A comparison of metabolic labeling and statistical methods to study genome-wide dynamics of RNA turnover
Differential exon usage

Exon usage was analyzed using DEXSeq [1] for each labeling duration by comparing the 'labeled' and 'unlabeled' samples of all nucleotide conversion protocols, as determined from the pulseR workflow (Methods). The supporting scripts and complete results are available at https://github.com/dieterich-lab/ComparisonOfMetabolicLabeling.

In this analysis, we used a threshold on the log2 fold change (LFC > 1.5), and a stringent p-value cutoff (padj $< 1e-5$). We used the independent filtering implemented in DEXSeq, and did not prefilter isoforms before construction of the counting bins, although this could improve performance. We did not use the default aggregation of overlapping genes.

Overall, a larger number of differentially used (DU) exons and associated transcripts were identified at shorter labeling durations (Fig. 1). The number of non-significant exons (and associated transcripts) were 280,312 (159,741) at 1h, 318,690 (170,837) at 2h, 325,094 (171,435) at 4h, and 333,371 (173,245) at 8h. Transcripts can contain both DU and non-significant exons. We were thus particularly interested in DU exons that were uniquely associated with a given transcript isoform, which could indicate the existence of decay rate-dependent 'isoform switching'. After 1h of labeling, and at the given detection threshold, 1,250 unique isoforms were identifed with at a single defining DU exon (Fig. 2a). The associated host gene set was enriched in terms related to cell division, mRNA export from nucleus, protein ubiquitination and deubiquitination. After 2h, 1,255 unique isoforms were identifed, and the host genes were enriched for terms such as intracellular protein transport, mRNA splicing, via spliceosome, and protein stabilization. After 4h of labeling, only 308 isoforms were identified, and their host genes were enriched for terms such as neutrophil degranulation, regulation of macroautophagy, post-translational protein modification, protein peptidyl-prolyl isomerization, and protein folding. Consistent with previous observations, a large proportion of the fastest isoforms were retained introns, but there were a relatively important number of protein coding transcripts (Fig. 2b). Finally, to illustrate how isoform switching is dependent on exon usage, we show the example of the SMAD2 gene, which appears to express a longer isoform that is rapidly transcribed and degraded, and which is replaced with more stable or slower isoforms (Fig. 3). All these observations raise the possibility that isoform switching may be influenced by varying and coordinated decay rates, and potentially involved in much broader functions than *e.g.* development, or disease progression like cancer. However, further work and data would be required to fully characterize these mechanisms, *e.g.* using shorter pulse labeling, but short (or ultra-short) pulse labeling may not be suitable with nucleotide conversion protocols, which typically require longer labeling times.

References

[1] Anders, s. *et al.* Detecting differential usage of exons from RNA-seq data. *Genome Research* **22**, 4025, http://www.genome.org/cgi/doi/10.1101/gr.133744.111 (2012).

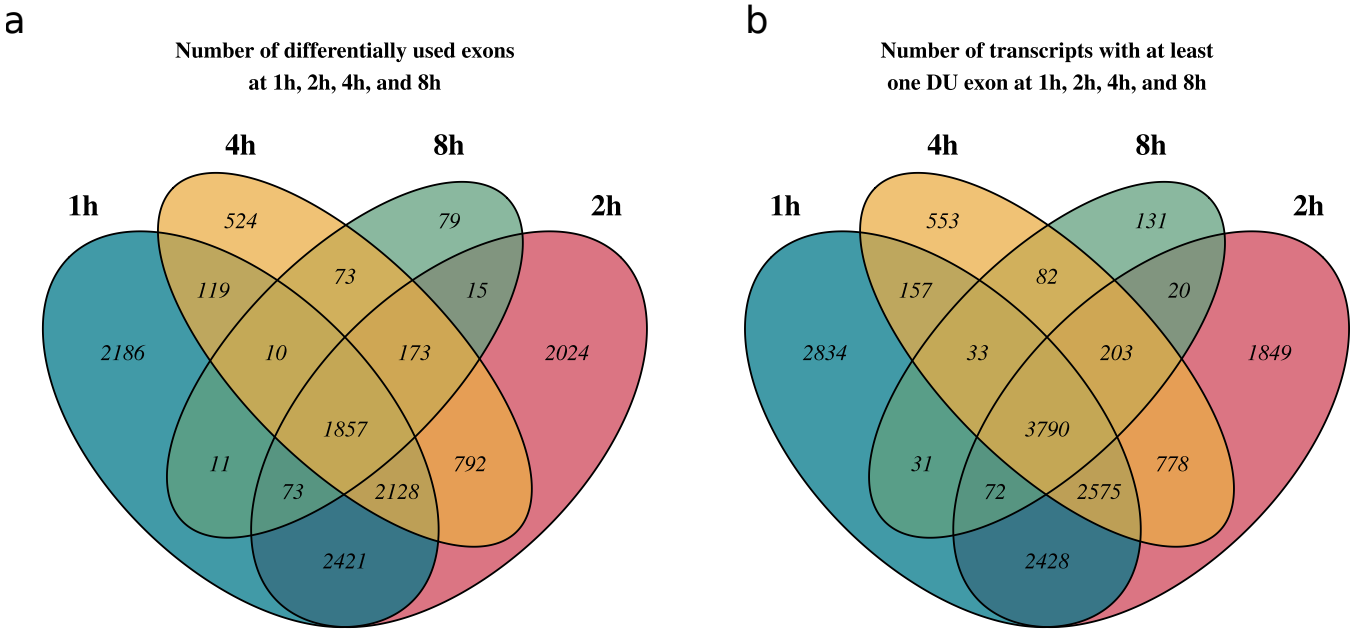

Figure 1: **a** Differentially used (DU) exons and **b** associated transcripts (with at least one DU exon)

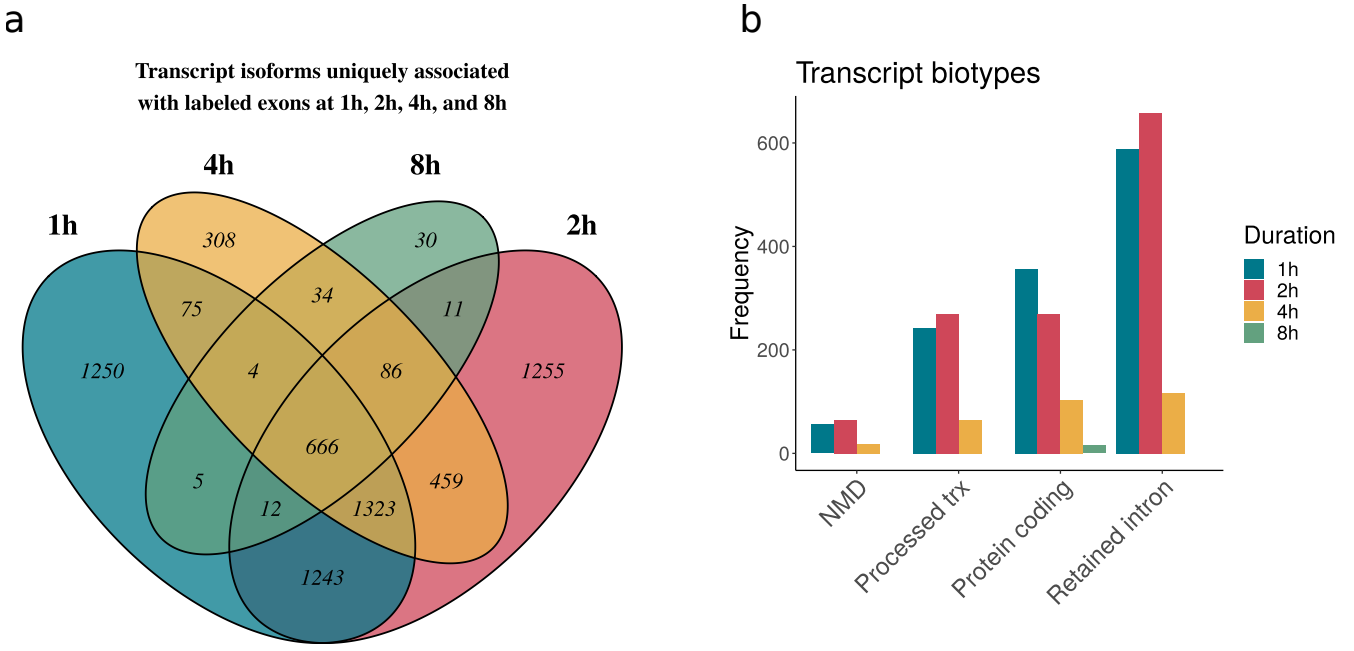

Figure 2: **a** DU exons-defining transcript isoforms. **b** Transcript biotypes for isoforms uniquely found at 1h, 2h, 4h, and 8h. See panel a.

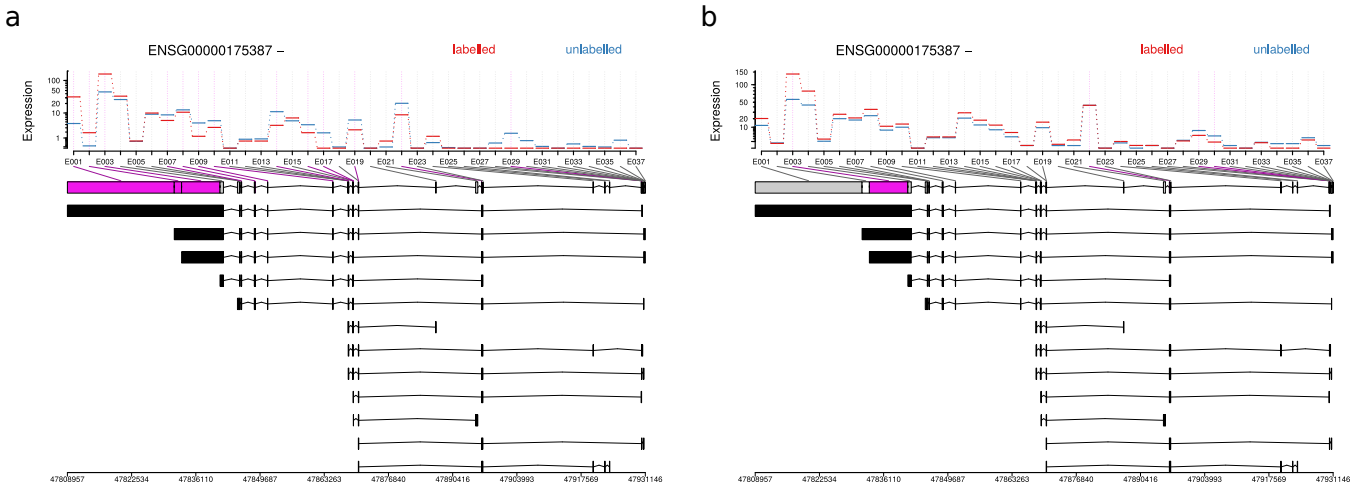

Figure 3: Example of the SMAD2 gene which shows a shift in isoform usage due to a DU exon at **a** 1h and **b** 8h.