunningham F, et al. Ensembl 2015. *Nucleic Acids Res*. 2015; 43:D662–D669. [PubMed: 25352552]
43. Andreev DE, et al. Translation of 5' leaders is pervasive in genes resistant to eIF2 repression. *Elife*. 2015; 4:e03971. [PubMed: 25621764]
44. Guo H, Ingolia NT, Weissman JS, Bartel DP. Mammalian microRNAs predominantly act to decrease target mRNA levels. *Nature*. 2010; 466:835–840. [PubMed: 20703300]
45. Rooijers K, Loayza-Puch F, Nijtmans LG, Agami R. Ribosome profiling reveals features of normal and disease-associated mitochondrial translation. *Nat Commun*. 2013; 4:2886. [PubMed: 24301020]
46. Rouskin S, Zubradt M, Washietl S, Kellis M, Weissman JS. Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature*. 2014; 505:701–705. [PubMed: 24336214]
47. Merico D, Isserlin R, Stueker O, Emili A, Bader GD. Enrichment map: a network-based method for gene-set enrichment visualization and interpretation. *PLoS One*. 2010; 5:e13984. [PubMed: 21085593]
48. Stumpf CR, Moreno MV, Olshen AB, Taylor BS, Ruggero D. The translational landscape of the mammalian cell cycle. *Mol Cell*. 2013; 52:574–582. [PubMed: 24120665]
49. Langmead B, Trapnell C, Pop M, Salzberg SL. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol*. 2009; 10:R25. [PubMed: 19261174]
50. Labbe RM, et al. A comparative transcriptomic analysis reveals conserved features of stem cell pluripotency in planarians and mammals. *Stem Cells*. 2012; 30:1734–1745. [PubMed: 22696458]
51. Bray N, et al. Near-optimal RNA-seq quantification. *arXiv.org*. 2015 arXiv: 1505.02710.
52. Ingolia NT, Lareau LF, Weissman JS. Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell*. 2011; 147:789–802. [PubMed: 22056041]

53. Whitfield ML, et al. Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Mol Biol Cell*. 2002; 13:1977–2000. [PubMed: 12058064]
54. Santos A, Wernersson R, Jensen LJ. Cyclebase 3.0: a multi-organism database on cell-cycle regulation and phenotypes. *Nucleic Acids Res*. 2015; 43:D1140–D1144. [PubMed: 25378319]
55. Lee S, et al. Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. *Proc Natl Acad Sci U S A*. 2012; 109:E2424–E2432. [PubMed: 22927429]
56. Ingolia NT, et al. Ribosome profiling reveals pervasive translation outside of annotated protein-coding genes. *Cell Rep*. 2014; 8:1365–1379. [PubMed: 25159147]

Fig 1. Most cassette alternative splicing events in medium to high abundance transcripts are engaged by the ribosome

(a) Box plots showing AS frequency (scored as the fraction of annotated exons in canonical transcripts that show skipping) for genes with different levels of RNA-Seq or ribosome-profiling read coverage in human cells. Simple, complex and microexon (i.e. 3-27 nt) cassette events were analyzed and only genes with detected AS events in RNA-Seq data were included. EEJ, exon-exon junction (b) Stacked bar plot comparing percentages of constitutive versus alternatively spliced exons for all genes with at least three exons detected. Between-sample normalization (BSN) of corrected reads per kilobase of transcript per million (cRPKM)1 was performed using DESeq41. (c) Bar plot comparing fractions of total alternative spliced events identified in RNA-Seq data that are also identified as alternatively spliced in ribosome profiling (RP) data, at different expression levels. Events were only analyzed if both constitutive exons surrounding an alternative exon were detected in ribosome profiling data (n=5462). Error bar calculated from the standard error of the mean (d) Bar plots showing fractions of coding AS events detected in matched ribosome

profiling (n=2,431) and RNA-Seq (n=8,797) datasets comprising exons divisible by 3 and not encoding an in-frame premature stop codons, or not divisible by 3 but partially overlapping UTR sequence. Error bar calculated from the standard error of the mean and p-value calculated using Fisher's exact test (e) Bar plot showing fractions of coding AS events not divisible by 3 within CDS that display changes in PSI (percentage spliced in) between ribosomal profiling and matched RNA-Seq data (n=1226). 'ORF-Preserving' indicates that the PSI change promotes the inclusion of a frame-preserving exon; 'ORF-Disrupting' indicates that the inclusion of a frame-shifting exon with the potential to elicit NMD. Error bar calculated from the standard error of the mean and p-value calculated using Fisher's exact test. Data for all panels obtained from 22. See Figure 2 and Online Methods for description of boxplots and statistical tests used. Source data are available online.

Fig 2. Detection of intron retention events in ribosomal profiling data

(a) Box plots showing intron retention frequency (fraction of annotated introns that show retention in canonical transcripts) for genes with different RNA-Seq/ribosome profiling read coverage. EIJ = Exon-intron Junction (b) Bar plot showing fraction of total intron retention events identified in RNA-Seq data that are also detected as retained in ribosome profiling (RP) data (n=847). Error bar calculated using the standard error of the mean and p-value calculated using Fisher's exact test (c) Bar plot as in (b) showing intron retention events detected in 5' UTR sequences and other regions ('REST') of transcripts with >100 cRPKM coverage (n=123). Error bar calculated from the standard error of the mean and p-value calculated using Fisher's exact test (d) Bar plot showing the percentage change in detection of IR events in different transcript locations using ribosome profiling data (n=847) and RNA-Seq data from fractionated cells (Cytosol n=2980; Nuclear n=3810), as compared to whole cell RNA-Seq data. Error bar calculated from the standard error of the mean and p-value calculated using Fisher's exact test. Transcript locations are mapped based on Ensembl GTF annotations⁴². (e) Box plots comparing average lengths of 5'-UTR retained introns

identified in ribosome profiling data and for total retained introns. Error bar calculated from the standard error of the mean and p-value calculated using Wilcoxon-Test. (5'-UTR introns: n=123; All introns: n = 9,760) (f) Bar plot showing fractions of 5' UTR retained introns identified in ribosome profiling data and total retained introns with evidence of intronization. Error bar calculated from the standard error of the mean and p-value calculated using Fisher's exact test (5'-UTR introns: n=123; All introns: n = 9,760). Sequencing data obtained from 22,43–45. See Supplementary Table 1 for details. For boxplots, the median value for each group of proteins is shown with a horizontal black line. Boxes enclose values between the first and third quartile. Interquartile range is calculated by subtracting the first quartile from the third quartile. All values outside this range are considered to be outliers and were removed from the graphs to improve visualization. The smallest and highest values that are not outliers are connected with the dashed line. The notches correspond to ~95% confidence interval for the median. Source data are available online.

Fig 3. Ribosome engaged intron retention events are enriched within the 5' UTRs of essential genes

(a) Bar plots comparing the frequency of detection of 5'UTR retained introns identified in ribosome profiling data and total retained introns in genes scored as essential in cell viability assays²⁷ and in housekeeping genes²⁶. Error bar calculated from the standard error of the mean and p-value calculated using Fisher's exact test (5'-UTR introns: n=123; All introns: n = 9,760) (b) Enrichment map of functional categories identified for genes with ribosome-engaged intron retention events that overlap 5'-UTR sequences. Each node represents a Gene Ontology (GO) category with overlapping gene-set clusters linked together by edges and organized into clouds of similar function. (c) Heatmaps showing the percentage of transcripts with a percent intron retained (PIR) values for introns located within 5'-UTRs, coding sequence (CDS) and 3'-UTRs. Colour scales indicate PIR values and the colour shading reflects the gene function category as shown in panel (b). Sequencing data obtained from^{22,43–45}, see Supplementary Table 1 for details. Source data are available online.

Fig 4. Cell cycle-regulated AS events control ribosome engagement

(a) Cumulative frequency plot comparing changes in levels of ribosome engagement of transcripts at different cell cycle stages for genes with cell-cycle-dependent (periodic) splicing changes ($n=411$) compared to genes with general AS events ($n=1068$). General AS events are events found across ribosome-profiling datasets where no regulation has been assigned. Box plots below quantify these changes. (b) Box plots comparing peptide abundance changes between cell cycle stages for four different categories of genes: 'All Genes' are genes identified as expressed during the cell cycle ($n=6195$); 'Periodic GE' are

genes that are differentially expressed between cell cycle stages (n=379); 'Cell Type AS' are genes with cell/tissue-specific AS events (n=315)(see Online Methods). (c) Box plots showing expression changes of transcripts between different cell cycle stages in transcripts per million (TPM) (d) Box plots displaying absolute difference in fraction of nucleotides detected as structured⁴⁶ between an alternative exon and their surrounding constitutive exons. (e) Enrichment map⁴⁷ of functional categories identified for genes with cell-cycle-dependent AS events and changes in ribosomal engagement between cell cycle stages (see also Supplementary Fig. 4). Each node represents a Gene Ontology (GO) category with overlapping gene-set clusters linked together by edges and organized into clouds of similar function. See Figure 2 and Online Methods for description of boxplots and statistical tests used. Sequencing data obtained from 6,22,46,48; see Supplementary Table 1 for details.