Method

Genome-wide dynamics of RNA synthesis, processing, and degradation without RNA metabolic labeling

Mattia Furlan,[1,2] Eugenia Galeota,[1] Nunzio Del Gaudio,[1] Erik Dassi,[3] Michele Caselle,[2] Stefano de Pretis,[1,4] and Mattia Pelizzola[1,4]

[1]Center for Genomic Science, Fondazione Istituto Italiano di Tecnologia, 20139 Milan, Italy; [2]Physics Department and INFN, University of Turin, 10125 Turin, Italy; [3]Centre for Integrative Biology, University of Trento, 38123 Trento, Italy

The quantification of the kinetic rates of RNA synthesis, processing, and degradation are largely based on the integrative analysis of total and nascent transcription, the latter being quantified through RNA metabolic labeling. We developed INSPEcT−, a computational method based on the mathematical modeling of premature and mature RNA expression that is able to quantify kinetic rates from steady-state or time course total RNA-seq data without requiring any information on nascent transcripts. Our approach outperforms available solutions, closely recapitulates the kinetic rates obtained through RNA metabolic labeling, improves the ability to detect changes in transcript half-lives, reduces the cost and complexity of the experiments, and can be adopted to study experimental conditions in which nascent transcription cannot be readily profiled. Finally, we applied INSPEcT− to the characterization of post-transcriptional regulation landscapes in dozens of physiological and disease conditions. This approach was included in the INSPEcT Bioconductor package, which can now unveil RNA dynamics from steady-state or time course data, with or without the profiling of nascent RNA.

[Supplemental material is available for this article.]

Since the development of microarrays first, and high-throughput sequencing later on, the investigation of the transcriptional activity of genes has been mostly based on the quantification of total RNA (Mortazavi et al. 2008). While bringing about a revolution in the field of transcriptional regulation, the quantification of absolute and differential expression provides only a glimpse of the complexity of cellular gene expression programs. Indeed, abundance and responsiveness to modulations of premature and mature RNA species are set by the combined action of three key steps: premature RNA synthesis, processing of premature into mature RNA, and degradation of the latter (Orphanides and Reinberg 2002). These steps are governed by corresponding kinetic rates, which collectively determine the RNA dynamics of transcripts (Fig. 1A). However, the specific contribution of each step of the RNA life cycle cannot be deconvoluted from an aggregate quantity like the amount of total RNA because, in principle, infinite combinations of kinetic rates can generate the same absolute expression level.

For decades, the study of RNA dynamics relied solely on transcription blockage experiments. However, these methods allow the quantification of RNA half-lives only, are highly invasive, affect cell viability, and could alter various pathways, RNA decay included (Wada and Becskei 2017). To overcome these limitations, new methods have been developed that are based on the integrative analysis of total and nascent RNA. Nascent RNA can be metabolically labeled with biotinylated, 4-thiouridine (4sU)–modified nucleotides, purified with streptavidin, and then sequenced (Dolken et al. 2008; Miller et al. 2011; Rabani et al. 2011;

Wissink et al. 2019). Alternatively, if the modified nucleotides are chemically derivatized before sequencing, reads from nascent transcripts can be in silico separated from pre-existing RNA (Herzog et al. 2017; Baptista and Dölken 2018; Jürges et al. 2018; Schofield et al. 2018). A number of methods were developed for the quantification of RNA dynamics via metabolic labeling, including INSPEcT (de Pretis et al. 2015), DRUID (Lugowski et al. 2018), cDTA (Sun et al. 2012), GRAND-SLAM (Jürges et al. 2018), pulseR (Uvarovskii and Dieterich 2017), and DRiLL (Rabani et al. 2014). Eventually, these approaches have started to unveil how the modulation of RNA dynamics can determine gene-specific regulatory modes and elicit complex transcriptional responses (Rabani et al. 2014; de Pretis et al. 2015, 2017; Furlan et al. 2019; Tesi et al. 2019).

Despite their advantages and popularity, methods based on RNA metabolic labeling are affected by various pitfalls, especially when a limited amount of nascent RNA is produced and when aiming at studying very short responses (Baptista and Dölken 2018). Moreover, these methods cannot be readily applied to model organisms, be it mammals (Matsushima et al. 2018) or plants (Sidaway-Lee et al. 2014), in vivo. For all these reasons, being able to study RNA dynamics from just total RNA would be a valuable alternative. A few studies have moved in this direction by using an integrative analysis of premature and mature RNA abundances (Zeisel et al. 2011; Gray et al. 2014; La Manno et al. 2018; Bergen et al. 2020), yet they have fallen short of quantifying the full set of RNA kinetic rates and their modulation. Specifically, the key limitation of all these studies is having considered intronic expression as a proxy of synthesis rates. Although this greatly simplifies the problem from a mathematical point of view, it neglects
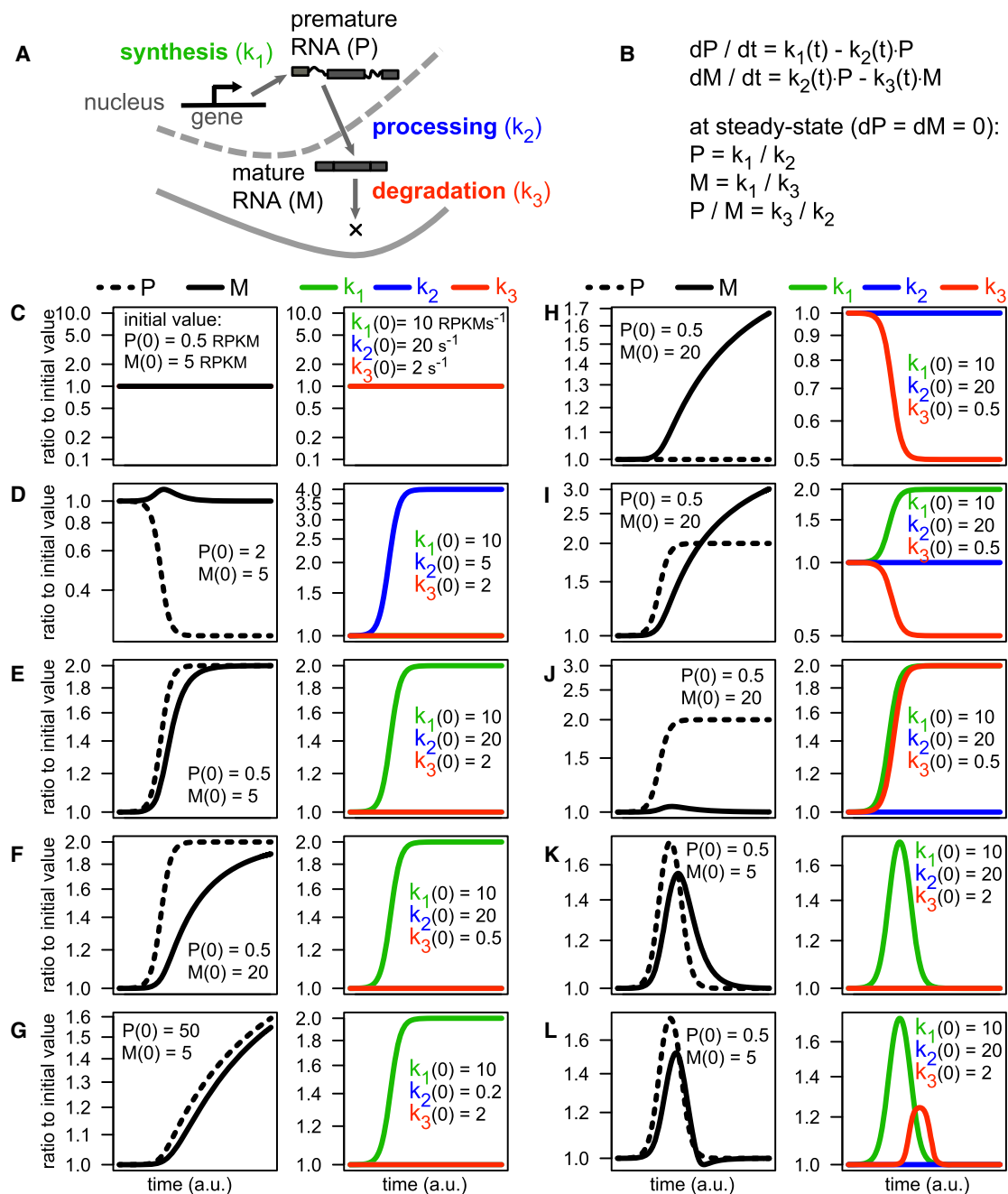

Figure 1. The influence of RNA kinetic rates on RNA abundance and responsiveness. (*A*) Schematic representation of the RNA life cycle, governed by the kinetics rates of synthesis, processing and degradation. (*B*) Deterministic mathematical model of the RNA life cycle based on ordinary differential equations (ODEs), including the solution of the system at steady state. (*C–L*) Solutions of the ODE system following the modulation of the kinetic rates: Each example reports, for premature and mature RNA species (*left*) and for the kinetic rates (*right*), the ratio to the initial time point. Initial values are indicated within each panel.

that intronic RNA-seq signals result from the joint action of two processes: the synthesis of premature RNA and its processing into the mature form. Therefore, these approaches neglected the contribution of RNA processing and relied on the strong assumption that the rate of processing is constant.

To cope with these key limitations, while avoiding all the downsides of RNA metabolic labeling, we developed INSPEcT–, a computational approach that determines RNA dynamics from to-

tal RNA-seq data. INSPEcT– quantifies the full set of kinetic rates from time course RNA-seq data sets and enables the study of post-transcriptional regulation between steady-state conditions. We used INSPEcT– to analyze different time course RNA-seq data sets, ranging from conditions in which gene expression programs are mostly controlled by transcriptional changes to conditions in which post-transcriptional regulation prevails. Finally, we used this method to characterize post-transcriptional regulation

landscapes in dozens of tissue types and disease conditions. INSPEcT– is available within the INSPEcT Bioconductor package (http://bioconductor.org/packages/INSPEcT/), formerly developed by us for the analysis of RNA metabolic labeling data (de Pretis et al. 2015), which now allows the user to study RNA dynamics on steady-state or time course data, with or without nascent RNA profiling.

Results

The quantification of RNA dynamics unveils the complexity of gene expression programs

At steady state, the abundance of premature RNA is equal to the ratio of its synthesis to its processing rate, and the quantity of its mature form is given by its synthesis to degradation rate ratio (Fig. 1A, B). Thus, although the rate of RNA synthesis influences the abundance of both premature and mature RNAs, processing and degradation rates impact just on premature and mature forms, respectively.

At the transition between steady states, both the new level of transcript abundance and the speed of the transition (responsiveness) depend on RNA kinetic rates. In the most straightforward case, differential expression—the regulation of the cellular abundance of premature (P) and mature (M) RNA species—derives from changes in the rate of RNA synthesis only (k_1). This implies a change in the amount of nascent RNA for a given gene. Although it is often assumed, this should be experimentally confirmed by RNA metabolic labeling before concluding that changes in P or M are transcriptional in nature. In all other cases, differential expression can entail more complex co- and post-transcriptional mechanisms, each governed by processing (k_2) and/or degradation (k_3) rate. Solving the system depicted in Figure 1B permits the determination of the impact one or more kinetic rates can have on the abundance of P and M when modulated over time:

- Constant kinetic rates define steady states where P and M abundances are calculated as k_1/k_2 and k_1/k_3 ratios, respectively (Fig. 1B,C).
- Modulations in the processing rate k_2 cause just transient variations in M abundance but permanent alterations in P abundance (Fig. 1D).
- M responsiveness to changes in k_1 depends on the level of k_3 (Fig. 1, cf. E and F; Friedel et al. 2009; Zeisel et al. 2011), and of k_2 (Fig. 1, cf. E and G).
- k_1 and k_3 can separately generate the same type of M variation if changing in opposite directions (Fig. 1F,H), whereas only the modulation of k_1 can affect P (Fig. 1F).
- k_1 and k_3 reinforce each other's modulation of M when changing simultaneously in opposite directions (Fig. 1, cf. E and I), whereas they neutralize each other's impact on M if simultaneously adjusted in the same direction (Fig. 1J).
- Transient alterations in M induced by a temporary change in k_1 (Fig. 1K) can be made sharper by a concomitant change in k_3 (Fig. 1L; Rabani et al. 2011).

First, these examples indicate that measurements of mature RNA are in themselves poorly informative of the real transcriptional state of genes. For instance, the detection of a mature RNA is typically taken as indication that the corresponding gene is transcriptionally active. This is not necessarily the case for highly stable RNAs, which might persist long after the gene has become silent. Second, these examples illustrate how difficult it is to deci-

pher the mechanism responsible for modulating mature RNAs without determining the corresponding RNA dynamics. For instance, the modulation of mature RNA species is typically seen as indication of transcriptional regulation, whereas it could originate from changes in the dynamics of processing and/or degradation, without any change in the rate of synthesis taking place. Ultimately, these examples show the necessity to develop methods for the quantification of RNA kinetic rates in order to fully disclose the mechanisms behind complex responses in gene expression.

Experimental and computational pitfalls of RNA metabolic labeling experiments

The steady-state solution introduced in Figure 1B is underdetermined (two equations and three unknown kinetic rates), and the original ordinary differential equations system does not allow the identification of a unique set of kinetic rates as well. The key to solving these systems is to use RNA metabolic labeling with short time pulses, so that the quantification of nascent RNA can be used as a proxy for the synthesis rate (de Pretis et al. 2015). There are two main types of RNA metabolic labeling experiments: one involving the purification of labeled RNA species (Dolken et al. 2008) and the other requiring their chemical derivatization before in silico identification (Baptista and Dölken 2018). Both categories of methods are characterized by specific pitfalls, and special care is needed when designing these experiments, particularly when deciding on number of replicates, sequencing coverage, and length of labeled nucleotides pulse (Uvarovskii et al. 2019).

Methods based on the purification of nascent RNA present three main drawbacks: (1) higher costs owing to the need to sequence both total and labeled RNA populations, (2) the need to normalize the signal from the nascent RNA population to that from the total (or pre-existing) RNA population, and (3) the contamination of labeled with unlabeled (pre-existing) RNA molecules. The need for normalization has been partially addressed by introducing internal standards (Sun et al. 2012) or through computational normalization procedures (Rabani et al. 2014; de Pretis et al. 2015). Rather, the problem of contamination issue is typically acknowledged but left unsolved. To quantify contamination and to verify whether it varies with the duration of the 4sU pulse, we measured the amount of labeled RNA that can be recovered with pulses of 4sU lasting from 10 min to 2 h (Fig. 2A,B). A model based on a constant rate of 4sU incorporation into nascent transcripts did not fit our data, suggesting that the incorporation rate depends on the pulse length (Fig. 2C). Indeed, a model based on an exponential increase of the incorporation rate did fit the data better (log likelihood-ratio test $P = 2 \times 10^{-27}$) (Fig. 2C,D). A model assuming a constant contamination rate, not dependent on the 4sU pulse length, further increased data fitting ($P = 3.1 \times 10^{-7}$). Rather, a model in which the contamination increased linearly with pulse length did not improve the fitting any further, and reverted to the constant-contamination hypothesis ($P = 1$; $a \approx 0$ in Fig. 2C). Altogether, we determined that 10-min-long 4sU pulses, which were often used in these studies (Miller et al. 2011; Rabani et al. 2011, 2014; Sun et al. 2012; Fuchs et al. 2014, 2015; Sabò et al. 2014; de Pretis et al. 2015; Marzi et al. 2016; de Pretis et al. 2017; Michel et al. 2017), led to 30% of the labeled fraction being originated through contamination of the pre-existing RNA population. In an independent study in which dendritic cells were subjected to 10-min-long 4sU pulses, 30% of the unlabeled RNA was estimated to contaminate the labeled fraction, suggesting that the percentage of labeled RNA being contaminated is even higher
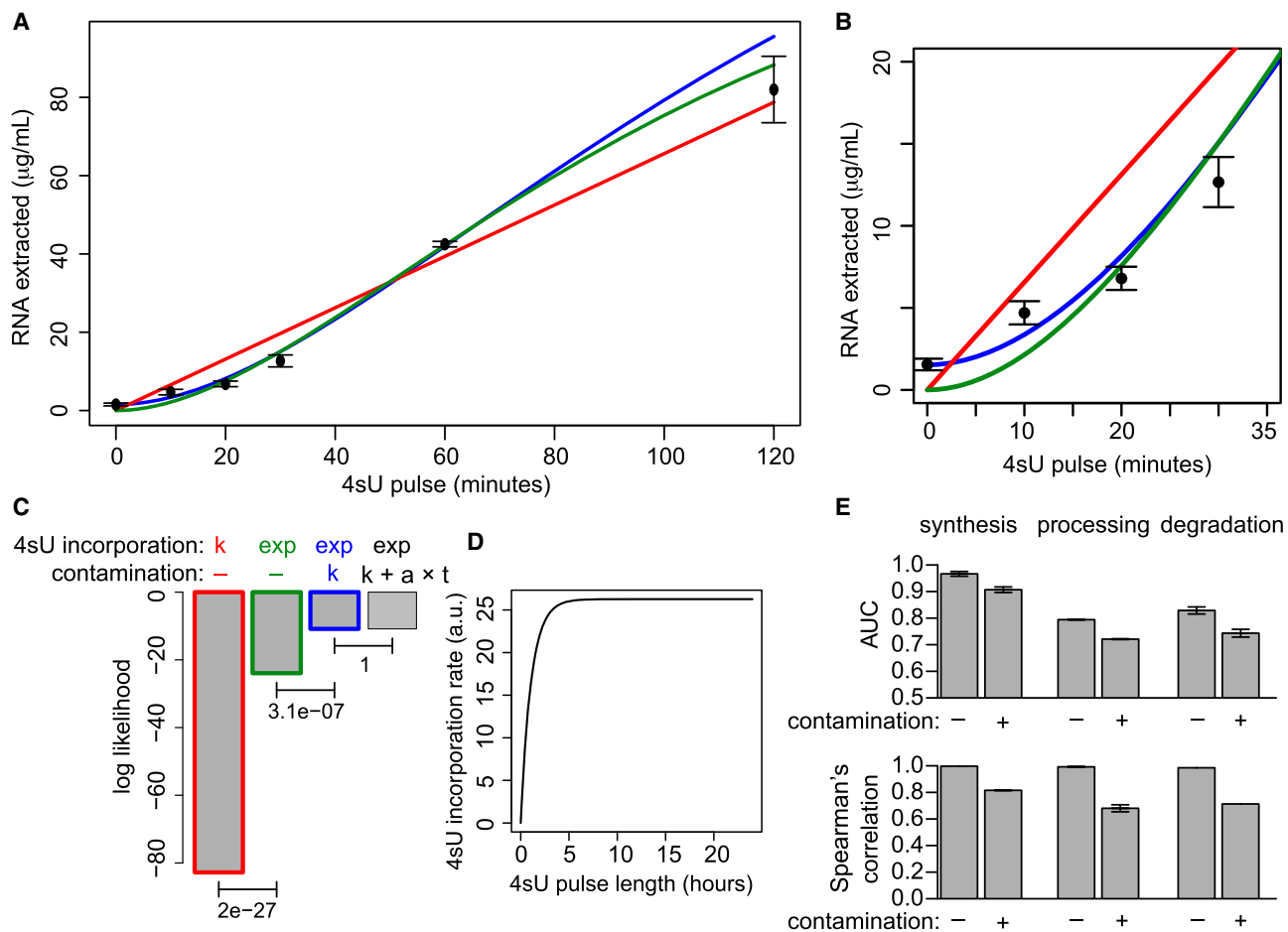
Figure 2. Contamination of 4sU labeled RNA with unlabeled RNA. (*A*) Yield of labeled RNA at 4sU pulses of different length in 3T9 mouse fibroblast cells. (*B*) A magnification of A. (*C*) Log likelihood score for the fit of four alternative models, considering constant (k) or exponential (exp) 4sU incorporation rates, combined with a contamination that is absent (−), constant (k), or linear to the 4sU pulse length (k + a × t). *P*-values of the indicated log likelihood-ratio tests are reported. (*D*) The estimated trend of 4sU incorporation based on the green model in panel *A*. (*E*) Changes in ROC areas under the curve (AUC) and in Spearman's correlation following the introduction of 30% contamination in simulated data.

(Rabani et al. 2014). Finally, a 30% contamination rate was also reported by Baptista and Dölken (2018). As the contamination rate is likely to depend on the cell type and on the specific protocol used, it should be reassessed at every experiment, thus further complicating the design of RNA metabolic labeling experiments.

Methods of RNA metabolic labeling that involve chemical derivatization do not rely on the purification of the labeled fraction and therefore do not count normalization and contamination among their downsides. However, although able to detect labeled transcripts with excellent specificity, these methods are hampered by low sensitivity and the need for a prolonged pulse time (typically >60 min). For example, it has been calculated that 2.4% T > C conversion rates obtained following 24-h-long pulses of 4sU in mouse embryonic stem cells (Herzog et al. 2017) permit to identify labeled RNAs at a sensitivity of 23% and 60% for read lengths of 50 bp or 150 bp, respectively (Neumann et al. 2019). Conversion rates decrease rapidly when the 4sU pulse length is reduced in order to increase temporal resolution, dropping to 0.5% for a 4-h-long pulse (Herzog et al. 2017). A reduced conversion rate is likely to worsen sensitivity. Finally, methods based on RNA metabolic labeling cannot be readily applied to model organisms, mammals (Matsushima et al. 2018) or plants (Sidaway-Lee et al. 2014), in vivo.

We recently developed INSPEcT (de Pretis et al. 2015), a Bioconductor package that, together with DRiLL (Rabani et al. 2014), combines analyses on total and nascent transcriptomes to allow, for the first time, quantification of RNA synthesis, processing, and degradation rates. Briefly, for each gene, INSPEcT compares eight different models, corresponding to all the possible combinations of each of the three kinetic rates in two alternative analytical forms (constant and impulse/sigmoid). Each model is plugged within a system of ordinary differential equations (Fig. 1B). The free parameters associated with the rates' functional forms are optimized to minimize the error when fitting premature and mature RNAs experimental data. Three key aspects of this method have been now updated. First, we have introduced a fully derivative approach able to speed up the execution by 20-fold (Supplemental Fig. S1). Second, model selection has been streamlined, as it now relies on fitting the model in which all rates are variable, avoiding the pair-wise comparison between all nested alternative models. Third, for each kinetic rate, confidence intervals are now determined in order to be exploited for model selection and to give critical information to the user. As before, INSPEcT is suitable for the analysis of both steady-state (Austenaa et al. 2015) and time course experiments (de Pretis

et al. 2017). Currently, only INSPEcT allows both the quantification of all kinetic rates at steady state and their temporal resolution in time course.

INSPEcT's ability to quantify the kinetic rates absolute values, and to identify genes with variable RNA dynamics, was benchmarked using simulated data that closely reproduced signal and noise of a real data set (Supplemental Figs. S2, S3; de Pretis et al. 2015). However, those data failed to include contamination of the labeled fraction with unlabeled pre-existing RNA, as an important source of bias in RNA metabolic labeling. To measure the importance of contamination, we generated simulated data with and without it. At a 30% contamination level, the correlation with expected rate values and the areas under the curve (AUC) from ROC analysis decreased by up to 30% and 12%, respectively (Fig. 2E), indicating that methods based on RNA metabolic labeling are severely affected by contamination of the labeled RNA fraction and prompting the search for alternative approaches.

Temporal quantification of RNA dynamics without RNA metabolic labeling

As illustrated in Figure 1, the modulation of one or more RNA kinetic rates leaves specific marks on the temporal profiles of premature and mature RNAs. Conversely, the temporal quantification of these RNA species should allow the deconvolution of the underlying RNA dynamics. Based on this rationale, we extended INSPEcT to include a novel computational approach able to quantify RNA dynamics using time course total RNA-seq data, without relying on any RNA metabolic labeling (Fig. 3A; Supplemental Methods). To keep it simple, INSPEcT+ and INSPEcT− will be used to refer to the application of the INSPEcT package to total and nascent or to just total RNA-seq data, respectively.

Briefly, INSPEcT− follows a three-step procedure in which the ODE system (Fig. 1B) is solved adopting various constraints on the functional shapes of the RNA kinetic rates (Fig. 3B). In the first step (priors estimation), processing (k_2) and degradation (k_3) rates are forced to be constant and optimized to reduce the chi-squared error on the mature RNA (M), assuming that premature RNA (P) behaves linearly between the experimental observations. The resulting k_1 priors, together with P and M, are used in the second step (first-guess estimation) to analytically solve the ODE system. This returns k_2 and k_3, which are now constant just between experimental time points (constant piecewise). In the last step, M, k_2, and k_3 are modeled through a combination of smooth functions (constant/sigmoid/impulsive), minimizing both error and complexity of the model according to the Akaike information criterion (AIC) framework. Finally, k_1 rates are updated accordingly, and confidence intervals are determined for all kinetic rates. The whole procedure takes ~10 sec per gene per core (Supplemental Fig. S1).
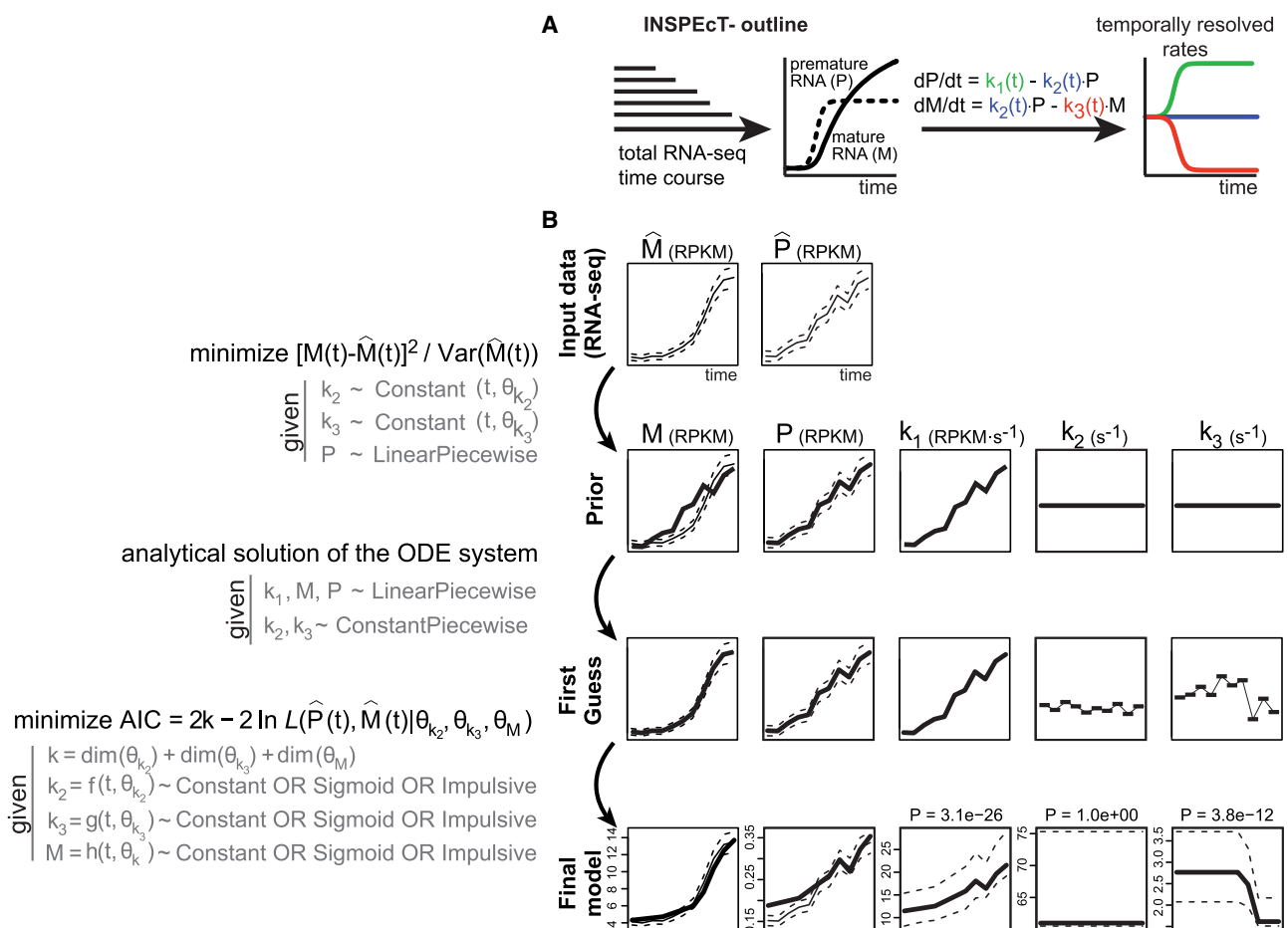

Figure 3. Temporal quantification of RNA dynamics without RNA metabolic labeling. (*A*) Outline of INSPEcT−. (*B*) INSPEcT− workflow; for details, see text. (AIC) Akaike information criterion; (*L*) likelihood function.

Validation of INSPEcT− RNA kinetic rates

We compared the RNA kinetic rates of 3T9 mouse fibroblasts cells that were quantified without (using INSPEcT−) or with metabolic labeling (using INSPEcT+) (de Pretis et al. 2017). Figure 4A exemplifies INSPEcT− output for the *H2bc6* gene in 3T9 cells after acute MYC activation, which closely matches the output of INSPEcT+ both in terms of fold change and time of response (for additional examples, see Supplemental Fig. S4). At the genome-wide level, the rates of synthesis, processing, and degradation quantified through INSPEcT− had a Spearman's correlation of 0.86, 0.61, and 0.69, with those quantified through INSPEcT+, respectively (upper-tail Spearman's Rho $P < 1 \times 10^{16}$) (Fig. 4B).

To further validate INSPEcT− kinetic rates without comparing them to the closely related INSPEcT+ approach, we focused on the rates of synthesis and degradation. INSPEcT− synthesis rates are expected to closely correspond to the quantification of nascent RNA. Indeed, when we performed a correlation analysis on 3T9 untreated cells, on those cells following 4 h of MYC activation, and on the $\log_2$ fold changes between those conditions, we obtained Spearman's correlations ranging between 0.87 and 0.90 (Supplemental Fig. S5). INSPEcT− degradation rates were compared with the rates determined by TimeLapse-seq, which relies on 4sU chemical derivatization (Schofield et al. 2018). Even though INSPEcT− degradation rates were determined on 3T9 mouse fibroblast cells, which are related but not identical to the mouse embryonic fibroblast cells used in the TimeLapse study (total RNA expression Spearman's correlation 0.67), the degradation rates determined with the two methods are in good agreement (0.50) (Supplemental Fig. S5). Rather, TimeLapse degradation rates have a lower correlation with INSPEcT+ degradation rates (0.47). Finally, we reanalyzed with INSPEcT− a time course of IL7-induced differentiation in WT and *Mettl3*-KO mouse T cells (Li et al. 2017; Furlan et al. 2019). METTL3 is the main m6A writer (Roundtree et al. 2017), and its KO reduced m6A bulk levels to 28% of WT levels. One of the key functions of m6A is to mediate the recruitment of marked RNAs to the degradation machinery. Therefore, a reduction in m6A is expected to lead to reduced degradation rates (Wang et al. 2014). Indeed, when we compared INSPEcT− degradation rates between WT and *Mettl3*-KO cells, they were reduced specifically for RNAs that were marked by m6A in the WT (Fig. 4C).

Altogether, the reanalysis of experimental data previously generated by us and others (genome-wide correlations in 3T9 cells, single gene examples, comparison with nascent RNA, TimeLapse, and the confirmation of reduced decay in the context of *Mettl3*-KO cells), validates INSPEcT− kinetic rates, indicating that their quantification is possible even in the absence of RNA metabolic labeling data.

To further and more comprehensively validate INSPEcT− ability to quantify rates changes, we used simulated data for 1000 genes (Supplemental Figs. S2, S3). For each gene, both nascent and total gene expression time course simulated data were included and analyzed using the INSPEcT+ (considering both types of data) and INSPEcT− (considering total RNA data only) approaches. Moreover, the simulated data included matching temporal profiles of RNA kinetic rates, which represented the ground truth for their pattern of modulation ("expected"). INSPEcT− kinetics rates changed over time similarly to INSPEcT+'s and closely recapitulated the expected response (Fig. 4D). The ability of our procedure of model selection to correctly classify variable rates was quantified using F1 scores, the harmonic mean of precision and recall (Fig. 4E). These results were in line or superior to those obtained with the INSPEcT+ approach, especially for degradation rates. In particular, both approaches have a specificity higher than 0.8, implying a low number of false positives (Supplemental Fig. S6).

A reduction in the number of time points only partially affects the quality of the classification, regardless of the availability of nascent RNA profiling (Fig. 4E). In particular, we found that the classification of genes modulated with sigmoidal functions are particularly resistant to a reduction in the number of time points. Rather, genes modulated with impulse functions most benefited from an increasing number of time points. This suggests that the cost of increasing the number of time points is not always justified by a corresponding growth in performance. Additional details on the impact of time series design on the quality of classification and practical hints for the design of these experiments are provided in the Supplemental Methods and Supplemental Figures S7–S9.

A possible problem with the INSPEcT− approach lies in the underdetermination issue affecting the equations presented in Figure 1B. Indeed, the modulation of mature RNA can be potentially explained by changes in either synthesis or degradation. Analogously, the modulation of premature RNA can be potentially explained by changes in either synthesis or processing rates. Although this ambiguity can be solved by profiling nascent RNA, which is a proxy for the rate of synthesis, it remains a potential confounding factor when only total RNA is considered. To quantify the importance of this issue, we repeated the ROC analyses by predicting the change in each rate based on the score of the other rates. Swapping the scores decreased INSPEcT− AUCs close to random levels (0.5) (Supplemental Fig. S10), indicating that the information gained for different rates is not interchangeable and showing that indetermination is not a major issue of our approach. This analysis also revealed that INSPEcT+ is more affected by the indetermination issue (Supplemental Fig. S10). In particular, when nascent RNA is profiled, changes in degradation rates can be attributed by error to synthesis and/or processing rates. This is likely because of contamination of labeled RNA with unlabeled transcripts. Indeed, when using simulated data not affected by contamination, the indetermination of INSPEcT+ is fully resolved (AUCs close to 0.5).

Altogether, these analyses indicated that the rates' absolute values and their changes over time could be estimated even in the absence of nascent RNA data.

Reanalysis of public data sets illustrates the additional information gained with INSPEcT−

We used INSPEcT− to reanalyze four publicly available RNA-seq time course data sets, corresponding to conditions with varying proportions of transcriptional and post-transcriptional regulation (Fig. 5A,B). The analysis of time course RNA-seq data sets is typically limited to the quantification of absolute and differential gene expression, as depicted in Figure 5C. Rather, INSPEcT− returned the quantification of the temporal changes in premature RNA and of the RNA kinetic rates (Fig. 5D), markedly extending what can be gained from the original data.

First, we focused on the temporal response to MYC acute activation in mouse fibroblasts, which we had recently characterized by profiling both total and nascent RNA (de Pretis et al. 2017). In that study, the integrative analysis of both data types with the INSPEcT+ approach revealed that MYC acts predominantly by modulating the rate of synthesis of its target genes, with an important, albeit less prevalent, impact on processing and degradation
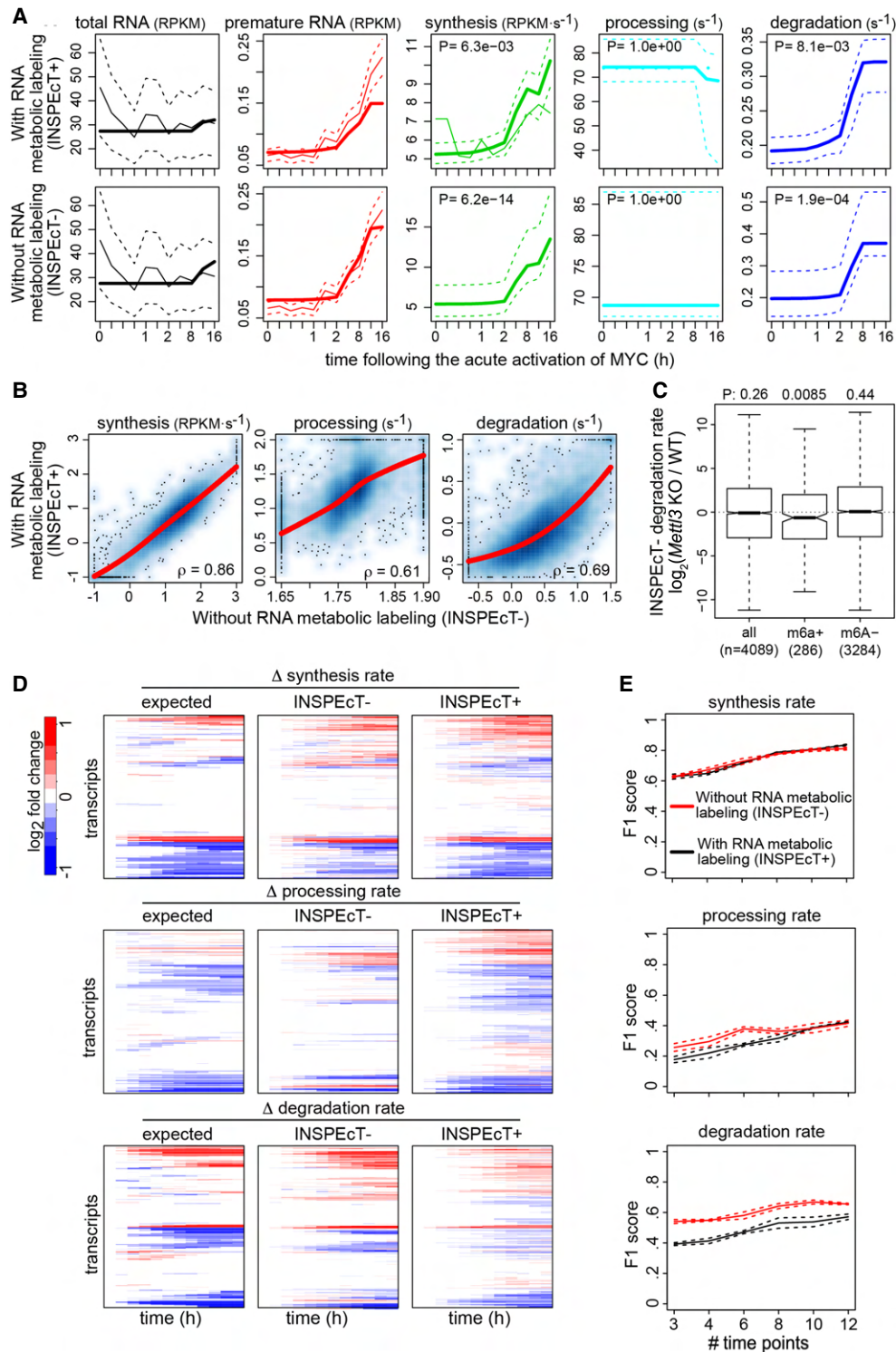
Figure 4. Validation of INSPEcT− kinetic rates. (*A*) *H2bc6* RNA dynamics quantified in 3T9 mouse fibroblast cells following the acute activation of MYC with (INSPEcT+) and without (INSPEcT−) RNA metabolic labeling. Solid bold lines indicate the model fit; thin solid and dashed lines indicate mean and standard deviation of experimental data for total and premature RNA; dashed lines indicate 95% confidence intervals for the kinetic rates models. (*B*) Scatter plots of RNA kinetic rates quantified in untreated 3T9 cells using INSPEcT+ and INSPEcT−. Regression curves and Spearman's correlation coefficients are indicated within each panel. (*C*) Boxplot of the changes in degradation rates during the differentiation of T cells quantified with INSPEcT−. Rates changes are displayed for m6A+, m6A−, or all RNAs in untreated cells. One-tailed Wilcoxon test *P*-values are displayed on the *top*. (*D*) Temporal changes of the RNA kinetic rates for simulated genes, relative to the initial time point (*left*), compared with those quantified through INSPEcT+ (*middle*) and INSPEcT− (*right*). (*E*) For each kinetic rate, quantified with or without metabolic labeling data, F1 scores are reported that measure the quality of the classification (*P*-value cutoff 0.05), considering both precision and recall. Score means and standard deviations are reported based on three simulated data sets obtained at increasing number of time points.
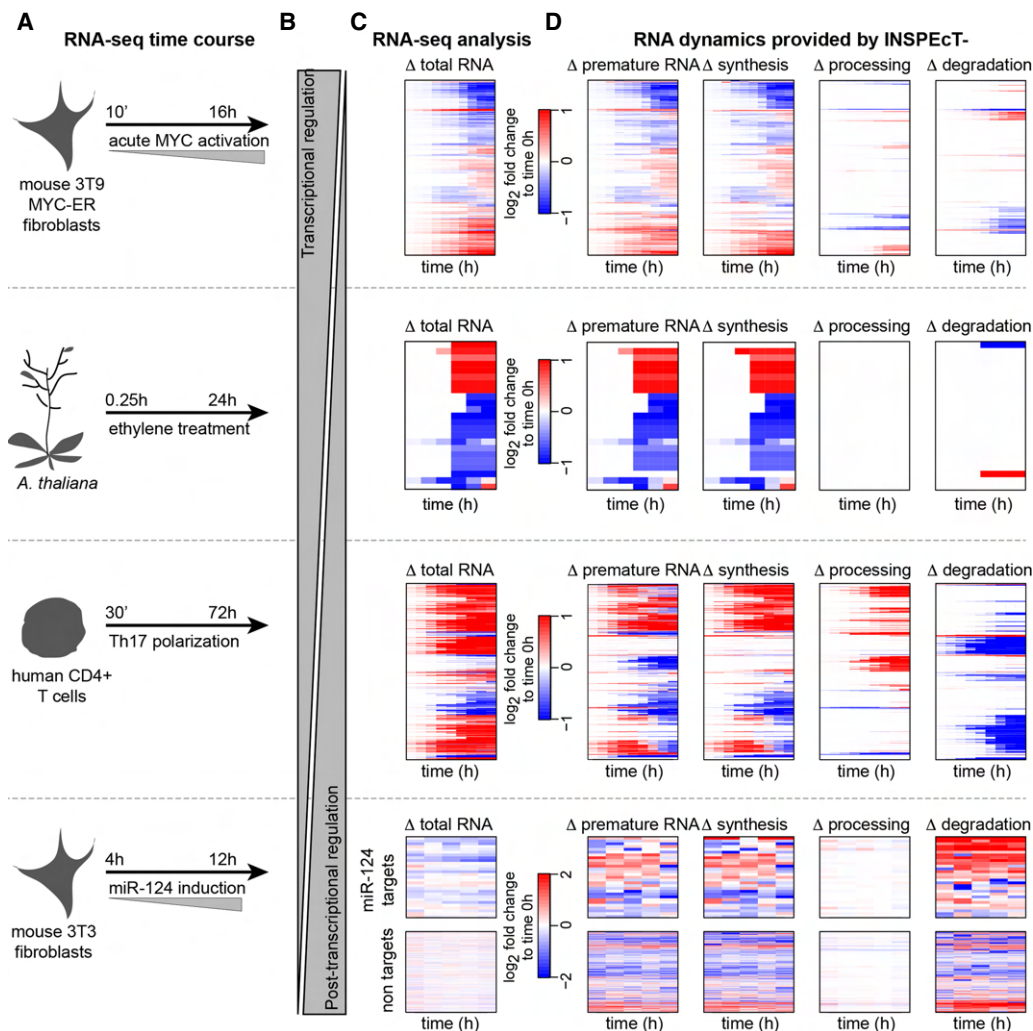
Figure 5. Characterization of time course RNA dynamics: reanalysis of four published data sets with INSPEcT−. (*A*) Experimental design of the considered RNA-seq time courses. (*B*) Expected balance between transcriptional and post-transcriptional responses in the different experiments. (*C*) The typical output of differential RNA-seq analyses: heatmap of differentially expressed genes. (*D*) The additional information gained by reanalyzing those data with INSPEcT−, which includes the gene-level modulation of premature RNAs, as well as the temporal changes of the kinetic rates of synthesis, processing, and degradation. For the miR-124 data set, reported rates are first-guess estimates, owing to lack of replicates in the time course.

involving around one-third of targets (Sabò et al. 2014; de Pretis et al. 2017). In agreement with those results, reanalysis with INSPEcT− (which neglects any available nascent RNA data) confirmed that 85% of MYC targets were impacted at the level of their synthesis rate, whereas 32% of them were affected in either processing or degradation (Fig. 5D).

Second, we quantified for the first time all kinetic rates in plants, focusing on the temporal response to ethylene in *Arabidopsis thaliana* (Chang et al. 2013). Ethylene causes growth inhibition, which is initially independent and then dependent on the EIN3 transcriptional regulator. After 4 h, EIN3 binding reaches its maximum, leading to a strong transcriptional response (Chang et al. 2013). Indeed, our analysis confirmed that ethylene response is primarily controlled at the transcriptional level (Fig. 5C).

Third, we reanalyzed the total RNA-seq data set on the temporal polarization of CD4[+] T cells with polarizing cytokines from Tuomela et al. (2016). As expected, in comparison to the one elicited by a master transcription factor of the likes of MYC, the response was more mixed and less dependent on the modulation

of synthesis rates: 72% of genes were modulated at the level of their processing and/or degradation rates (Fig. 5D).

Finally, we reanalyzed the total RNA-seq temporal response to the activation of miRNA-124 (Eichhorn et al. 2014). We expected to see a strong and specific post-transcriptional regulation of the miRNA target transcripts, and, indeed, these were seen to be primarily controlled at the level of their stability, leading to a reduction in total RNA, whereas nontarget transcripts remained mostly unaffected (Fig. 5C).

Altogether, these analyses illustrate how the quantification of RNA dynamics from total RNA-seq data sets can unveil the underlying mechanisms controlling premature and mature RNA abundances and their variations.

Temporal quantification of RNA dynamics without assumptions on the functional form

In this study, changes in premature and mature RNA, as well as in the kinetic rates, are modeled by fitting sigmoid or impulse

functions. Sigmoids are the most elementary nonlinear functions for modeling a smooth transition between two steady states. Impulse models, which combine an early response followed by an additional transition to a steady state, were previously proposed and successfully used for the modeling of transcriptional responses (Chechik et al. 2008; Chechik and Koller 2009). Moreover, they were already adopted in the context of RNA dynamics modeling (Rabani et al. 2011, 2014). However, despite their flexibility and broad applicability, these functional forms place a constraint on the modeling, which may poorly adapt to other temporal response patterns, such as oscillatory or more complex responses.

To deal with these cases without introducing additional or overly complicated functions, we implemented a modeling approach based on linear piece-wise functions, available for both INSPEcT+ and INSPEcT− (see Supplemental Methods). Briefly, confidence intervals are determined for first-guess kinetic rates (Fig. 3B), thus revealing the degree of dissimilarity from a constant model without assuming alternative functional forms. To test this approach, we built data sets including simulated genes modulated by a circadian oscillation of synthesis rates or by a circadian oscillation of both synthesis and degradation rates opportunely out of phase (Fig. 6A). Models returned by both INSPEcT+ and INSPEcT− and obtained by fitting sigmoid or impulse functions had a poor goodness of fit for the oscillating genes (Fig. 6B). Rather, we found that both approaches can successfully model these circadian oscillatory patterns when agnostic of a priori knowledge of the functional forms (Fig. 6C).

RNA-dynamics from steady-state total RNA-seq data

At steady state and in the absence of nascent RNA profiling, no information on the rate of synthesis is available. However, the ratio of premature to mature RNA abundance is equal to the ratio of degradation to processing rate (k_3/k_2) (Fig. 1B). Although this ratio does not allow the deconvolution of the individual contributions of the two rates, its change over different conditions indicates alterations in post-transcriptional regulation. INSPEcT− uses the ratio of premature to mature RNA species to provide an excellent estimate of the k_3/k_2 ratio (Fig. 7A). The modulation of the ratio across conditions, such as time points, is also accurate (Fig. 7B). This suggests that steady-state post-transcriptional regulation can be studied even in the absence of RNA metabolic labeling.

Based on this rationale, we used INSPEcT− to characterize the landscape of human post-transcriptional regulation with an unprecedented breadth, covering 35,000 genes and more than 600 RNA-seq samples. By leveraging natural language processing approaches that we had recently implemented in the Onassis Bioconductor package (Galeota et al. 2020), each data set was assigned to a specific tissue type and disease condition, ultimately covering 26 tissues and 24 diseases. We focused on RNA-seq data sets depleted of ribosomal RNA species and therefore enriched of both premature and mature RNAs. Moreover, we relied on RNA-seq coverage data that had been homogeneously reanalyzed across data sets as a part of the recount2 project (Collado-Torres et al. 2017), thus minimizing potential batch effects owing to different analysis pipelines and normalization methods. We found that the amount of premature RNA (P) increases with the abundance of mature RNA (M) following a power law that depends on the gene type (protein coding, pseudo, or long noncoding genes) (Fig. 7C). Noncoding transcripts have a higher proportion of premature RNA compared with other gene types. One possible reason for this is that RNA processing rates are particularly low for noncoding

genes, which was indeed recently reported using metabolic labeling (Mukherjee et al. 2017). Significant deviations from these trends, for each gene class, point to post-transcriptionally regulated genes (Fig. 7C; Supplemental Methods).

To validate the null model implemented in INSPEcT, which relies on the global power law relationship between the expression of premature and mature RNAs, we first verified that it does not depend on the level of gene expression. Indeed, the proportion of regulated genes is similar at different levels of expression (Supplemental Fig. S11). Moreover, we reasoned that the genes deviating from this model, if they were post-transcriptionally regulated, should be enriched in miRNA targets. Indeed, their enrichment is maximum in correspondence of the power law slope identified with the INSPEcT null model (Supplemental Fig. S12).

Each gene, within each sample, was classified as post-transcriptionally regulated (red in Fig. 7D), nondifferential (white), or not expressed (blue). Unsupervised clustering of the heatmap columns resulted in the spontaneous grouping of samples from similar tissues and disease conditions (Fig. 7D; Supplemental Fig. S13), suggesting that post-transcriptional regulation is coordinated across similar biological conditions. The observed sample clustering did not simply arise from gene expression patterns of tissue-specific genes, because it was 30% different from the clustering obtained based on mature RNA (Supplemental Fig. S14). This analysis allowed us to rank samples and genes according to their propensity to be regulated at the post-transcriptional level. On one hand, this revealed that post-transcriptional regulation is particularly common in specific conditions (Fig. 7D,E). On the other hand, this indicated that the three gene classes were markedly different in terms of post-transcriptional regulation, with protein coding and pseudo genes being regulated more frequently than noncoding ones (Fig. 7D). Finally, we analyzed the function of the 1000 genes with the lowest frequency of post-transcriptional regulation and found them to be associated with basic cellular processes such as protein folding, organelle organization, and metabolic processes ($P < 1 \times 10^{30}$). On the contrary, the 1000 genes with the highest frequency were enriched in miRNA targets and were found to be related to more specific cellular processes, including various diseases, B cell activation, autoimmune response, differentiation, and morphology ($P < 1 \times 10^2$).

We analyzed more closely the functionality of the genes undergoing post-transcriptional regulation under specific conditions. Genes altered in T cell samples were associated with the regulation of T cell number and proliferation and with immunodeficiency. Genes altered in heart samples were associated with cardiac hypertrophy, abnormal contractility, and cardiomyopathy. Indeed, a subset of these samples could be associated with the cardiomyopathy disease (Fig. 7D). Focusing on the RNAs regulated in the brain, the corresponding genes were associated with several diseases, including glioma, autism, and neoplasm of the nervous system, and with biological processes such as hormone secretion and synaptic transmission. Compared with genes expressed in the brain, the 3′ and 5′ UTR regions in the subset of the regulated transcripts are longer, have a lower percentage of CGs, and have lower free energy, (Fig. 7F), indicating a higher likelihood of harboring regulatory motifs. In particular, their 3′ UTRs are enriched in motifs containing the ACA sequence (Fig. 7G). In mammals, ACA is where the majority of N6-methyladenosines (m6As) occur, m6A being the most abundant RNA modification and an important determinant of post-transcriptional regulation (Linder et al. 2015; Roundtree et al. 2017). We used AURA (Dassi et al. 2012, 2014) to search for motifs of RNA-binding proteins within these
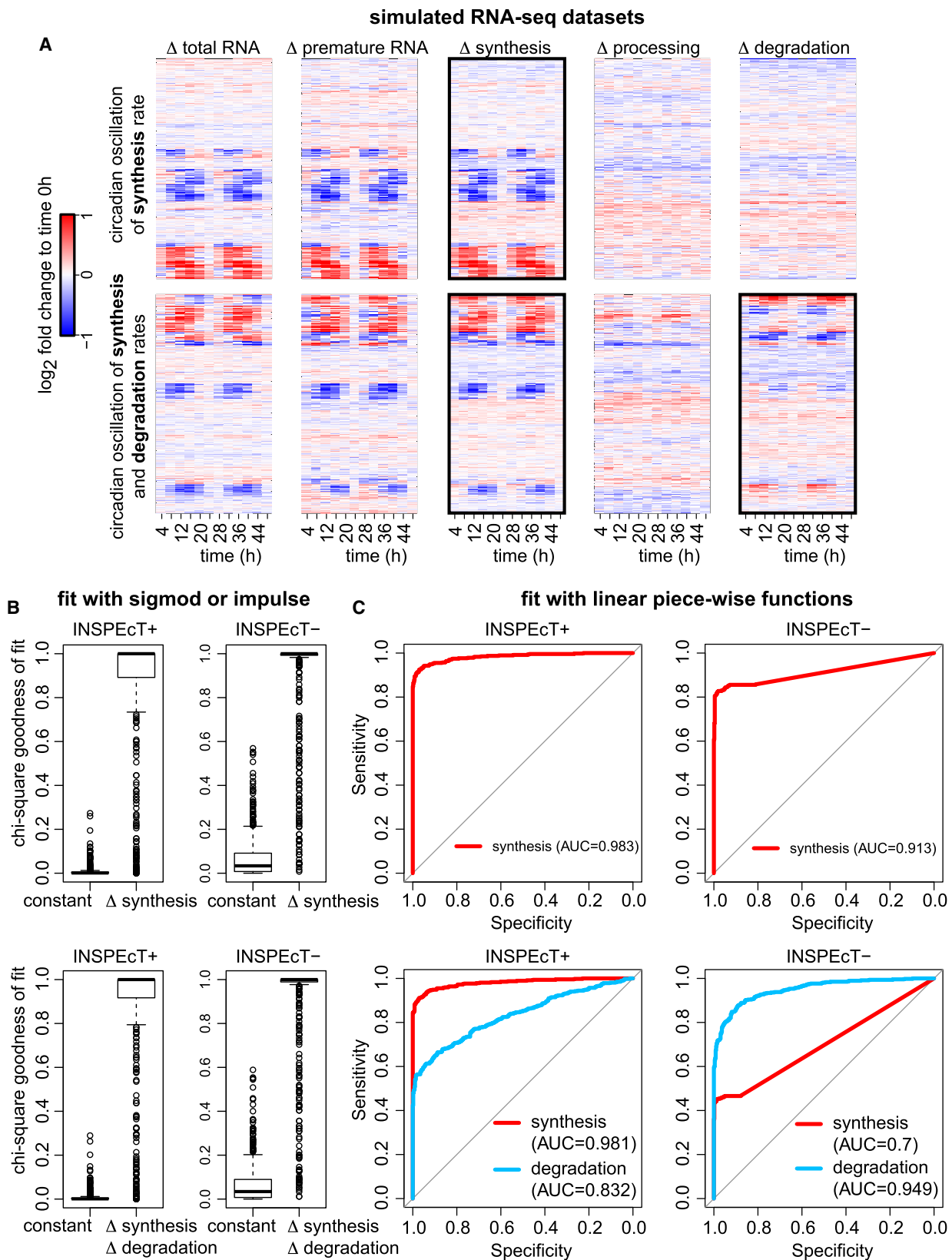
Figure 6. Modeling RNA dynamics without assumptions on the functional form. (*A*) Simulated data composed by 500 constant genes and 500 genes subject to the circadian oscillation of synthesis (*top*) or synthesis followed by degradation rates (*bottom*). (*B*) Chi-square goodness of fit of sigmoid or impulse models on the data sets in *A* using INSPEcT+ or INSPEcT−. (*C*) ROC analysis of the classification of synthesis (*top*) or synthesis and degradation rates (*bottom*) using INSPEcT+ or INSPEcT− with linear piecewise functions.
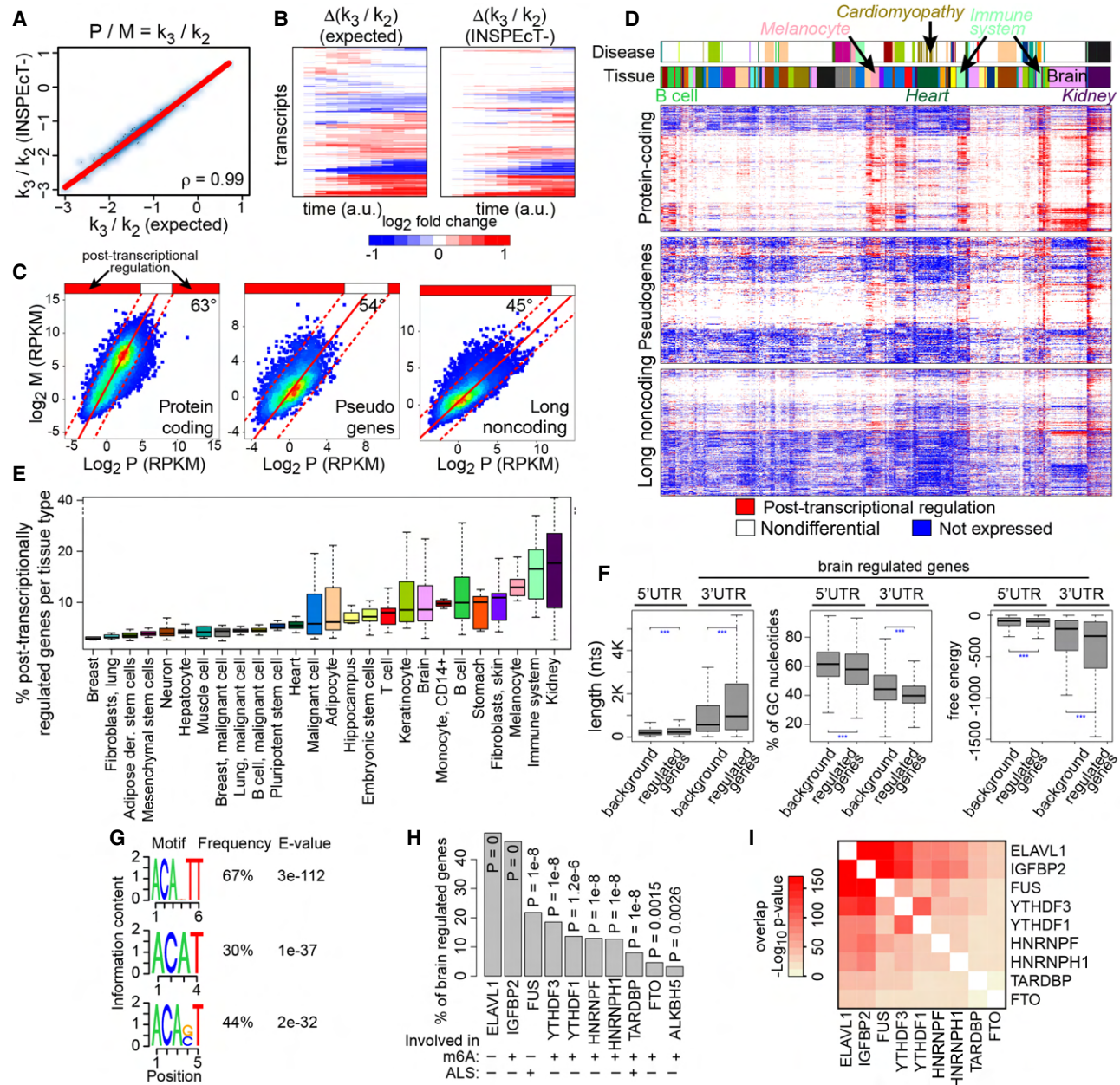
Figure 7. Characterization of steady-state RNA dynamics: reanalysis of 620 RNA-seq data sets with INSPEcT−. At steady state, the ratio between premature (P) and mature (M) RNA corresponds to the ratio between degradation (k_3) and processing (k_2) rates. Absolute values (*A*) and the temporal variation (*B*) of k_3/k_2 ratios determined by INSPEcT− were compared to the ground truth ("expected"). (*C*) Median abundances of premature and mature RNAs per gene across 620 RNA-seq data sets for the indicated gene classes. Density scatter plots were fitted with a linear model, whose slope is reported. (*D*) Heatmaps displaying the classification in terms of post-transcriptional regulation for each gene (row) in each sample (column). k_3/k_2 ratios were quantified for each gene in each sample and compared with the global trend depicted in *C*. Each gene is either not expressed (blue), is not differential (white; ratio *between* the dashed lines in *C*), or is differentially post-transcriptional regulated (red; ratio *above* the dashed lines in *C*). *Above* the heatmaps, two color bars indicate the tissue type and disease conditions of each sample. (*E*) Boxplot of the percentage of genes that are post-transcriptionally regulated for the samples associated to each cell type, color matched with *D*. (*F*) Distributions of length, %GC, and free energy for 3′ and 5′ UTRs of genes post-transcriptionally regulated in the brain compared with all genes expressed in the brain (background). (*G*) Sequence logo of the selected RNA-binding protein motifs in the 3′ UTR regions of brain regulated genes. (*H*) Frequency and *P*-values of enrichment for selected motifs of RNA-binding proteins found in UTR regions of brain regulated genes. (*I*) Hypergeometric *P*-value for the overlap between the genes associated to the factors in *H*.

UTR regions (Fig. 7H). Among the enriched motifs we found those for ELAVL1, also known as HuR, an important regulator of transcripts stability (Mukherjee et al. 2011). In addition, we identified motifs for several m6A readers and erasers (Edupuganti et al. 2017). Finally, we identified the binding proteins FUS and TARDBP, im-

portant factors in amyotrophic lateral sclerosis (ALS) (Paez-Colasante et al. 2015). The genes associated with the motifs of these factors have a marked overlap (Fig. 7I). For example, >92% of the genes containing the FUS motif in their 3′ UTR and >78% of those containing the TARDBP motif also contain the ELAVL1

and IGFBP2 motifs ($P < 2 \times 10^{41}$). Collectively, these data confirm that m6A-directed post-transcriptional regulation is pervasive in the brain (Yoon et al. 2018) and potentially relevant for ALS. Finally, this analysis provided sets of candidate regulated genes, as well as RNA-binding proteins that could be responsible for their atypical dynamics of expression.

Altogether, these results illustrated the type and range of information that INSPEcT− is able to provide from the study of RNA dynamics when individual conditions are compared in the absence of nascent RNA data.

The impact of different RNA-seq protocols

The measurement of the abundance of both premature and mature RNA is pivotal in all the approaches that aim to quantify RNA dynamics, with or without nascent RNA. Premature RNA is typically quantified by intronic RNA-seq signals, whereas the abundance of mature transcripts is obtained by subtracting intronic from exonic signals. Numerous studies support the concept that intronic RNA-seq reads are a robust proxy for the abundance of premature RNA and the rate of RNA production (Ameur et al. 2011; Rabani et al. 2011, 2014; Zeisel et al. 2011). In particular, in a recent report (Gaidatzis et al. 2015), a comprehensive analysis was conducted that showed the high correspondence between intronic RNA-seq signals and both nascent and chromatin-associated RNA signals. To further confirm the notion that intronic and exonic signals are closely related to premature and mature RNA, respectively, we took advantage of a study in which the nuclear and cytoplasmic RNA fractions were distinctly profiled. As expected, intronic reads are markedly enriched in the nuclear fraction and depleted in the cytoplasmic fraction, and our quantifications of premature and mature expression have Spearman's correlations of 0.75 and 0.88 with the abundance of nuclear and cytoplasmic RNA, respectively (Supplemental Fig. S15).

Throughout this study, in order to maximize intronic signal, we conservatively decided to take into consideration only total RNA-seq experiments in which RNA molecules had not been poly(A)-selected. However, we found that standard coverage (20 million aligned reads) RNA-seq libraries prepared with various protocols, including poly(A) selection, are also suitable for these analyses (Supplemental Fig. S16A; Adiconis et al. 2013). Indeed, Spearman's correlations between Ribo-Zero and poly(A) selection protocols are in the order of 0.85–0.9 for both premature RNAs and their ratios to mature RNAs.

To test INSPEcT− on a poly(A)-selected RNA-seq data set, we reanalyzed the temporal response to the induction of RAF. Additional data from the same study revealed that the gene expression response was primarily controlled at the transcriptional level (Uhlitz et al. 2017). Despite the low coverage in the time course of the total RNA-seq samples, INSPEcT− confirmed a modulation in the synthesis rate of 90% of the genes with altered kinetic rates (Supplemental Fig. S16B).

These data and results indicate that there is enough intronic signal available in samples subjected to poly(A) selection, despite the depletion of premature RNA species, and that the quantification of premature and mature RNA species is robust to the choice of the RNA-seq protocol, thus broadening the scope of our approaches.

Comparison with existing methods

The analysis of premature and mature RNA abundances, without the quantification of nascent RNA, has been already used to study RNA dynamics (Table 1). Few methods were developed that only allow characterizing steady-state RNA dynamics. SnapShot-Seq enables the inference of splicing kinetics by the differential coverage that introns have in their 5′ and 3′ ends, whereas this is not amenable for gene-level analyses (Gray et al. 2014). Alternatively, by assuming invariant splicing kinetics SnapShot-Seq allows the quantification of absolute RNA synthesis (∼ intron RNA-seq signal) and decay rates (∼ exon/intron signals) for individual genes. With similar assumptions, two additional tools were developed: EISA (Gaidatzis et al. 2015), which infers changes in synthesis and degradation, and REMBRANDTS (Alkallas et al. 2018), which only focuses on the latter and includes a term to manage the coupling between transcription and processing. The only method able to deal with time courses is the one described by Zeisel et al. (2011), which models RNA synthesis and degradation dynamics from a

Table 1. Comparison of INSPEcT− with available tools for the analysis of RNA dynamics based on total RNA-seq data

			INSPEcT−	EISA	SnapShot-Seq	REMBRANDTS	Zeisel
Steady state	Absolute rates	Synthesis	—	—	•	—	—
		Processing	•[a]	—	•[b]	—	—
		Degradation	•[a]	—	•	—	—
	Differential rates	Synthesis	—	•	—	—	—
		Processing	•[a]	—	—	—	—
		Degradation	•[a]	•	—	•	—
Time course	Absolute rates	Synthesis	•	—	—	—	•[c]
		Processing	•	—	—	—	—
		Degradation	•	—	—	—	•[c]
	Differential rates	Synthesis	•	—	—	—	—
		Processing	•	—	—	—	—
		Degradation	•	—	—	—	—
Rates coupling[d]			•	—	—	•	—
Unbounded response complexity[e]			•	—	—	—	—
Simulated data			•	—	—	—	—
Software			•	—	—	•	—

[a]Aggregated datum.
[b]Cannot be resolved at the gene level.
[c]Constrained to the experimental measure of RNA processing rate.
[d]Managing the coupling between transcriptional and post-transcriptional responses.
[e]Assumption-free on the maximum complexity of the responses (e.g., compatible with circadian oscillations).

time course of premature and mature RNA abundances. This method requires knowing a priori the rate of RNA processing; it assumes that this rate is constant throughout the time course and is not suitable for steady-state analyses.

To quantitatively compare the results obtained with INSPEcT– against other tools, we took advantage of a recently published study in which changes in RNA degradation were independently quantified through the block-of-transcription approach (Slobodin et al. 2020). As part of that study, the investigators showed that the drug camptothecin (CPT) slows down RNA polymerase II elongation and reduces RNA degradation, mediated by changes in m6A RNA modifications. When those data were analyzed with EISA and REMBRANDTS, they both returned changes in RNA stability that were opposite to those experimentally measured by block of transcription (Spearman's correlation of −0.35 and −0.31, respectively) (Supplemental Fig. S17), suggesting that RNA degradation was actually increased for most transcripts instead of being reduced. This is likely because of impact of CPT on the RNA processing machinery, which invalidates the basic assumptions of both EISA and REMBRANDTS. In fact, EISA assumes that the processing rates are invariant between conditions, whereas REMBRANDTS assumes that changes in the processing rates are opposed to changes in the synthesis rate. Instead, the analysis of those data with INSPEcT– confirmed that substantial changes in post-transcriptional regulation occurred (Supplemental Fig. S17), whereas our method does not distinguish the contribution of processing or degradation rates. In addition, the INSPEcT– enrichment in miRNA targets is higher than the one obtained with REMBRANDTS regulated genes (Wilcoxon test P-value 4.7×10^{42}) (Supplemental Fig. S12). Overall, at steady state, INSPEcT– relies on the modulation of the ratio of premature to mature RNA abundance as previously proposed (Gaidatzis et al. 2015), implements a novel null model to find significant deviations, and avoids assumptions regarding the step of premature RNA processing (La Manno et al. 2018). The analyses presented in Supplemental Figures S12 and S17 suggest that the procedure implemented in INSPEcT– for the analysis of steady-state conditions safeguards from the confounding effect of a modulation of both processing and degradation machineries.

Altogether, INSPEcT– is compatible with the broadest range of experimental designs, can generate and take advantage of simulated data, is available as a well-documented software, and offers a graphical user interface (Table 1; de Pretis et al. 2020).

Discussion

The deconvolution of RNA dynamics from transcriptional genomics data is an emerging field of research, which the development of RNA metabolic labeling has fuelled by enabling the analysis of nascent transcription (Dolken et al. 2008; Rabani et al. 2011; Baptista and Dölken 2018). We recently developed INSPEcT, a Bioconductor package that, through mathematical modeling of nascent and total RNA-seq data sets, allows the quantification of the kinetic rates governing the RNA life cycle (de Pretis et al. 2015). We extensively used this tool for the analysis of the RNA dynamics controlling several classes of coding and noncoding transcripts (Austenaa et al. 2015; Marzi et al. 2016; de Pretis et al. 2017). Aware of the challenges the integrative analysis of nascent and total RNA-seq data poses, we have now expanded the package with INSPEcT– to include the possibility to use total RNA-seq data sets only, without requiring any information on nascent transcripts.

Based on experimental data generated by us and others, the RNA kinetic rates calculated by INSPEcT–, using time course total RNA-seq experiments, were validated through comparison with those obtained by using RNA metabolic labeling. Moreover, degradation rates quantified through INSPEcT– were validated in *Mettl3*-KO cells, where, as expected, they are reduced following the depletion of m6A RNA modifications. Finally, the ability of INSPEcT– to quantify changes in all the kinetic rates was benchmarked on simulated data sets. In particular, INSPEcT– quantifications of transcripts half-lives were found to be improved compared with INSPEcT+, which is affected by the contamination of unlabeled RNA. By reanalyzing various time course data sets of total RNA-seq, we illustrated INSPEcT–'s ability to unravel underlying RNA dynamics and hence provide a deeper understanding of the resulting gene expression programs. INSPEcT– prevents all the additional experimental work required in nascent RNA profiling and safeguards from a number of pitfalls afflicting RNA metabolic labeling experiments, primarily the difficulty in working with limited RNA amounts and/or tight temporal resolutions, the necessity to normalize the quantification of pre-existing transcripts to that of nascent transcripts, and the contamination of the latter with the former. Although at steady state these downsides could be accepted in exchange for the ability to deconvolute all RNA kinetic rates, in time course conditions they might not be justified when considering INSPEcT– straightforwardness.

Finally, we also showed that INSPEcT– could unveil RNA-seq dynamics under steady-state conditions by providing the first comprehensive analysis of post-transcriptional regulation using hundreds of publicly available data sets, covering a multitude of tissues and disease conditions. The analysis revealed a signature of brain genes, some of which are involved in ALS, which is potentially post-transcriptionally regulated by m6A RNA modifications.

In conclusion, the characterization of RNA dynamics can uncover the mechanistic details underlying complex transcriptional responses. INSPEcT– allows, for the first time, to quantify the magnitude and the modulation of all RNA kinetic rates without requiring RNA metabolic labeling data. Hence, it provides a new perspective on what knowledge can be gained from total RNA-seq data sets, including those previously published, which can now be used not only for measuring abundance and variation in expression but also for unveiling the contribution of the different phases in the RNA metabolism. We expect that our approach will be useful for the analysis of RNA dynamics in the context of single cells, as well as direct RNA sequencing data (e.g., Nanopore-based), with (Erhard et al. 2019; Furlan et al. 2020; Maier et al. 2020) or without RNA metabolic labeling. Finally, INSPEcT– is ideal for the identification or prioritization of conditions that are likely to be of high interest to the study of RNA modifications and of their pivotal role in controlling RNA metabolism (Roundtree et al. 2017; Furlan et al. 2019). Altogether, INSPEcT is a unifying computational tool able to unfold these layers of regulation in most experimental scenarios, independently from the availability of information on nascent transcription, and is suitable for both steady-state and time course profiling of total RNA-seq.

Methods

Expression data quantification

Premature, mature, and total RNA expression levels were quantified through a dedicated routine of the INSPEcT package. Premature and total RNA were estimated as length and library

size normalized read counts that overlap gene introns and exons, respectively. Mature RNA was estimated as the difference between total and premature RNA. If a gene had multiple isoforms, we collapsed the exons of its transcripts and defined introns as the gaps between adjacent collapsed exons.

The RNA-seq data sets reanalyzed in this study can be found under the following NCBI Gene Expression Omnibus (GEO; https://www.ncbi.nlm.nih.gov/geo) or Sequence Read Archive (SRA; https://www.ncbi.nlm.nih.gov/sra) accession numbers: MYC activation (Fig. 5), (GEO) GSE98420; *A. thaliana*, (SRA) SRP017925; T cell differentiation, (GEO) GSE52260; and miRNA induction, (GEO) GSE60426. RNA-seq data for Figure 7 were retrieved using the recount R/Bioconductor package available at https://bioconductor.org/packages/recount/.

For additional details, see Supplemental Methods Section 1.

Mathematical modeling of the RNA life cycle

We modeled the RNA life cycle through a set of two ordinary differential equations, which describe the modulation of premature (P) and mature (M) RNA, respectively, as functions of synthesis (k_1), processing (k_2), and degradation (k_3) rates. This is the core of the inference procedures implemented in INSPEcT. For additional details, see Supplemental Methods Section 2.

Temporal inference of RNA kinetic rates

The time course inference procedure starts with the fit of k_1 as a piecewise linear function and of k_2, k_3 as piecewise constant functions. This step overfits the expression levels, but it also provides a fast solution to check the quality of input data and a first quantification of the kinetic rates that is used to initialize the parameters of further modeling steps. For additional details, see Supplemental Methods 3.1.

The second stage of the inference procedure aims at controlling the noise associated with the experimental data and to statistically assess the rates responsible for premature and mature RNA modulation. Three alternative routines are available to perform this task (for additional details, see Supplemental Methods 3.2):

1. The integrative functional approach, which restricts the shape of kinetic rates to constant, sigmoid, or impulse functions and exploits the parameterization to numerically solve the ODE system. The comparison between inferred and experimental expression levels guides models optimization (standard chi-square minimization) and selection (AIC minimization by default).
2. The derivative functional approach, which is similar to the integrative one, but the parameterization regards one RNA species, either mature or total RNA, and two kinetic rates. The missing quantities needed to estimate the cost function for model optimization and selection are expressed as functions of the parameterized quantities and their time derivatives. This allows bypassing the numerical solution of the ODE system, reducing the computational cost, and is the default method in the INSPEcT package.
3. The nonfunctional approach, which is able to detect gene responses of any shape as it relies on the piecewise parameterization but is also more affected by noise than the first two alternatives.

Validation of the temporal inference

Simulated data were exploited to characterize the performance of INSPEcT+ and INSPEcT− on the classification of constant and variable rates. They were generated through a revised approach included in the INSPEcT package, which now takes into account nascent RNA contamination (for additional details, see Supplemental Methods Section 4.2). We characterized the contamination process with a dedicated experiment based on nascent RNA profiling at different labeling times. Experimental and computational details are available in Supplemental Methods Section 4.1.

The performance in the classification of the kinetic rates as constant or variable was evaluated through specificity, sensitivity, area under the ROC curve, and/or F1 score, comparing observed to expected (simulated) classification results. For additional details, see Supplemental Methods Section 4.3.

RNA kinetic rates inference at steady state

At steady state, the only quantity that can be inferred regarding the RNA life cycle kinetics is the ratio between post-transcriptional rates (k_3 over k_2), which is equal to P over M. A modulation of this ratio between conditions indicates an uneven regulation of processing and degradation rates.

INSPEcT− identifies significant post-transcriptional regulations as data points that deviate from a linear model fitted in the $\log_2 P$, $\log_2 M$ space. This approach allows filtering out trivial regulations owing to the coupling of synthesis, processing, and degradation machineries. For additional details, see Supplemental Methods Section 5.

We applied the steady-state INSPEcT− approach on a large data set (669 samples and 35125 genes) of non-poly(A)-selected RNA-seq experiments retrieved querying the SRA database. The corresponding data were retrieved using the R/Bioconductor package recount, and the corresponding metadata were annotated through the Onassis R/Bioconductor package. For additional details, see Supplemental Methods Section 6.1.

We determined distinct null models in the $\log_2 P$, $\log_2 M$ space for protein coding genes, pseudogenes, and noncoding genes according to the GENCODE annotation (for additional details, see Supplemental Methods Section 6.2). We identified sets of genes atypically regulated in samples sharing the same tissue and/or disease annotations, which we investigated through functional enrichment analysis (for details, see Supplemental Methods Sections 6.3 and 6.4). Finally, we characterized the features of UTR regions of protein coding genes post-transcriptionally regulated in brain by comparing their length, GC content, and free energy to the background. We applied the regulatory enrichment tool of the AURA2 database to search for known motifs of RNA-binding proteins (see Supplemental Table S1). For additional details, see Supplemental Methods Section 6.5. Gene and samples identifications of the recount data set, as well as the Gene Ontology enrichment analysis results, are available in Supplemental Table S2.

Software availability

The INSPEcT R/Bioconductor package, including both the INSPEcT+ and INSPEcT− approaches, is available at https://bioconductor.org/packages/INSPEcT/, together with software documentation and instructions for its installation.

Data access

R scripts that allow reproducing all main and supplemental figures and other key results included in this study are available as Supplemental Code.

Competing interest statement

The authors declare no competing interests.

Acknowledgments

References

Adiconis X, Borges-Rivera D, Satija R, DeLuca DS, Busby MA, Berlin AM, Sivachenko A, Thompson DA, Wysoker A, Fennell T, et al. 2013. Comparative analysis of RNA sequencing methods for degraded or low-input samples. *Nat Methods* **10:** 623–629. doi:10.1038/nmeth.2483

Alkallas R, Fish L, Goodarzi H, Najafabadi HS. 2018. Inference of RNA decay rate from transcriptional profiling highlights the regulatory programs of Alzheimer's disease. *Nat Commun* **8:** 909. doi:10.1038/s41467-017-00867-z

Ameur A, Zaghlool A, Halvardson J, Wetterbom A, Gyllensten U, Cavelier L, Feuk L. 2011. Total RNA sequencing reveals nascent transcription and widespread co-transcriptional splicing in the human brain. *Nat Struct Mol Biol* **18:** 1435–1440. doi:10.1038/nsmb.2143

Austenaa LMI, Barozzi I, Simonatto M, Masella S, Chiara Della G, Ghisletti S, Curina A, de Wit E, Bouwman BAM, de Pretis S, et al. 2015. Transcription of mammalian *cis*-regulatory elements is restrained by actively enforced early termination. *Mol Cell* **60:** 460–474. doi:10.1016/j.molcel.2015.09.018

Baptista MAP, Dölken L. 2018. RNA dynamics revealed by metabolic RNA labeling and biochemical nucleoside conversions. *Nat Methods* **15:** 171–172. doi:10.1038/nmeth.4608

Bergen V, Lange M, Peidli S, Wolf FA, Theis FJ. 2020. Generalizing RNA velocity to transient cell states through dynamical modeling. *Nat Biotechnol* doi:10.1038/s41587-020-0591-3

Chang KN, Zhong S, Weirauch MT, Hon G, Pelizzola M, Li H, Huang SSC, Schmitz RJ, Urich MA, Kuo D, et al. 2013. Temporal transcriptional response to ethylene gas drives growth hormone cross-regulation in arabidopsis. *eLife* **2:** e00675. doi:10.7554/eLife.00675

Chechik G, Koller D. 2009. Timing of gene expression responses to environmental changes. *J Comput Biol* **16:** 279–290. doi:10.1089/cmb.2008.13TT

Chechik G, Oh E, Rando O, Weissman J, Regev A, Koller D. 2008. Activity motifs reveal principles of timing in transcriptional control of the yeast metabolic network. *Nat Biotechnol* **26:** 1251–1259. doi:10.1038/nbt.1499

Collado-Torres L, Nellore A, Kammers K, Ellis SE, Taub MA, Hansen KD, Jaffe AE, Langmead B, Leek JT. 2017. Reproducible RNA-seq analysis using recount2. *Nat Biotechnol* **35:** 319–321. doi:10.1038/nbt.3838

Dassi E, Malossini A, Re A, Mazza T, Tebaldi T, Caputi L, Quattrone A. 2012. AURA: atlas of UTR regulatory activity. *Bioinformatics* **28:** 142–144. doi:10.1093/bioinformatics/btr608

Dassi E, Re A, Leo S, Tebaldi T, Pasini L, Peroni D, Quattrone A. 2014. AURA 2: empowering discovery of post-transcriptional networks. *Translation (Austin)* **2:** e27738. doi:10.4161/trla.27738

de Pretis S, Kress T, Morelli MJ, Melloni GEM, Riva L, Amati B, Pelizzola M. 2015. INSPEct: a computational tool to infer mRNA synthesis, processing and degradation dynamics from RNA- and 4sU-seq time course experiments. *Bioinformatics* **31:** 2829–2835. doi:10.1093/bioinformatics/btv288

de Pretis S, Kress TR, Morelli MJ, Sabò A, Locarno C, Verrecchia A, Doni M, Campaner S, Amati B, Pelizzola M. 2017. Integrative analysis of RNA polymerase II and transcriptional dynamics upon MYC activation. *Genome Res* **27:** 1658–1664. doi:10.1101/gr.226035.117

de Pretis S, Furlan M, Pelizzola M. 2020. INSPEcT-GUI reveals the impact of the kinetic rates of RNA synthesis, processing, and degradation, on premature and mature RNA species. *Front Genet* **11:** 230. doi:10.3389/fgene.2020.00759

Dolken L, Ruzsics Z, Radle 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. doi:10.1261/rna.1136108

Edupuganti RR, Geiger S, Lindeboom RGH, Shi H, Hsu PJ, Lu Z, Wang S-Y, Baltissen MPA, Jansen PWTC, Rossa M, et al. 2017. N^6-methyladenosine (m^{6}A) recruits and repels proteins to regulate mRNA homeostasis. *Nat Struct Mol Biol* **24:** 870–878. doi:10.1038/nsmb.3462

Eichhorn SW, Guo H, McGeary SE, Rodriguez-Mias RA, Shin C, Baek D, Hsu S-H, Ghoshal K, Villén J, Bartel DP. 2014. mRNA destabilization is the dominant effect of mammalian microRNAs by the time substantial repression ensues. *Mol Cell* **56:** 104–115. doi:10.1016/j.molcel.2014.08.028

Erhard F, Baptista MAP, Krammer T, Hennig T, Lange M, Arampatzi P, Jürges CS, Theis FJ, Saliba A-E, Dölken L. 2019. scSLAM-seq reveals core features of transcription dynamics in single cells. *Nature* **571:** 419–423. doi:10.1038/s41586-019-1369-y

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

Fuchs G, Voichek Y, Benjamin S, Gilad S, Amit I, Oren M. 2014. 4sUDRB-seq: measuring genomewide transcriptional elongation rates and initiation frequencies within cells. *Genome Biol* **15:** R69–D875. doi:10.1186/gb-2014-15-5-r69

Fuchs G, Voichek Y, Rabani M, Benjamin S, Gilad S, Amit I, Oren M. 2015. Simultaneous measurement of genome-wide transcription elongation speeds and rates of RNA polymerase II transition into active elongation with 4sUDRB-seq. *Nat Protoc* **10:** 605–618. doi:10.1038/nprot.2015.035

Furlan M, Galeota E, de Pretis S, Caselle M, Pelizzola M. 2019. m6A-Dependent RNA dynamics in T cell differentiation. *Genes (Basel)* **10:** 28. doi:10.3390/genes10010028

Furlan M, Tanaka I, Leonardi T, de Pretis S, Pelizzola M. 2020. Direct RNA sequencing for the study of synthesis, processing, and degradation of modified transcripts. *Front Genet* **11:** 394. doi:10.3389/fgene.2020.00394

Gaidatzis D, Burger L, Florescu M, Stadler MB. 2015. Analysis of intronic and exonic reads in RNA-seq data characterizes transcriptional and post-transcriptional regulation. *Nat Biotechnol* **33:** 722–729. doi:10.1038/nbt.3269

Galeota E, Kishore K, Pelizzola M. 2020. Ontology-driven integrative analysis of omics data through Onassis. *Sci Rep* **10:** 703. doi:10.1038/s41598-020-57716-1

Gray JM, Harmin DA, Boswell SA, Cloonan N, Mullen TE, Ling JJ, Miller N, Kuersten S, Ma Y-C, McCarroll SA, et al. 2014. SnapShot-Seq: a method for extracting genome-wide, *in vivo* mRNA dynamics from a single total RNA sample. *PLoS One* **9:** e89673. doi:10.1371/journal.pone.0089673

Herzog VA, Reichholf B, Neumann T, Rescheneder P, Bhat P, Burkard TR, Wlotzka W, Haeseler von A, Zuber J, Ameres SL. 2017. Thiol-linked alkylation of RNA to assess expression dynamics. *Nat Methods* **14:** 1198–1204. doi:10.1038/nmeth.4435

Jürges C, Dölken L, Erhard F. 2018. Dissecting newly transcribed and old RNA using GRAND-SLAM. *Bioinformatics* **34:** i218–i226. doi:10.1093/bioinformatics/bty256

La Manno G, Soldatov R, Zeisel A, Braun E, Hochgerner H, Petukhov V, Lidschreiber K, Kastriti ME, Lönnerberg PLX, Furlan A, et al. 2018. RNA velocity of single cells. *Nature* **560:** 494–498. doi:10.1038/s41586-018-0414-6

Li H-B, Tong J, Zhu S, Batista PJ, Duffy EE, Zhao J, Bailis W, Cao G, Kroehling L, Chen Y, et al. 2017. M6a mRNA methylation controls T cell homeostasis by targeting the IL-7/STAT5/SOCS pathways. *Nature* **548:** 338–342. doi:10.1038/nature23450

Linder B, Grozhik AV, Olarerin-George AO, Meydan C, Mason CE, Jaffrey SR. 2015. Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. *Nat Methods* **12:** 767–772. doi:10.1038/nmeth.3453

Lugowski A, Nicholson B, Rissland OS. 2018. DRUID: a pipeline for transcriptome-wide measurements of mRNA stability. *RNA* **24:** 623–632. doi:10.1261/rna.062877.117

Maier KC, Gressel S, Cramer P, Schwalb B. 2020. Native molecule sequencing by nano-ID reveals synthesis and stability of RNA isoforms. *Genome Res* **30:** 1332–1344. doi:10.1101/gr.257857.119

Marzi MJ, Ghini F, Cerruti B, de Pretis S, Bonetti P, Giacomelli C, Gorski MM, Kress T, Pelizzola M, Muller H, et al. 2016. Degradation dynamics of microRNAs revealed by a novel pulse-chase approach. *Genome Res* **26:** 554–565. doi:10.1101/gr.198788.115

Matsushima W, Herzog VA, Neumann T, Gapp K, Zuber J, Ameres SL, Miska EA. 2018. SLAM-ITseq: sequencing cell type-specific transcriptomes without cell sorting. *Development* **145:** dev164640. doi:10.1242/dev.164640

Michel M, Demel C, Zacher B, Schwalb B, Krebs S, Blum H, Gagneur J, Cramer P. 2017. TT-seq captures enhancer landscapes immediately after T-cell stimulation. *Mol Syst Biol* **13:** 920–913. doi:10.15252/msb.20167507

Miller C, Schwalb BOR, Maier K, Schulz D, Dümcke SDU, Zacher B, Mayer A, Sydow J, Marcinowski L, Martin DE, 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. doi:10.1038/nmeth.1226

Mukherjee N, Corcoran DL, Nusbaum JD, Reid DW, Georgiev S, Hafner M, Ascano M, Tuschl T, Ohler U, Keene JD. 2011. Integrative regulatory mapping indicates that the RNA-binding protein HuR couples pre-mRNA processing and mRNA stability. *Mol Cell* **43:** 327–339. doi:10.1016/j.molcel.2011.06.007

Mukherjee N, Calviello L, Hirsekorn A, de Pretis S, Pelizzola M, Ohler U. 2017. Integrative classification of human coding and noncoding genes through RNA metabolism profiles. *Nat Struct Mol Biol* **24:** 86–96. doi:10.1038/nsmb.3325

Neumann T, Herzog VA, Muhar M, Haeseler von A, Zuber J, Ameres SL, Rescheneder P. 2019. Quantification of experimentally induced nucleotide conversions in high-throughput sequencing datasets. *BMC Bioinformatics* **20:** 258. doi:10.1186/s12859-019-2849-7

Orphanides G, Reinberg D. 2002. A unified theory of gene expression. *Cell* **108:** 439–451. doi:10.1016/S0092-8674(02)00655-4

Paez-Colasante X, Figueroa-Romero C, Sakowski SA, Goutman SA, Feldman EL. 2015. Amyotrophic lateral sclerosis: mechanisms and therapeutics in the epigenomic era. *Nat Rev Neurol* **11:** 266–279. doi:10.1038/nrneurol.2015.57

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. doi:10.1038/nbt.1861

Rabani M, Raychowdhury R, Jovanovic M, Rooney M, Stumpo DJ, Pauli A, Hacohen N, Schier AF, Blackshear PJ, Friedman N, et al. 2014. High-resolution sequencing and modeling identifies distinct dynamic RNA regulatory strategies. *Cell* **159:** 1698–1710. doi:10.1016/j.cell.2014.11.015

Roundtree IA, Evans ME, Pan T, He C. 2017. Dynamic RNA modifications in gene expression regulation. *Cell* **169:** 1187–1200. doi:10.1016/j.cell.2017.05.045

Sabò A, Kress TR, Pelizzola M, de Pretis S, Gorski MM, Tesi A, Morelli MJ, Bora P, Doni M, Verrecchia A, et al. 2014. Selective transcriptional regulation by Myc in cellular growth control and lymphomagenesis. *Nature* **511:** 488–492. doi:10.1038/nature13537

Schofield JA, Duffy EE, Kiefer L, Sullivan MC, Simon MD. 2018. TimeLapse-seq: adding a temporal dimension to RNA sequencing through nucleoside recoding. *Nat Methods* **15:** 221–225. doi:10.1038/nmeth.4582

Sidaway-Lee K, Costa MJ, Rand DA, Finkenstädt B, Penfield S. 2014. Direct measurement of transcription rates reveals multiple mechanisms for configuration of the arabidopsis ambient temperature response. *Genome Biol* **15:** R45. doi:10.1186/gb-2014-15-3-r45

Slobodin B, Bahat A, Sehrawat U, Becker-Herman S, Zuckerman B, Weiss AN, Han R, Elkon R, Agami R, Ulitsky I, et al. 2020. Transcription dynamics regulate poly(A) tails and expression of the RNA degradation machinery to balance mRNA levels. *Mol Cell* **78:** 434–444.e5. doi:10.1016/j.molcel.2020.03.022

Sun M, Schwalb B, Schulz D, Pirkl N, Etzold S, Larivière L, Maier KC, Seizl M, Tresch A, Cramer P. 2012. Comparative dynamic transcriptome analysis (cDTA) reveals mutual feedback between mRNA synthesis and degradation. *Genome Res* **22:** 1350–1359. doi:10.1101/gr.130161.111

Tesi A, de Pretis S, Furlan M, Filipuzzi M, Morelli MJ, Andronache A, Doni M, Verrecchia A, Pelizzola M, Amati B, et al. 2019. An early Myc-dependent transcriptional program orchestrates cell growth during B-cell activation. *EMBO Rep* **20:** e47987. doi:10.15252/embr.201947987

Tuomela S, Rautio S, Ahlfors H, Öling V, Salo V, Ullah U, Chen Z, Hämälistö S, Tripathi SK, Äijö T, et al. 2016. Comparative analysis of human and mouse transcriptomes of Th17 cell priming. *Oncotarget* **7:** 13416–13428. doi:10.18632/oncotarget.7963

Uhlitz F, Sieber A, Wyler E, Fritsche-Guenther R, Meisig J, Landthaler M, Klinger B, Blüthgen N. 2017. An immediate-late gene expression module decodes ERK signal duration. *Mol Syst Biol* **13:** 928. doi:10.15252/msb.20177554

Uvarovskii A, Dieterich C. 2017. pulseR: versatile computational analysis of RNA turnover from metabolic labeling experiments. *Bioinformatics* **33:** 3305–3307. doi:10.1093/bioinformatics/btx368

Uvarovskii A, Naarmann-de Vries IS, Dieterich C. 2019. On the optimal design of metabolic RNA labeling experiments. *PLoS Comput Biol* **15:** e1007252. doi:10.1371/journal.pcbi.1007252

Wada T, Becskei A. 2017. Impact of methods on the measurement of mRNA turnover. *Int J Mol Sci* **18:** 2723–2714. doi:10.3390/ijms18122723

Wang X, Lu Z, Gomez A, Hon GC, Yue Y, Han D, Fu Y, Parisien M, Dai Q, Jia G, et al. 2014. N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature* **505:** 117–120. doi:10.1038/nature12730

Wissink EM, Vihervaara A, Tippens ND, Lis JT. 2019. Nascent RNA analyses: tracking transcription and its regulation. *Nat Rev Genet* **20:** 705–723. doi:10.1038/s41576-019-0159-6

Yoon K-J, Ming G-L, Song H. 2018. Epitranscriptomes in the adult mammalian brain: dynamic changes regulate behavior. *Neuron* **99:** 243–245. doi:10.1016/j.neuron.2018.07.019

Zeisel A, Köstler WJ, Molotski N, Tsai JM, Krauthgamer R, Jacob-Hirsch J, Rechavi G, Soen Y, Jung S, Yarden Y, et al. 2011. Coupled pre-mRNA and mRNA dynamics unveil operational strategies underlying transcriptional responses to stimuli. *Mol Syst Biol* **7:** 529. doi:10.1038/msb.2011.62

Genome-wide dynamics of RNA synthesis, processing, and degradation without RNA metabolic labeling

Mattia Furlan, Eugenia Galeota, Nunzio Del Gaudio, et al.

Supplemental Material	http://genome.cshlp.org/content/suppl/2020/09/25/gr.260984.120.DC1
P<P	Published online September 25, 2020 in advance of the print journal.
Open Access	Freely available online through the *Genome Research* Open Access option.
Creative Commons License	This article, published in *Genome Research*, is available under a Creative Commons License (Attribution-NonCommercial 4.0 International), as described at http://creativecommons.org/licenses/by-nc/4.0/.
Email Alerting Service	Receive free email alerts when new articles cite this article - sign up in the box at the top right corner of the article or **click here.**

Advance online articles have been peer reviewed and accepted for publication but have not yet appeared in the paper journal (edited, typeset versions may be posted when available prior to final publication). Advance online articles are citable and establish publication priority; they are indexed by PubMed from initial publication. Citations to Advance online articles must include the digital object identifier (DOIs) and date of initial publication.