Method

Ultrashort and progressive 4sU-tagging reveals key characteristics of RNA processing at nucleotide resolution

Lukas Windhager,[1,7] Thomas Bonfert,[1,7] Kaspar Burger,[2] Zsolt Ruzsics,[3] Stefan Krebs,[4] Stefanie Kaufmann,[1] Georg Malterer,[3] Anne L'Hernault,[5] Markus Schilhabel,[6] Stefan Schreiber,[6] Philip Rosenstiel,[6] Ralf Zimmer,[1] Dirk Eick,[2] Caroline C. Friedel,[1,8] and Lars Dölken[5,8]

[1]Institute for Informatics, Ludwig-Maximilians-Universität München, Munich 80333, Germany; [2]Department of Molecular Epigenetics, Helmholtz-Zentrum München, Center of Integrated Protein Science (CIPSM), Munich 81377, Germany; [3]Max von Pettenkofer-Institute, Ludwig-Maximilians-Universität München, Munich 80336, Germany; [4]Gene Center, Ludwig-Maximilians-Universität München, Munich 81377, Germany; [5]Department of Medicine, University of Cambridge, Addenbrooke's Hospital, Cambridge CB2 0QQ, United Kingdom; [6]Institute of Clinical Molecular Biology, Christian-Albrechts-Universität Kiel, Kiel 24105, Germany

RNA synthesis and decay rates determine the steady-state levels of cellular RNAs. Metabolic tagging of newly transcribed RNA by 4-thiouridine (4sU) can reveal the relative contributions of RNA synthesis and decay rates. The kinetics of RNA processing, however, had so far remained unresolved. Here, we show that ultrashort 4sU-tagging not only provides snapshot pictures of eukaryotic gene expression but, when combined with progressive 4sU-tagging and RNA-seq, reveals global RNA processing kinetics at nucleotide resolution. Using this method, we identified classes of rapidly and slowly spliced/degraded introns. Interestingly, each class of splicing kinetics was characterized by a distinct association with intron length, gene length, and splice site strength. For a large group of introns, we also observed long lasting retention in the primary transcript, but efficient secondary splicing or degradation at later time points. Finally, we show that processing of most, but not all small nucleolar (sno)RNA-containing introns is remarkably inefficient with the majority of introns being spliced and degraded rather than processed into mature snoRNAs. In summary, our study yields unparalleled insights into the kinetics of RNA processing and provides the tools to study molecular mechanisms of RNA processing and their contribution to the regulation of gene expression.

[Supplemental material is available for this article.]

RNA levels in a cell are determined by the rates of transcription, RNA processing, and RNA decay. Regulation may occur at all three levels providing substantial flexibility for adaption to alterations in environmental conditions (Jing et al. 2005; Kim et al. 2009; Nilsen and Graveley 2010). Most studies focus on regulation at the transcriptional level but changes in RNA degradation rates may also significantly alter gene expression of coding and noncoding RNAs (Shalem et al. 2008; Cazalla et al. 2010; Miller et al. 2011). So far, little is known about the contribution of alterations in RNA processing to gene expression. Furthermore, despite the knowledge on the occurrence of multiple isoforms of transcripts, the dynamic mechanisms guiding tissue- and context-specific regulation of RNA processing (e.g., alternative splicing events) remain unknown. Research has been severely hampered by the lack of proper tools to study these processes with sufficient resolution. Next-generation sequencing of total cellular RNA (RNA-seq) allows studying the outcome of RNA processing at whole-transcriptome level at a given time point (Pan et al. 2008; Wang et al. 2008). This has recently resulted in the discovery of numerous new alternative isoforms of mammalian transcripts indicating that most multi-exon genes are alternatively spliced (Nilsen and Graveley 2010). The kinetics of RNA splicing and processing and thus the underlying regulatory mechanisms, however, can hardly be resolved with these techniques.

Metabolic labeling of newly transcribed RNA using 4-thiouridine (4sU-tagging), a naturally occurring uridine derivative, provides direct access to newly synthesized transcripts with minimal interference to cell growth and gene expression (Melvin et al. 1978; Cleary et al. 2005; Kenzelmann et al. 2007; Dölken et al. 2008; Friedel et al. 2009; Weintz et al. 2010). Following isolation of total cellular RNA and thiol-specific biotinylation, this can be quantitatively separated into tagged (newly transcribed) and untagged (preexisting) RNA using streptavidin-coated magnetic beads. This allows bias-free analysis of RNA synthesis and decay at high resolution. We and others have demonstrated that this approach provides access to the dynamics of RNA production and degradation in eukaryotic cells. Furthermore, it is directly compatible with microarray analysis (Dölken et al. 2008; Friedel and Dölken 2009; Friedel et al. 2009) and RNA-seq (Rabani et al. 2011; Schwanhäusser et al. 2011). However, only relatively long durations of 4sU-tagging were used together with RNA-seq so far. Here, we show that ultrashort 4sU-tagging with as little as 5-min labeling time can be combined with RNA sequencing to provide

high-quality sequencing data. The combination of ultrashort and progressive 4sU-tagging from 5- to 60-min labeling time then allows unparalleled insights into the kinetics of RNA processing, in particular RNA splicing and processing of noncoding RNAs.

Results

Ultrashort 4sU-tagging is compatible with RNA-seq in human B-cells

Newly transcribed RNA obtained by 4sU-tagging contains substantially greater amounts of large, unprocessed transcripts than regularly found in total cellular RNA. This is readily visualized by electrophoretic analysis (Dölken et al. 2008). When shortening the duration of 4sU-tagging the average age of nascent transcripts in newly transcribed RNA decreases. We thus hypothesized that RNA-seq combined with progressive reduction of the duration of 4sU-tagging could be employed to study the kinetics of RNA processing. For this purpose, we performed a time course experiment of 4sU-tagging in DG75 human B-cells consisting of five samples with 60, 20, 15, 10, and 5 min of 4sU-tagging. At the end of 4sU exposure, cells were harvested using TRIzol, total cellular RNA was prepared, and newly transcribed RNA was purified. The relative abundance of newly transcribed, tagged RNA in total cellular RNA decreased from 3.5% of total RNA after 1-h 4sU-tagging to ~0.8% after 5 min (Fig. 1A). Newly transcribed RNA from all five labeling conditions was subjected to RNA-seq analysis using sequencing by ligation (SOLiD II, Applied Biosystems). In the following, we will refer to these samples as '5-min 4sU-RNA' to '60-min 4sU-RNA.' In addition, total and untagged RNA following 60 min of 4sU-tagging were sequenced. As newly transcribed RNA may not yet be poly-adenylated, no poly(A) selection was performed. For each sample, between 2.8 and 4.4 million reads could be uniquely mapped by aligning them first to the human tran-
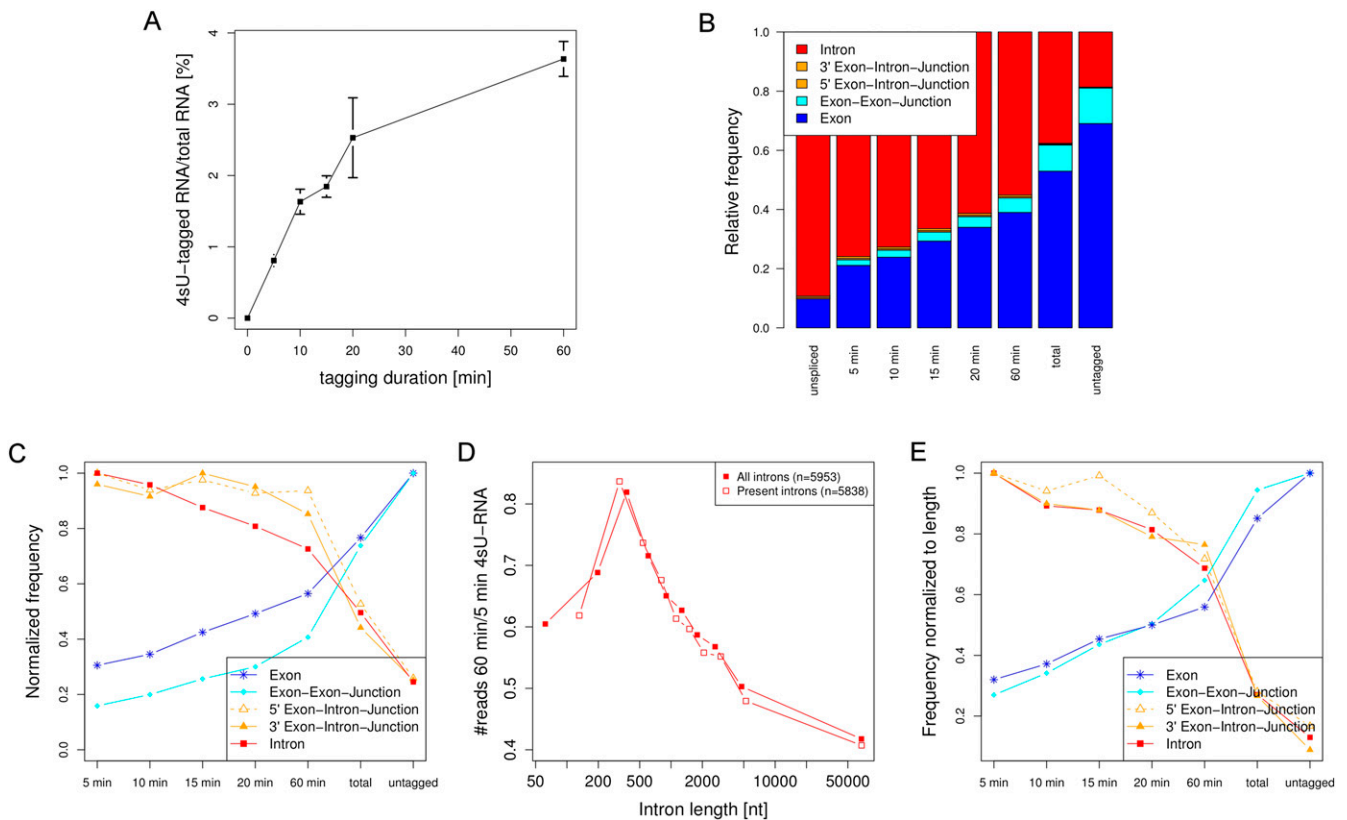

Figure 1. Ultrashort and progressive 4sU-seq reveals the kinetics of RNA splicing. (A) Contribution of newly transcribed RNA to total RNA levels following different durations of 4sU exposure. The relative contributions of purified newly transcribed RNA (4sU-tagged RNA) to input RNA (total RNA) measured spectrophotometrically (OD$_{260}$) are shown (combined data of three biological replicates). Standard error is indicated by error bars. (B) Distribution of the number of reads mapped to exons, exon–exon junctions, exon–intron junctions, and intron regions for 5- to 60-min 4sU-RNA, total and untagged RNA. Samples are ordered according to increasing age of RNA in these samples from 5- to 60-min 4sU-tagging to total RNA and finally untagged RNA. RNA in the untagged RNA samples is at least 60 min old. This visualizes the maturation of transcripts over time. The expected distribution of reads for completely unspliced RNA is shown in the *left*-most column (see Supplemental Methods). (C) Normalized read frequencies were calculated by first dividing read numbers by the total number of reads on protein-coding genes in the corresponding sample. Subsequently, frequencies for a specific read type were divided by the maximum frequency observed for the corresponding read type in any sample. Despite the overall smaller number of exon–exon junction and exon–intron junction reads, their normalized read frequencies are similar to normalized frequencies of exon and intron reads, respectively. (D) Introns were binned into 10 equally sized groups according to intron length. For each group, average ratios of intron read counts in 60-min compared with 5-min 4sU-RNA were calculated. Results are shown both for all introns of the 525 most highly expressed genes (excluding only those with 0 reads in 5-min 4sU-RNA) and introns present in 5-min 4sU-RNA (RPKM ≥ 0.5). Ratios were highest for introns in a range of 300–400 nt and decreased both for longer and shorter introns. This was observed independently of gene length (see Supplemental Fig. 1E). (E) After normalizing the number of intronic reads to intron length and exonic reads to exon length, the delay between the decrease in intronic reads and exon–intron junction reads, on the one hand, and the increase in exonic reads and exon–exon junction reads, on the other hand, disappeared. For this analysis, only the 525 most highly expressed genes were considered. After normalization to intron and exon length, respectively, the same normalization was applied as in C.

scriptome and the remaining unaligned reads to the human genome. All samples passed standard quality controls.

We first assessed the contribution of intronic and exonic sequences in the seven RNA samples. As predicted, the number of reads mapping to intronic sequences increased with reduced duration of 4sU-tagging from 18.9% in untagged RNA to 75.9% in 5-min 4sU-RNA (Fig. 1B). As excised introns are generally believed to be rapidly degraded (Lamond et al. 1988; Nam et al. 1997; Clement et al. 1999) this indicates the presence of large amounts of unspliced pre-mRNAs in the newly transcribed RNA samples. Our data thus provide solid evidence that as little as 5 min of 4sU-tagging yields high-quality newly transcribed RNA fully compatible with next-generation sequencing. If none of the transcripts in 5-min 4sU-RNA had undergone any splicing events, the intronic reads would have been predicted to contribute ~89% instead of 75.9% of all reads (Fig. 1B). Thus, a substantial fraction of cellular transcripts in 5-min 4sU-RNA has already undergone splicing events with >65% (conservative estimate) of all introns already decayed (see Supplemental Methods on how this estimate was obtained).

Similar to the changes in the contribution of intronic reads over time, we observed a strong correlation between the number of reads crossing exon–intron or exon–exon junctions and the duration of 4sU-tagging (Fig. 1C). Exon–intron junction reads result from unspliced or partially spliced transcripts. Accordingly, their contribution considerably decreased with the duration of 4sU-tagging (from 1.1% in 5-min 4sU-RNA to 0.29% in untagged RNA). Conversely, the frequency of exon–exon junction reads increased from 1.9% in 5-min 4sU-RNA to 12% in untagged RNA. In conclusion, these results show that progressive 4sU-tagging visualizes the maturation of transcripts over time.

Intron decay is correlated to intron length, gene length, and splice site strength

Interestingly, the decrease in the number of exon–intron junction reads appeared to be delayed compared with the decrease in the intronic reads themselves (Fig. 1C). As the relative contribution of long introns to the number of intronic reads is higher than to the number of exon–intron reads (Supplemental Fig. 1A), we hypothesized this to be due to a faster decay of long introns. In this case, intronic read numbers, which are disproportionately determined by long introns, would decrease faster with increasing labeling time than exon–intron junction read numbers. Indeed, no delay could be observed when read numbers were analyzed selectively for three approximately equally sized groups of introns with similar length (Supplemental Fig. 1B–D).

To quantify the rate of decay for each intron, ratios of intron reads in 60-min 4sU-RNA compared with 5-min 4sU-RNA (60/5 min ratios) were calculated and correlated with intron length. Thus, high 60/5 min ratios correspond to low decay rates and vice versa. On the individual intron level, a correlation with intron length was basically nonexistent (spearman rank correlation $r_s = -0.037$, P-value for a correlation different from zero = 0.0039). However, when introns were binned according to length and average 60/5 min ratios were calculated for each bin, a very clear trend was observed (Fig. 1D). Interestingly, the highest 60/5 min ratios, i.e., the lowest decay rates, were not observed for the shortest introns, but for introns in a range of ~300–400 nucleotides (nt). Decay rates increased both for introns longer as well as shorter than this range, with the highest decay rates observed for introns longer than ~1500 nt. Accordingly, after normalizing

intronic read counts to intron length, the lag between the loss of intronic reads and exon–intron junction reads over time disappeared (Fig. 1E).

To investigate whether these findings were influenced by any bias introduced by our experimental setup or computational analysis, we correlated 60/5 min ratios and intron length with other intron features. While there was no significant correlation between the relative position of an intron in a gene and its 60/5 min ratio or length, gene length was correlated to both characteristics ($r_s = -0.24$ for 60/5 min ratios and $r_s = 0.38$ for intron length, P-value $< 10^{-15}$ in both cases). Here, the correlation between gene length and 60/5 min ratio was very clear when binning according to gene length ($r_s = -0.94$): As average gene length increased, average 60/5 min ratios decreased. An analysis of 60/5 min ratios binned first according to gene length into three groups, and then for each group binned according to intron length, indicated that the contributions of gene and intron length were mostly independent of each other (Supplemental Fig. 1E). Irrespective of gene length, intron length and 60/5 min ratios showed the same characteristic correlation with a peak of 60/5 min ratios in the low-to-medium intron length range (300–400 nt) and decreasing ratios on either side.

Finally, we investigated whether splice site strength had any effect on splicing kinetics. Splice site strength was quantified in terms of splice site scores calculated with MaxEntScan (Yeo and Burge 2004). Similar to intron length, there was only an extremely weak correlation between splice site scores (5′ only, 3′ only, and 5′ + 3′ scores) and 60/5 min ratios ($r_s = -0.055, -0.044,$ and -0.072, respectively) or intron length ($r_s = 0.055, 0.024,$ and $0.046,$ respectively). However, when binning introns according to the splice site strength, there was a trend toward higher 60/5 min ratios (slower splicing kinetics) for very weak splice sites (Supplemental Fig. 1F). This effect was mostly independent of intron length (Supplemental Fig. 1G).

Distinct classes of introns are defined by their splicing kinetics

To further investigate differences in the kinetics of intron processing, we first focused on the most highly expressed genes as intron expression levels in newly transcribed RNA, although substantially higher than in total RNA, are much lower than the expression levels of the surrounding exons. This is due to the large fraction of introns (>65%) already spliced and decayed in 5-min 4sU-RNA. Expression levels of genes were quantified in terms of reads per kilobase of gene per million mapped reads (RPKM) after normalizing for mappability (see Methods), and the analysis was focused on genes with an RPKM ≥ 11 in all RNA samples (525 genes). Since ribosomal proteins are generally highly expressed, they, as well as other genes involved in translation and ribosome biogenesis, were enriched in this set. Other enriched functions included RNA splicing, ATPase activity, cell cycle, and regulation of apoptosis (see Supplemental Table 1 for a full list). For these genes, we distinguished between introns absent (RPKM < 0.5: 1014 introns, Supplemental Table 2) or present (5838 introns, Supplemental Table 3) in 5-min 4sU-RNA. Even after excluding 50 absent introns (~5%) that were shorter than the read length (35 nt) and, thus, could not contain any intronic reads, absent introns were significantly shorter than the present ones (Wilcoxon test, P-value $< 10^{-15}$). Furthermore, they were located closer to the 3′ end of the gene than present introns (Wilcoxon test, P-value = 0.0042) with 12% of the absent introns being the last intron of the gene compared with 7% for present introns (Fisher's exact test, P-value $< 10^{-6}$).

This suggests that at least some of these introns were part of longer transcript versions that were not transcribed in this form in the DG75 cells. For other introns, possible explanations for their absence in 5-min 4sU-RNA might be (1) very fast co-transcriptional splicing, (2) problems in sequencing, or (3) problems in mapping these parts of the pre-mRNA, e.g., due to repetitive sequences. Interestingly, in many absent introns, both neighboring exons were well expressed and precisely delimited. This indicates rapid co-transcriptional splicing and intron degradation rather than sequencing bias. In addition, there was no significant increase for the absent introns in the frequency of repetitive sequences identified by RepeatMasker (Smit et al. 1996) or the frequency of nonunique read mappings. Notably, the fraction of absent intron positions contained within repetitive sequences was actually significantly smaller than for present ones (Wilcoxon test, P-value $< 10^{-9}$). These analyses confirm that numerous transcripts in 5-min 4sU-RNA (>65%) had already been spliced and their introns had been degraded.

Remarkably, exon–intron junction reads for these absent introns were regularly observed, although only at ~60% of the level of introns present in 5-min 4sU-RNA (Wilcoxon test, P-value $< 10^{-15}$). The frequency of exon–exon junction reads was reduced to a similar degree. This indicates that the proportion of completed splicing events including the fusion of neighboring exons was not significantly higher than for present introns. A possible explanation for this observation is that major parts of the absent introns are already degraded while the exon–intron junctions are still connected to the splicing machinery in a state prior to exon fusion.

An example for such an absent intron is shown for the *RPL23A* gene (Fig. 2A). In this case, the absent intron is surrounded by a retained intron contained in an alternative transcript of *RPL23A*. According to the Ensembl annotation, this alternative transcript is not translated. It is important to note that the absence of this intron is not due to repetitive sequences in this region, which might result in nonuniquely mapped reads and their subsequent exclusion during mapping. Although two repeats were identified by RepeatMasker in this absent intron, sequence divergence to the corresponding repeat consensus was rather high (6.2% and 11.3%, respectively). Accordingly, only one additional nonuniquely mapping read was found in all samples. This confirms that the low expression of this intron is not an artifact created by the removal of nonunique mappings. Interestingly, the intron retention in *RPL23A* is still observed at considerable levels in 60-min 4sU-RNA, but hardly detectable in total and untagged RNA. This suggests that this alternative splicing variant is erroneously produced at considerable levels but quality control mechanisms of the cell later on result in its delayed but nevertheless highly efficient removal. This may either result from decay of the whole transcript or secondary splicing events removing only the retained intron.

To systematically mine for such and other distinctive kinetics of intron processing, we used cluster analysis on the introns still present in 5-min 4sU-RNA. Introns were clustered according to their relative abundance compared with the expression level of the gene (intron/gene ratio). The intron/gene ratio approximately represents the fraction of pre-mRNAs or transcripts still containing the intron. A stable clustering was obtained by repeated k-means clustering (see Methods). The resulting 20 largest clusters represented 85% of introns (Supplemental Fig. 2A). Although distinct differences between introns of the same gene were observed, they were significantly enriched in the same clusters (Fisher's exact test, P-value $< 10^{-15}$). This indicates that introns of the same genes tend to have similar splicing kinetics.

A visual inspection of the clusters identified distinctive subgroups of clusters with different absolute 5-min intron/gene ratios but similar trends across the time course, i.e., similar relative ratio changes over time. For random data sets analyzed as controls, no distinctive subgroups were observed (Supplemental Fig. 2B). To systematically identify such subgroups of splicing patterns, the 20 largest clusters were clustered again after normalizing the intron/gene ratios to 5-min 4sU-RNA. Thus, exon/intron ratios in 5-min 4sU-RNA were set to 100%. This resulted in four distinct classes of intron splicing (Fig. 2B; Supplemental Table 3). The largest class, Class 1, representing 3908 introns, was characterized by an almost linear slope in intron/gene ratios up to 20-min 4sU-RNA. Class 2 (580 introns) showed a smaller slope than Class 1, suggesting a lower splicing rate. In contrast, Class 3 (338 introns) was characterized by a more rapid decrease in intron/gene ratios between 5- and 10-min 4sU-RNA, indicating a faster splicing rate. Finally, Class 4 (109 introns) showed a very interesting pattern. Here, intron/gene ratios remained remarkably stable from 5- to 60-min 4sU-RNA, but then dropped significantly some time later to reach similar low levels in total and untagged RNA as observed in the other classes. Exemplary introns for Class 1 and 4 are shown in Figure 2C,D. Please note that the division into different classes was not always clear-cut. For instance, although cluster 3 (238 introns) is assigned to Class 2, it shows a similar trend as Class 4 introns with stable intron levels up to 20–60 min and subsequent decay. In general, Class 2 constitutes a weaker version of the delayed decay in Class 4, while Class 3 contains extreme cases of rapid intron removal and decay.

Monitoring the fate of alternative splicing products

Interestingly, Class 4 introns and to a lesser degree Class 2 introns showed a pattern very similar to the known intron retention example shown in Figure 2A. This suggests that these introns likely constitute novel alternative splicing events resulting in retained introns. To provide evidence for this hypothesis, degradation patterns for known alternative splicing products were investigated. For this purpose, exon boundaries were recalculated based on Ensembl transcripts and the redefined exons were classified either as core exons, retained introns, or alternative 3′ or 5′ exon ends (see Supplemental Material). Exon expression levels were divided by the overall gene expression level (exon/gene ratios) and normalized to 5-min 4sU-RNA (Fig. 2E). As expected, core exon/gene ratios were stable across all time points. In contrast, for retained introns the same pattern was observed as for Class 4 and some Class 2 introns. The same pattern was also seen for alternative 3′ and 5′ ends, although much less pronounced. This implies that alternative transcripts containing retained introns are present at substantial numbers (>30%) until 60-min 4sU-RNA but are subsequently degraded quite rapidly, e.g., by nonsense-mediated decay or by secondary splicing events.

Properties of splicing classes

Intron length, gene length, and splice site strength were analyzed to reveal their contribution in defining the kinetics of splicing in the four classes of introns. Interestingly, significant differences in the distribution of these features were observed (Fig. 3A). Class 1 introns were on average longer (by 36%, Wilcoxon test, FDR corrected P-value $< 10^{-12}$), contained in longer genes (by 76%, P-value $< 10^{-15}$), and showed increased splice site strength (by 4.5%, P-value $< 10^{-6}$) compared with introns of the other classes. As all of these
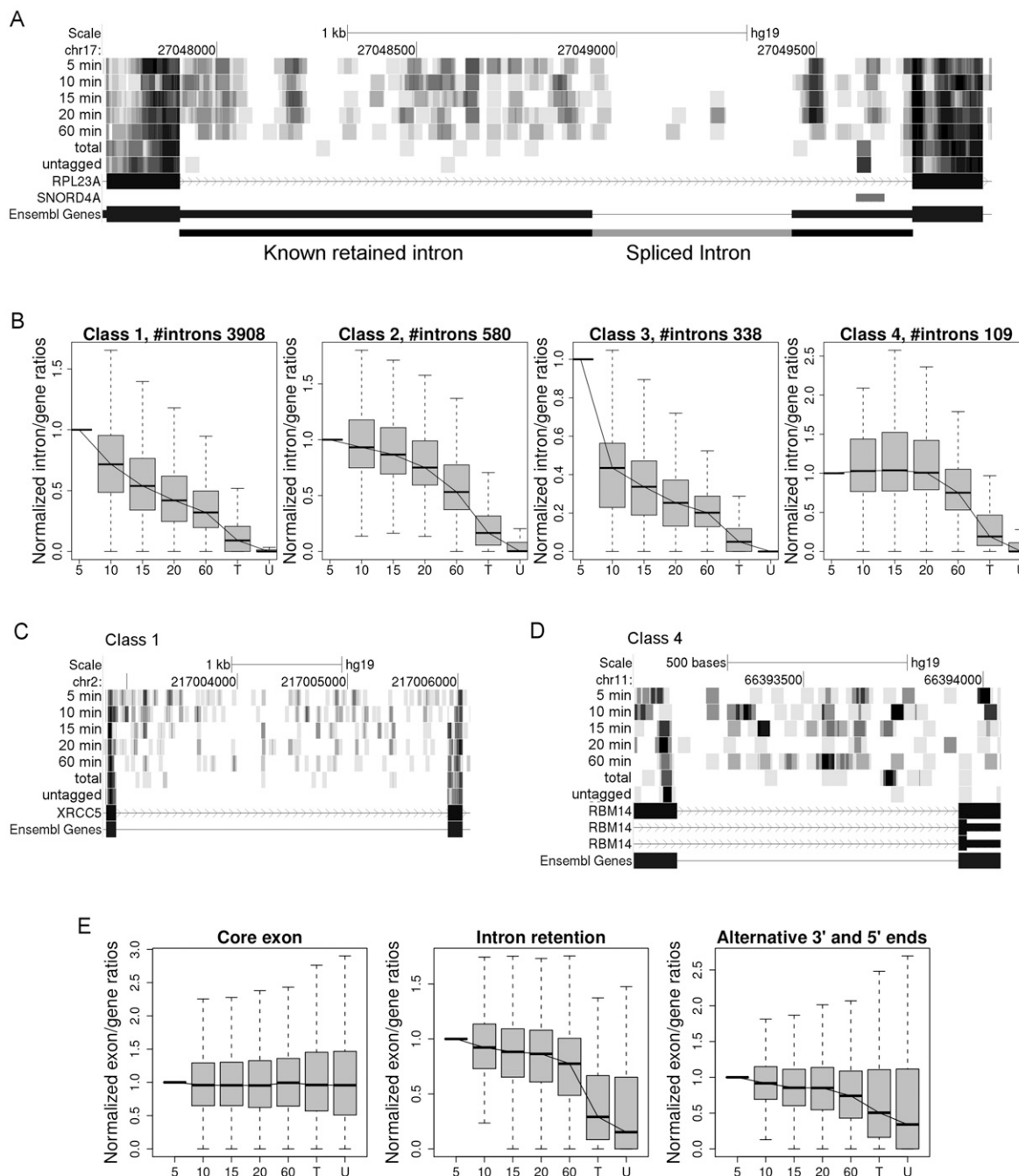
Figure 2. Characterization of distinct intron splicing kinetics. (*A*) Decay of an alternative splicing isoform of the *RPL23A* gene. Read densities for 5- to 60-min 4sU-RNA as well as total and untagged RNA are shown in shades of gray. This figure and all others showing read densities were created using the UCSC Genome Browser (Kent et al. 2002). The corresponding UCSC Genome Browser session containing read density values for all 525 genes analyzed can be accessed via http://www.bio.ifi.lmu.de/en/4sU-seq. Exons are indicated by boxes, introns by lines. Regions that are part of noncoding transcripts are indicated by boxes of smaller height. The known retained intron not translated is additionally indicated in the last line in black. The intron marked in gray is absent in all 4sU-RNA samples (RPKM < 0.5), which is indicative of very fast splicing. In total and untagged RNA, the retained intron is also absent, indicating either nonsense-mediated decay of the alternative splicing product or a secondary splicing event. (*B*) Higher level clustering of the intron clusters identified in the first clustering step (see also Supplemental Fig. 2). Clustering was performed based on intron/gene ratios of cluster representatives normalized to 5-min 4sU-RNA levels. Four classes of intron processing were identified: (Class 1) "Normal" decay with a linear slope in the first 20 min; (Class 2) reduced decay rates; (Class 3) increased decay rates; (Class 4) retained introns that are stable during the first 60 min but are eventually spliced and/or degraded. (*C*) Read density in all samples for an exemplary Class 1 intron. (*D*) Read density for an exemplary Class 4 intron. (*E*) Exon/gene ratios in all samples for core exons, exons constituting retained introns, as well as alternative 3′ and 5′ ends. Again, exon/gene ratios were normalized to ratios in 5-min 4sU-RNA.
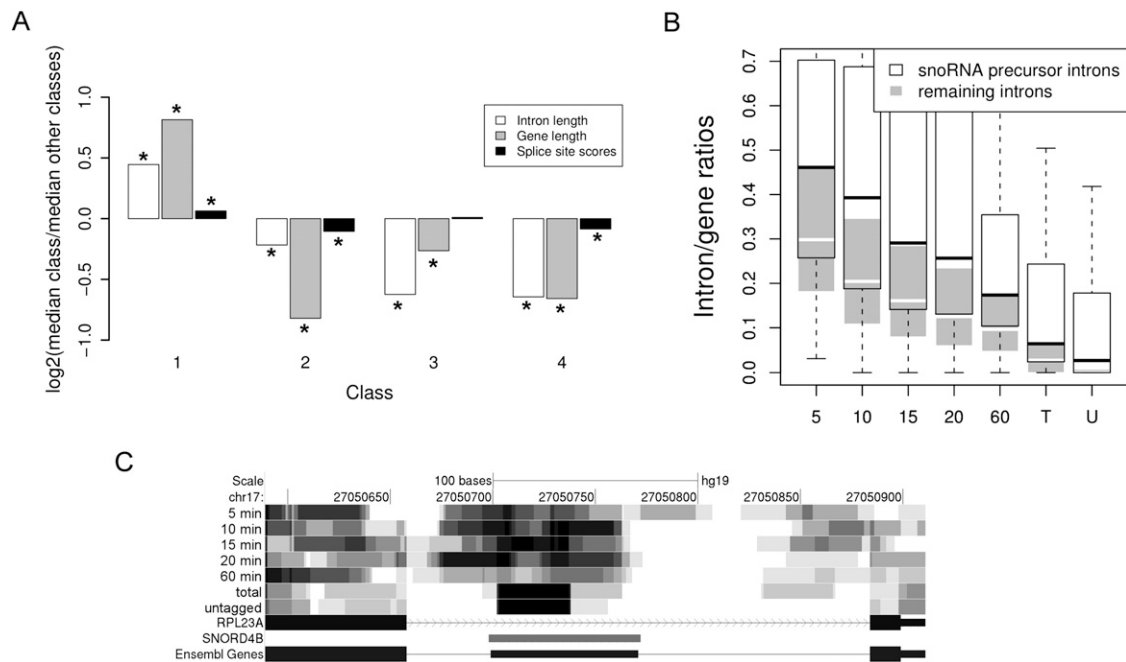
Figure 3. Properties of intron classes and snoRNA processing. (*A*) Log$_2$ fold-changes of the median in each class compared with the median of all other classes are shown for intron length (white), gene length (gray), and splice site scores (black), calculated using MaxEntScan (Yeo and Burge 2004). (*) Statistically significant difference compared with all other classes (Wilcoxon test, FDR corrected *P*-value ≤ 0.05). With the exception of splice site scores in Class 3, distributions showed significant differences for all classes and all features analyzed. (*B*) Intron/gene ratios are shown separately for snoRNA precursor introns (white) and all other introns (gray). Intron/gene ratios for precursor introns were significantly increased compared with all other introns at all times. (*C*) Read density plot illustrating the processing of an exemplary precursor intron to the mature snoRNA. One can observe both the substantially slower decay of the intronic sequences surrounding the mature snoRNAs, which was typical for intronic snoRNAs, and the characteristic 5′ read start pattern of mature snoRNAs in total and untagged RNA.

factors were positively correlated with intron decay rates, it is not surprising that these introns are lost relatively fast compared with the majority of other introns (Classes 2 and 4). Strikingly, Class 3, which was characterized by the most rapid intron decay, contained shorter introns and shorter genes than Class 1, while splice site strength was neither significantly increased nor decreased. In contrast, Classes 2 and 4, which also contained significantly shorter introns, were characterized by both significantly shorter gene length and weaker splice sites compared with Classes 1 and 3 (Wilcoxon test, *P*-value < 10^{-5}). These results suggest that for short introns, gene length and splice site strength make a marked difference, resulting in either very fast or very slow splicing and intron decay. In particular, reduced splice site strength may result in the retention of introns (resulting in Class 4 or Class 2 introns) due to poor recognition of splice sites by the splicing machinery. For the latter two classes, the strongest difference was in intron length with Class 2 having significantly longer introns than Class 4. Thus, the four classes are characterized by distinct combinations of intron length, gene length, and splice site strength.

Characteristics of snoRNA processing

Interestingly, both Class 2 and Class 4 were significantly enriched for small nucleolar RNA (snoRNA) precursor introns. SnoRNAs are small noncoding RNAs mainly involved in the chemical modification of other noncoding RNAs, in particular rRNAs. They are mostly encoded within introns of protein-coding or noncoding genes and are excised from these larger RNAs during splicing (Hirose et al. 2003; Dieci et al. 2009). In our case, 22 out of 580

Class 2 introns (3.8%, Fisher's exact test, FDR corrected *P*-value = 0.0027) and seven out of 109 Class 4 introns (6.4%, Fisher's exact test, FDR corrected *P*-value = 0.011) contained snoRNAs. In contrast, snoRNA-containing introns were underrepresented in Class 1 (52 out of 3908 introns, 1.3%, *P*-value = 0.00024). To analyze this observation in more detail, the RNA-seq data were used to investigate processing of snoRNA transcription units. Almost all human snoRNAs (>90%) are processed from introns with only a single snoRNA generated from each intron (Dieci et al. 2009). For these snoRNAs, splicing is generally required to make the 5′ and 3′ ends of the snoRNA-containing intron accessible to trimming by nuclear RNAses. As the majority of snoRNA-containing genes were found to be among the most highly transcribed genes, this group of rather small noncoding RNAs was ideally suited for further analysis. Again, we focused on the 5838 well expressed introns of which 121 contained snoRNAs. Of these snoRNA precursor introns, 88 (72%) were assigned to one of the four classes. The remaining ones were not assigned to the 20 largest clusters. Please note that in this and all previous analyses reads mapping to snoRNAs were assigned only to the snoRNAs and not used for calculating the intron levels.

As Classes 2 and 4 showed reduced intron decay rates, we wondered whether this might be a general feature of snoRNA-containing introns. Indeed, intron/gene ratios of snoRNA precursor introns were significantly increased in all RNA fractions compared with all other introns (Fig. 3B). This indicates that introns containing snoRNAs are characterized by significantly reduced splicing and degradation rates. Nevertheless, the shape of the curve is the same as for Class 1 with a linear decrease within the

first 20 min and no "bulge" as observed for Classes 2 and 4. Accordingly, snoRNA intron processing generally appears to be a continuous process and not a case of temporary intron retention and delayed processing/decay. Remarkably, splice site scores of snoRNA-containing introns were not significantly reduced compared with other introns. This suggests that the reduced decay rates are not due to poor recognition of the splice site by splicing factors. In contrast, both intron length and length of the corresponding genes were reduced on average by ~40% (Wilcoxon test, P-value = 0.0005) and 60% (P-value < 10^{-15}), respectively. As both short-to-medium intron length and short gene length have been found to be associated with reduced intron decay rates, even if snoRNA encoding introns were excluded from the analysis, this may provide a possible explanation for the smaller splicing rates of snoRNA introns.

One representative example of a snoRNA precursor intron is shown in Figure 3C. Here, the *SNORD4B* snoRNA is excised from an intron of the *RPL23A* gene. In this case, large parts of the precursor intron were clearly present until 60-min 4sU-RNA, whereas only

the mature snoRNA was detected in total and untagged RNA. Interestingly, the mature snoRNAs can be identified by reads starting predominantly at the 5′ end of the snoRNA. These sequences most likely represent cloning products of mature snoRNAs not affected by the fragmentation procedure applied during library preparation. Any fragmentation of these small noncoding RNAs probably resulted in the loss of the resulting snoRNA fragments as the standard SOLiD protocol includes a size selection during library preparation to remove cDNAs without inserts.

Due to this apparent cloning/sequencing bias, determination of snoRNA stability using the RNA-seq data is not possible. Thus, we analyzed snoRNA half-lives obtained using Affymetrix Gene ST 1.0 microarrays in DG75-eGFP and two other human B-cell lines (DG75-10/12 and BCBL-1) (Dölken et al. 2010) based on newly transcribed/total RNA ratios. Surprisingly, snoRNAs were highly enriched among the most short-lived transcripts in all three B-cell lines in the microarray data (Fig. 4A). Here, snoRNAs represented >20 of the 40 most short-lived transcripts in all three human B-cell lines. This was surprising as these small noncoding RNAs
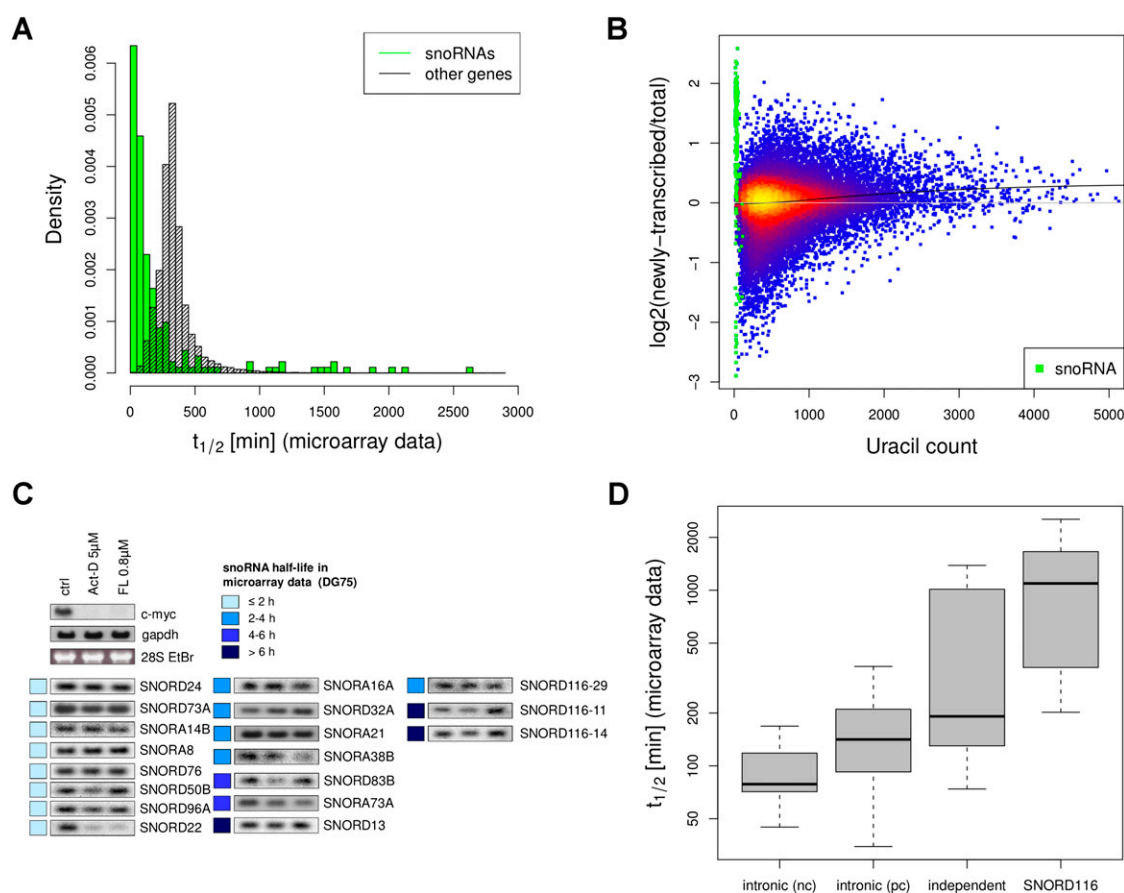

Figure 4. Poor processing efficiency of many, but not all, human snoRNAs. (*A*) Distribution of RNA half-lives ($t_{1/2}$) based on the DG75 microarray data for snoRNAs (green) compared with all other genes (dashed). (*B*) Comparison of the transcript's uracil count against the $\log_2$ (newly transcribed RNA/total RNA ratios) allows the identification of reduced capture rates for short, i.e., uracil-poor, transcripts. The colors indicate the local density of points (yellow, high density; blue, low density). For genes with several transcripts, uracil number was calculated as the average of all transcripts. If no bias in capture rate were observed, log ratios should average around zero (gray line), indicative of no association of RNA half-life with uracil content. Local regression (black line) suggests a slight bias in this example, reflecting reduced capture efficiency of small transcripts, as seen in previous studies. For snoRNAs (green), log ratios are considerably higher than expected for their short length. (*C*) Northern blot data for 15 snoRNAs after transcriptional inhibition by treatment with either 5 μM Act-D or 800 nM FL for 6 h. 28S EtBr served as loading control. For almost all snoRNAs, expression levels after transcriptional inhibition remained unaltered, indicative of their high stability. (*D*) Distribution of RNA half-lives based on the DG75 microarray data for snoRNAs with different expression strategies: intronic (nc), snoRNA expressed as part of an intron of a noncoding gene; intronic (pc), snoRNA expressed as a part of an intron of a protein-coding gene; independent, snoRNA with independent promoter; *SNORD116*, *SNORD116* cluster snoRNAs.

are involved in metabolic processes, which are usually associated with long-lived transcripts (Friedel et al. 2009). In addition, there is so far no evidence for snoRNAs being used up and metabolized while exerting their function in rRNA maturation. Furthermore, snoRNAs are small RNA molecules (60–300 nt; Kiss 2002). Therefore, they are prone to reduced capture rates during 4sU-tagging due to their very low uridine (and thus low 4sU) content (Friedel et al. 2009; Miller et al. 2009). Capture rates can be assessed by comparing the $\log_2$ (newly transcribed/total RNA) ratios with the uridine content of transcripts (Fig. 4B). As expected, average log ratios decreased for very short transcripts with very low uridine content. For virtually all snoRNAs, however, these ratios were not only not reduced, but rather surprisingly high for transcripts of such short length. If we had corrected for size-dependent reduced capture efficiency using computational methods (Miller et al. 2009, 2011), these ratios would have been even higher, resulting in even shorter RNA half-lives.

Remarkably, the apparent short snoRNA half-lives from microarrays could also not be verified by Northern blot analysis. Here, snoRNA levels for eight short-lived, seven medium-lived, and three long-lived snoRNAs (according to the microarray data) were determined following 6-h transcriptional arrest with either Flavipiridol (FL) or Actinomycin D (Act-D). The very short-lived *c-myc* transcript and the long-lived *glyceraldehyde-3-phosphate dehydrogenase* (*GAPDH*) served as controls. While a reduction in snoRNA levels following transcriptional arrest was observed for three snoRNAs (*SNORD22*, *SNORA38B*, and *SNORA73A*), expression of the majority of snoRNAs remained unaltered (Fig. 4C). Although a rapid stabilization of these snoRNAs due to transcriptional arrest cannot be ruled out completely, the alterations were much smaller than for other short-lived RNAs (e.g., *MYC* or *FOS*) known to be rapidly stabilized after transcriptional arrest. We conclude that most mature snoRNAs molecules are not highly unstable, in contrast to what was implied by the microarray data. However, as the majority of snoRNAs are well expressed in human B-cells and were highly consistent in three different cell lines, these measurements are very unlikely to be due to technical problems or artifacts of the microarray measurements.

As microarray probes cannot distinguish between mature snoRNAs and their precursors, we hypothesized that microarrays measured the short RNA half-lives of instable snoRNA precursors, rather than of the mature snoRNAs. Accordingly, the very short half-lives observed for many "snoRNA" probe sets in the microarray data likely represent the half-lives of their parental introns. Thus, they correspond to rapid splicing and decay of the parental introns, which are degraded rather than processed to a (stable) mature snoRNA. In conclusion, the short snoRNA half-lives observed by microarray analysis actually reflect inefficient processing of many snoRNAs, i.e., only a small fraction of introns is actually processed to mature snoRNAs while the majority of introns are simply spliced and degraded. Interestingly, very short microarray-derived half-lives were observed only for snoRNAs encoded within introns of noncoding and protein-coding genes, but not for snoRNAs with independent promotors, which provides further support to this conclusion (Fig. 4D).

In contrast to most other snoRNAs, snoRNAs of the *SNORD116* cluster were found to possess rather long RNA half-lives (>5 h) in the microarray data (Fig. 4D). This would be consistent with substantially more efficient snoRNA processing. It is important to note that their long RNA half-lives were not due to low signal intensities of the corresponding probe sets. These snoRNAs all derive from introns of the *SNURF–SNRPN* transcription unit (Runte et al. 2001), which is expressed from the imprinted *SNURF–SNRPN* domain. It encodes for at least two long, paternally expressed noncoding RNAs, which both encode for multiple intronic snoRNAs (Royo and Cavaille 2008; Vitali et al. 2010). Little is known about their biogenesis and function. The second cluster of snoRNAs (*SNORD115*) encoded in this domain is brain-specific (Cavaille et al. 2000) and was hardly expressed at all in the B-cells under study. Although found only at relatively low levels in our RNA-seq data, the mature snoRNAs of the *SNORD116* cluster were expressed at greater levels than the surrounding intronic sequences (Fig. 5A). Furthermore, reads derived from the *SNORD116* cluster generally started at the 5′ end of the snoRNA even in 5-min 4sU-RNA (Fig. 5B), which is in stark contrast to the other intronic snoRNAs (Fig. 5C). This strongly indicates that these reads derive from mature snoRNAs. Accordingly, most *SNORD116* snoRNAs were already fully processed in 5-min 4sU-RNA, which indicates highly efficient snoRNA processing within the *SNORD116* cluster.
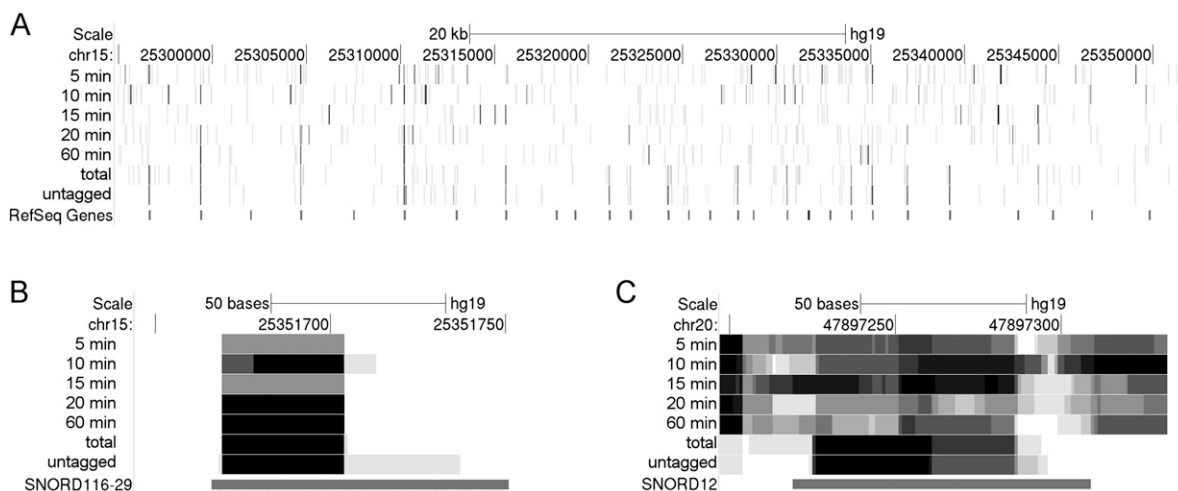

Figure 5. Processing of the SNORD116 cluster. (*A*) Read density for the *SNORD116* cluster region (overview). (*B*) Read density plot for a typical snoRNA of the *SNORD116* cluster only shows reads at the 5′ end of the mature snoRNA, but no intronic sequences, thus representing sequencing of the mature, unfragmented snoRNA. (*C*) Read density plot for a typical intronic snoRNA. Here, intronic sequences were observed until 60-min 4sU-RNA.

Discussion

Metabolic labeling of newly transcribed RNA with 4-thiouridine (4sU-tagging) is superior to simply analyzing total cellular RNA and allows distinguishing alterations in RNA synthesis and decay (Dölken et al. 2008; Friedel et al. 2009; Rabani et al. 2011). In this study, we show that characteristics of RNA processing can be examined by performing ultrashort and progressive 4sU-tagging combined with RNA-seq. In human B-cells, as little as 5 min of 4sU-tagging provided sufficient amounts of newly transcribed RNA to perform RNA-seq. Previously, 4sU-tagging was shortened to 10–15 min but only combined with microarray (Dölken et al. 2008) and nCounter technology (Rabani et al. 2011) as not enough RNA material was thought to be recoverable for sequencing. RNA-seq was only performed following much longer labeling (e.g., 2 h) (Schwanhäusser et al. 2011). In order to shorten the duration of 4sU labeling down to 5 min and still obtain sufficient newly transcribed RNA for next-generation sequencing, we increased the amount of total RNA with which we started from 70 to 150 μg. In addition, we benefited from a slightly increased efficiency of 4sU uptake by cells growing in suspension (B-cells) compared with adherent cells (murine fibroblasts). As 4sU incorporation is strongly dependent on the employed 4sU concentration, this approach is readily adaptable to other cell types by simply increasing the 4sU concentration for ultrashort labeling.

The high purity of the newly transcribed RNA samples was confirmed by the much higher percentage of intronic reads and exon–intron junction reads inversely correlated with the duration of labeling. Interestingly, our data showed that >65% of intronic sequences had already been spliced and decayed in 5-min 4sU-RNA, which highlights the need to perform ultrashort 4sU-tagging. To further characterize processing of introns, we analyzed the relative loss of intronic reads over time (5-min 4sU-RNA to 60-min 4sU-RNA) to define splicing kinetics for >5800 introns and correlated this with various features. Remarkably, intron decay rates were found to be influenced by both intron and gene length. The correlation to intron length was mostly independent of gene length as it was also observed for genes of similar length. As these observations were based on the comparison of read counts in 60-min 4sU-RNA to 5-min 4sU-RNA, it is unlikely that they are due to sequencing artifacts or insufficient length normalization. These problems would affect intron read counts in both samples to a similar degree and, thus, they would not be relevant when calculating ratios of read counts for each intron. In conclusion, although we cannot completely exclude a minor bias introduced during sample and library preparation negatively affecting sequence reads derived from very large genes/pre-mRNAs, our data indicate that any bias introduced did not substantially affect the major findings of this study.

Interestingly, numerous introns were already absent in 5-min 4sU-RNA. Although some of these probably represented non-transcribed variants at the 3′ end of a gene and some may simply reflect sequencing bias, a considerable number were surrounded by well-expressed exons, suggestive of very fast splicing. Remarkably, we still observed exon–intron junction reads for these absent introns albeit at a lower frequency than for introns present in 5-min 4sU-RNA. Furthermore, the number of exon–exon junction reads was not increased compared with the number of exon–intron junction reads. This hints at degradation of these introns concurrently with splicing, which would be consistent with previous results showing that tethering of exons to the RNA polymerase II results in correct splicing of these exons even though the intron connecting them may no longer be continuous, e.g., due to ribozyme-mediated cleavage and partial degradation (Dye et al. 2006).

When looking more closely at the kinetics of splicing, we identified clusters of introns spliced with distinct kinetics accompanied by an enrichment for similar splicing kinetics within a given gene. It is important to note that the reported splicing kinetics were also observed for less abundant genes with more diverse functions. When classes were extended by so far unclassified present introns (RPKM ≥ 0.5) of genes with a lower minimum expression value (RPKM ≥ 5) (2468 genes; 29,947 introns), the same trends were observed for each class as before (Supplemental Fig. 2C). The most interesting kinetics were observed in introns of Class 4 and less pronounced in Class 2. These showed relative high intron levels even in 60-min 4sU-RNA—indicative of retained introns. This was associated with reduced splice site strength most likely resulting in missed splice site recognition contributing to intron retention. Nevertheless, intron levels dropped substantially in total and untagged RNA. This is consistent with a delayed but nevertheless efficient removal of these intronic sequences. Many known and at least 100 novel alternatively spliced transcripts containing retained introns showed this pattern. This temporally delayed but nevertheless eventually efficient removal of retained introns either may be due to RNA degradation by nonsense-mediated RNA decay or may represent novel secondary, i.e., post-transcriptional splicing events. In principle, the two could be distinguished, as secondary splicing events would only remove the initially retained introns while nonsense-mediated decay would also affect levels of the surrounding exons and thus overall transcript RNA half-lives. Unfortunately, the fraction of transcripts of a gene still containing retained introns in 5-min 4sU-RNA was not much larger than 30% on average. Thus, the effect on half-lives was not large enough to differentiate the two alternatives based on the current data. Further studies will thus need to employ knock-out/down approaches targeting nonsense-mediated RNA decay to answer this interesting question. Nevertheless, both alternatives suggest highly effective cellular quality control measures to ensure correct splicing, as the level of transcripts containing retained introns in total and untagged RNA was very low.

Another interesting finding of this study was the poor processing efficiency of many but not all human snoRNAs. In microarray data of newly transcribed/total RNA ratios derived from 1-h 4sU-tagging (Dölken et al. 2010) snoRNAs seemed to be the most short-lived cellular transcripts in three human B-cell lines. Northern blot analysis of mature snoRNAs after 6 h of transcriptional arrest clearly excluded such short RNA half-lives for most of the mature snoRNAs, thereby at first contradicting the microarray-based measurements. However, as microarray probe sets cannot differentiate between a small mature snoRNA and its much larger precursor, these "seemingly" short RNA half-lives only reflect inefficient processing of snoRNAs from their much larger precursors, i.e., degradation of the parental introns without processing to the mature snoRNAs. Poor processing efficiency of snoRNAs derived from intronic sequences indicates competition between splicing factors (resulting in subsequent intron degradation) and the snoRNA processing machinery. Evidence for this hypothesis is provided by the observation that introns containing snoRNAs are spliced and degraded much slower than other introns. It is tempting to speculate that the low basal processing efficiency of many snoRNAs may offer the opportunity for significant regulation of snoRNA expression levels by modifying their processing efficiency. Indeed, up-regulation of snoRNA levels has recently been reported for cells expressing a mutant form of the carboxy-terminal

domain (CTD) of the large subunit of RNA polymerase II (Sims et al. 2011). Expression of a CTD mutant deficient in arginine methylation resulted in a significant increase of steady state levels of a variety of snoRNAs and snRNAs (small nuclear RNAs), while the levels of all other categories of RNAs, e.g., mRNAs, remained unaffected. It will be interesting to test whether this mutant shows alterations in RNA splicing, which might account for the alterations in snoRNA levels. An alternative explanation for the large number of snoRNA-containing introns being spliced and degraded rather than processed into mature snoRNAs would be that the expression levels of snoRNA-binding proteins may not be sufficiently high to bind all newly synthesized snoRNA precursors and thereby prevent their degradation. In this case, snoRNA levels would not be defined transcriptionally but by the abundance of their respective snoRNA-binding proteins.

Interestingly, RNA processing of intronic snoRNAs was not always found to be inefficient as exemplified by the majority of snoRNAs of the *SNORD116* cluster. Although we did observe weak expression of the intronic regions surrounding the snoRNAs, accumulation of reads derived from the 5′ end of the mature snoRNAs in all samples provided evidence that mature snoRNAs were already generated within the first few minutes of synthesis and thus became apparent even in 5-min 4sU-RNA. This also indicates that snoRNA processing can indeed be both fast and efficient. For other snoRNAs with less efficient processing, this clear bias of read starts characteristic for mature snoRNAs was only observed later on, indicating that the majority of their parental introns are not rapidly spliced and processed to snoRNAs. The underlying molecular mechanism of this substantial difference in processing efficiency of the paternally imprinted *SNORD116* snoRNAs remains to be elucidated.

RNA processing involves numerous complex and highly regulated molecular processes. This field of research has recently received a dramatic boost by the development of high-throughput technology for studying RNA–protein interactions at nucleotide resolution. Both HITS-CLIP and PAR-CLIP can now be employed to reveal thousands of protein–RNA interaction sites in a single experiment and thus serve to unravel the underlying functional networks and molecular mechanisms (Ule et al. 2003, 2006; Hafner et al. 2010). Our approach now provides the means to directly study the functional consequences of these RNA–protein interactions on RNA processing. In this study, we restricted our approach to RNA splicing and the processing of the highly abundant snoRNAs. With more in-depth sequencing and sequencing of shorter RNAs, RNA processing of other noncoding RNAs, such as miRNAs or lincRNAs, could also be studied in detail. One of the most important conclusions of this work is that RNA processing efficiency needs to be considered when studying the regulation of RNA expression levels. RNA processing is likely to provide a notable contribution to the regulation of both large and small noncoding RNAs. Application of RNA-seq combined with ultrashort and progressive 4sU-tagging will thus dramatically enhance our understanding of the underlying molecular mechanisms and regulatory networks.

Methods

Cell culture and 4sU-tagging

DG75-eGFP human B-cell lines were cultured in RPMI 1640 medium containing 10% fetal calf serum, 100 IU/mL Penicillin, 100 μg/mL Streptomycin, and 2 mM L-Glutamine. Newly transcribed RNA was labeled for 5, 10, 15, 20, or 60 min using 500 μM 4sU. Total cellular RNA was prepared from cells using TRIzol reagent (Invitrogen) following the modified protocol by Chomczynski and Mackey (1995). Newly transcribed RNA was purified from 150 μg (5- to 15-min 4sU), 100 μg (20-min 4sU), and 70 μg total RNA (60-min 4sU). Separation of total RNA into newly transcribed and untagged preexisting RNA was performed as described (Dölken et al. 2008) using 100 μL streptavidin beads (Miltenyi Biotec).

Next-generation sequencing

RNA sequencing was performed using the sequencing-by-ligation SOLiD II platform (Applied Biosystems). A modified whole-transcriptome analysis (WTAK) protocol was employed. Three-hundred nanograms of RNA was used in the library construction process. Volume of the samples was adjusted by vacuum centrifugation, as concentrations of the early newly transcribed RNA samples tended to be low after depletion. Integrity, size distribution, and yield were monitored after RNaseIII fragmentation (10 min, 37°C) on an Agilent RNA 6000 PicoChip. If newly transcribed RNA samples resulted in larger than expected fragments, samples were redigested using RNaseIII for additional 2 min at 37°C, repurified, and rechecked as described above. Libraries of individual samples were prepared using barcoded adaptors and sequenced on the SOLiD instrument using standard protocols. No deviation in quality or length of the newly transcribed RNA sequences was observed when comparing with other standard RNAs.

Read mapping

Reads were mapped in a three-step process using the Bowtie alignment program (Langmead et al. 2009). First, reads were aligned to pre-rRNA sequences (18S, 5.8S, 28S, and spacer regions). The remaining unmapped reads were aligned to all Ensembl transcripts (Ensembl version 60) excluding pseudo-genes and haplotypes to identify exonic and exon–exon junction reads (aligned reads overlapping an exon–exon junction by ≥1 nt). Reads that remained unmapped after step two were aligned to the human reference genome (GRCh37/hg19) to identify intron and exon–intron junction reads (overlapping an exon–intron junction by ≥1 nt). For each read the best alignment location was used. Reads mapping equally well to two different locations were discarded. The following Bowtie settings were used: seed region = first 20 nt, three mismatches allowed in the seed, five in the whole alignment.

Quantification of gene, exon, and intron expression levels

Expression levels of genes, exons, and introns were estimated using the standard RPKM measure (number of reads per kilobase of gene, exon, or intron per million mapped reads) (Mortazavi et al. 2008). Number of reads mapping to a gene was determined as the total number of exon and exon–exon junction reads for this gene. To calculate RPKM values for exons and introns respectively, only reads mapping completely within this region were used. To account for problems in mapping reads to repetitive sequence regions, the effective length of exons and introns was used instead of the actual length. The effective length was calculated in the following way. First, in silico reads were simulated by sliding a window across gene regions with the size of the read length in the experiment (35 nt). Thus, the simulated read set contains exactly one read from each position in each gene. The simulated reads were then mapped using the same three-step procedure described above. The effective length was then defined as the number of positions within the respective region (exon, intron, or gene) which

had exactly one correctly and uniquely mapped read starting at this position.

Intron clustering

Introns were clustered based on the ratios of intron RPKM to gene RPKM (intron/gene ratio) across all samples (feature vector for clustering = intron/gene ratios at 5-, 10-, 15-, 20-, 60-min 4sU-RNA, total and untagged RNA). Introns were first clustered 100 times using standard k-means clustering (k = 10) and Euclidean distance metrics. Final clusters were then defined as introns always clustered together in the 100 k-means runs. To identify classes among clusters, representative introns were determined first for each cluster as those with the smallest Euclidian distance to the cluster median. Exon/intron ratios of the representatives were then normalized to the exon/intron ratios in 5-min 4sU-RNA. Cluster representatives were again clustered 100 times with k-means for values of k between 3 and 10. For each k, the most frequent clustering was determined. For the final clustering, the k was chosen that maximized this frequency (in this case, k = 4 with a maximal frequency of ~90%).

Statistical tests

Significance of differences between two distributions was assessed using the Wilcoxon rank sum test. This test is based on the ranks of observations and, thus, does not require any assumptions on the type of distribution (nonparametric test). Significance of enrichment was evaluated using Fisher's exact test, which tests for a significant association between two different types of classification. Correction for multiple testing was performed using the method by Benjamini and Hochberg (1995) for control of the false discovery rate (FDR).

Data access

Sequencing data including raw sequencing reads have been submitted to the NCBI Gene Expression Omnibus (GEO) (http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE31653.

Acknowledgments

References

Benjamini Y, Hochberg Y. 1995. Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J R Stat Soc Ser B Methodol* **57:** 289–300.
Cavaille J, Buiting K, Kiefmann M, Lalande M, Brannan CI, Horsthemke B, Bachellerie JP, Brosius J, Huttenhofer A. 2000. Identification of brain-specific and imprinted small nucleolar RNA genes exhibiting an unusual genomic organization. *Proc Natl Acad Sci* **97:** 14311–14316.
Cazalla D, Yario T, Steitz JA. 2010. Down-regulation of a host microRNA by a Herpesvirus saimiri noncoding RNA. *Science* **328:** 1563–1566.
Chomczynski P, Mackey K. 1995. Short technical reports. Modification of the TRI reagent procedure for isolation of RNA from polysaccharide- and proteoglycan-rich sources. *Biotechniques* **19:** 942–945.
Cleary MD, Meiering CD, Jan E, Guymon R, Boothroyd JC. 2005. Biosynthetic labeling of RNA with uracil phosphoribosyltransferase allows cell-specific microarray analysis of mRNA synthesis and decay. *Nat Biotechnol* **23:** 232–237.
Clement JQ, Qian L, Kaplinsky N, Wilkinson MF. 1999. The stability and fate of a spliced intron from vertebrate cells. *RNA* **5:** 206–220.
Dieci G, Preti M, Montanini B. 2009. Eukaryotic snoRNAs: A paradigm for gene expression flexibility. *Genomics* **94:** 83–88.
Dölken L, Ruzsics Z, Rädle B, Friedel CC, Zimmer R, Mages J, Hoffmann R, Dickinson P, Forster T, Ghazal P, et al. 2008. High-resolution gene expression profiling for simultaneous kinetic parameter analysis of RNA synthesis and decay. *RNA* **14:** 1959–1972.
Dölken L, Malterer G, Erhard F, Kothe S, Friedel CC, Suffert G, Marcinowski L, Motsch N, Barth S, Beitzinger M, et al. 2010. Systematic analysis of viral and cellular microRNA targets in cells latently infected with human γ-herpesviruses by RISC immunoprecipitation assay. *Cell Host Microbe* **7:** 324–334.
Dye MJ, Gromak N, Proudfoot NJ. 2006. Exon tethering in transcription by RNA polymerase II. *Mol Cell* **21:** 849–859.
Friedel CC, Dölken L. 2009. Metabolic tagging and purification of nascent RNA: Implications for transcriptomics. *Mol Biosyst* **5:** 1271–1278.
Friedel CC, Dölken L, Ruzsics Z, Koszinowski UH, Zimmer R. 2009. Conserved principles of mammalian transcriptional regulation revealed by RNA half-life. *Nucleic Acids Res* **37:** e115. doi: 10.1093/nar/gkp542.
Hafner M, Landthaler M, Burger L, Khorshid M, Hausser J, Berninger P, Rothballer A, Ascano M Jr, Jungkamp AC, Munschauer M, et al. 2010. Transcriptome-wide identification of RNA-binding protein and microRNA target sites by PAR-CLIP. *Cell* **141:** 129–141.
Hirose T, Shu MD, Steitz JA. 2003. Splicing-dependent and -independent modes of assembly for intron-encoded box C/D snoRNPs in mammalian cells. *Mol Cell* **12:** 113–123.
Jing Q, Huang S, Guth S, Zarubin T, Motoyama A, Chen J, Di Padova F, Lin SC, Gram H, Han J. 2005. Involvement of microRNA in AU-rich element-mediated mRNA instability. *Cell* **120:** 623–634.
Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. 2002. The human genome browser at UCSC. *Genome Res* **12:** 996–1006.
Kenzelmann M, Maertens S, Hergenhahn M, Kueffer S, Hotz-Wagenblatt A, Li L, Wang S, Ittrich C, Lemberger T, Arribas R, et al. 2007. Microarray analysis of newly synthesized RNA in cells and animals. *Proc Natl Acad Sci* **104:** 6164–6169.
Kim HD, Shay T, O'Shea EK, Regev A. 2009. Transcriptional regulatory circuits: Predicting numbers from alphabets. *Science* **325:** 429–432.
Kiss T. 2002. Small nucleolar RNAs: An abundant group of noncoding RNAs with diverse cellular functions. *Cell* **109:** 145–148.
Lamond AI, Konarska MM, Grabowski PJ, Sharp PA. 1988. Spliceosome assembly involves the binding and release of U4 small nuclear ribonucleoprotein. *Proc Natl Acad Sci* **85:** 411–415.
Langmead B, Trapnell C, Pop M, Salzberg SL. 2009. Ultrafast and memory-efficient alignment of short DNA sequences to the human genome. *Genome Biol* **10:** R25. doi: 10.1186/gb-2009-10-3-r25.
Melvin WT, Milne HB, Slater AA, Allen HJ, Keir HM. 1978. Incorporation of 6-thioguanosine and 4-thiouridine into RNA. Application to isolation of newly synthesised RNA by affinity chromatography. *Eur J Biochem* **92:** 373–379.
Miller MR, Robinson KJ, Cleary MD, Doe CQ. 2009. TU-tagging: Cell type-specific RNA isolation from intact complex tissues. *Nat Methods* **6:** 439–441.
Miller C, Schwalb B, Maier K, Schulz D, Dümcke S, Zacher B, Mayer A, Sydow J, Marcinowski L, Dölken L, et al. 2011. Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast. *Mol Syst Biol* **7:** 458. doi: 10.1038/msb.2010.112.
Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* **5:** 621–628.
Nam K, Lee G, Trambley J, Devine SE, Boeke JD. 1997. Severe growth defect in a *Schizosaccharomyces pombe* mutant defective in intron lariat degradation. *Mol Cell Biol* **17:** 809–818.
Nilsen TW, Graveley BR. 2010. Expansion of the eukaryotic proteome by alternative splicing. *Nature* **463:** 457–463.
Pan Q, Shai O, Lee LJ, Frey BJ, Blencowe BJ. 2008. Deep surveying of alternative splicing complexity in the human transcriptome by high-throughput sequencing. *Nat Genet* **40:** 1413–1415.
Rabani M, Levin JZ, Fan L, Adiconis X, Raychowdhury R, Garber M, Gnirke A, Nusbaum C, Hacohen N, Friedman N, et al. 2011. Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. *Nat Biotechnol* **29:** 436–442.
Royo H, Cavaille J. 2008. Non-coding RNAs in imprinted gene clusters. *Biol Cell* **100:** 149–166.

Runte M, Huttenhofer A, Gross S, Kiefmann M, Horsthemke B, Buiting K. 2001. The IC-*SNURF-SNRPN* transcript serves as a host for multiple small nucleolar RNA species and as an antisense RNA for *UBE3A*. *Hum Mol Genet* **10:** 2687–2700.

Schwanhäusser B, Busse D, Li N, Dittmar G, Schuchhardt J, Wolf J, Chen W, Selbach M. 2011. Global quantification of mammalian gene expression control. *Nature* **473:** 337–342.

Shalem O, Dahan O, Levo M, Martinez MR, Furman I, Segal E, Pilpel Y. 2008. Transient transcriptional responses to stress are generated by opposing effects of mRNA production and degradation. *Mol Syst Biol* **4:** 223. doi: 10.1038/msb.2008.59.

Sims RJ III, Rojas LA, Beck D, Bonasio R, Schuller R, Drury WJ III, Eick D, Reinberg D. 2011. The C-terminal domain of RNA polymerase II is modified by site-specific methylation. *Science* **332:** 99–103.

Smit AFA, Hubley R, Green P. 1996–2010. RepeatMasker Open-3.0. http://www.repeatmasker.org.

Ule J, Jensen KB, Ruggiu M, Mele A, Ule A, Darnell RB. 2003. CLIP identifies Nova-regulated RNA networks in the brain. *Science* **302:** 1212–1215.

Ule J, Stefani G, Mele A, Ruggiu M, Wang X, Taneri B, Gaasterland T, Blencowe BJ, Darnell RB. 2006. An RNA map predicting Nova-dependent splicing regulation. *Nature* **444:** 580–586.

Vitali P, Royo H, Marty V, Bortolin-Cavaille ML, Cavaille J. 2010. Long nuclear-retained non-coding RNAs and allele-specific higher-order chromatin organization at imprinted snoRNA gene arrays. *J Cell Sci* **123:** 70–83.

Wang ET, Sandberg R, Luo S, Khrebtukova I, Zhang L, Mayr C, Kingsmore SF, Schroth GP, Burge CB. 2008. Alternative isoform regulation in human tissue transcriptomes. *Nature* **456:** 470–476.

Weintz G, Olsen JV, Frühauf K, Niedzielska M, Amit I, Jantsch J, Mages J, Frech C, Dölken L, Mann M, et al. 2010. The phosphoproteome of toll-like receptor-activated macrophages. *Mol Syst Biol* **6:** 371. doi: 10.1038/msb.2010.29.

Yeo G, Burge CB. 2004. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. *J Comput Biol* **11:** 377–394.