SI Appendix for: "Genome-wide modelling of transcription kinetics reveals patterns of RNA production delays"

Antti Honkela [*], Jaakko Peltonen [† ‡], Hande Topa [†], Iryna Charapitsa [§], Filomena Matarese [¶], Korbinian Grote [‖], Hendrik G. Stunnenberg [¶], George Reid [§], Neil D. Lawrence [**] and Magnus Rattray [††]

[*]Helsinki Institute for Information Technology HIIT, Department of Computer Science, University of Helsinki, Helsinki, Finland,[†]Helsinki Institute for Information Technology HIIT, Department of Computer Science, Aalto University, Espoo, Finland,[‡]School of Information Sciences, University of Tampere, Tampere, Finland,[§]Institute for Molecular Biology, Mainz, Germany,[¶]Radboud University, Department of Molecular Biology, Faculty of Sciences and Faculty of Medicine, Nijmegen, The Netherlands,[‖]Genomatix Software GmbH, Muenchen, Germany,[**]Department of Computer Science, University of Sheffield, Sheffield, UK, and [††]Faculty of Life Sciences, University of Manchester, Manchester, UK

This supplementary file provides detailed information on materials and methods including data acquisition and preprocessing, differential equation based modelling, Gaussian process based inference, and filtering of results.

In the following we provide details of data acquisition, processing of RNA-seq data and filtering of active genes, processing of pol-II ChIP-seq data, differential equation modelling of the connection between pol-II and mRNA, Gaussian process based inference of underlying time series, and summarisation and filtering of results. We then provide an explanation of how synthetic data were used to study accuracy of parameter estimation for mRNA half life, a measure of mRNA decay in the differential equation model between pol-II and mRNA. Lastly, we provide additional figures about tail probabilities of delays for alternative result filtering choices, an additional figure about long posterior mean delays with and without annotated exon skipping, and differences in pre-mRNA accumulation in short and long delay genes.

Data acquisition

MCF-7 breast cancer cells were treated with estradiol (E2). The cells were put in estradiol free media for three days. This is defined media devoid of phenol red (which is estrogenic) containing 2% charcoal stripped foetal calf serum. The charcoal absorbs estradiol but not other essential serum components, such as growth factors. This resulted in basal levels of transcription from E2 dependent genes. The cells were then incubated with E2 containing media, which resulted in the stimulation of estrogen responsive genes. Measurements were taken at logarithmically spaced time points 0, 5, 10, 20, ..., 1280 minutes after E2 stimulation.

At each time point, the pol-II occupancy was measured genome-wide by ChIP-seq. Raw reads were mapped onto the human genome reference sequence (NCBI_build37) using the Genomatix Mining Station (software version 3.5.2). The mapping software is an index-based mapper using a shortest unique subword index generated from the reference to identify possible read positions. A subsequent alignment step is then used to get the highest-scoring match(es) according to the parameters used. We used a minimum alignment quality threshold of 92% for mapping, reads were not trimmed. On average 84% percent of reads could be mapped uniquely.

At each time point, the pre-mRNA and mature mRNA abundances were measured for each human gene by RNA-seq. Total RNA was isolated and subjected to rRNA depletion with the Ribo-Zero Magnetic Gold Kit and processed further for strand-specific RNA-seq. The RNA-seq reads were mapped using Bowtie to a transcriptome constructed from Ensembl version 68 annotation allowing at most 3 mismatches and ignoring reads with more than 100 alignments. The transcriptome was formed by combining the cDNA and ncRNA transcriptomes with pre-mRNA sequences containing the full genomic sequence from the beginning of the first annotated exon to the end of the last annotated exon. On average 84.7% of the RNA-seq reads were mapped.

All the ChIP-seq and RNA-seq data are available from the NCBI Gene Expression Omnibus under accession number GSE62789.

RNA-seq data processing

RNA-seq data were analysed at each time point separately using BitSeq (1). The reads were first mapped to human reference transcriptome (Ensembl v68) using Bowtie version 0.12.7 (2). In order to separate pre-mRNA activity as well, we augmented the reference transcriptome with pre-mRNA transcripts for each gene that consisted of the genomic sequence from the beginning of the first exon to the end of the last exon of the gene.

BitSeq uses a probabilistic model to probabilistically assign multimapping reads to transcript isoforms (1), in our case also including the pre-mRNA transcripts. We obtained gene expression estimates by adding the corresponding mRNA transcript expression levels. In addition to the mean expression levels, BitSeq provides variances of the transcript isoform expression levels. We further used the biological variance estimation procedure from BitSeq differential expression analysis on the estimated gene expression levels by treating the first three time points (0, 5, 10 min) as biological replicates. Genes with similar mean expression levels (log-RPKM) were grouped together such that each group contained 500 genes except for the last group with 571 genes with the highest expression. Then, the biological variances were estimated for each group of genes by using the Metropolis–Hastings algorithm used in BitSeq stage 2 (1). Biological variances for the single measurements were determined according to the gene expression levels at each time point, where each gene was considered to belong to the closest gene group according to its expression level. The observation noise variance for each observation was defined as the sum of the technical (BitSeq stage 1) and biological (BitSeq stage 2) variances, and transformed from log-expression to raw expression using

$$\sigma^2_{\text{raw}} = \sigma^2_{\log} \exp(\mu_{\log})^2. \qquad \textbf{[S1]}$$

Different time points of the RNA-seq time series were normalised using the method of (3) as implemented in the edgeR R/Bioconductor package (4).

Statistics of RNA-seq mapping and distribution of reads for pre-mRNA and mRNA transcripts are presented in Tables S2 and S3 as well as Fig. S9.

Filtering of active genes

We removed genes with no clear time-dependent activity by fitting time-dependent Gaussian process models to the activity curves and only keeping genes with Bayes factor at least 3 in favour of the time-dependent model compared to a null model with no time dependence.

We also removed genes that had no pol-II observations at 2 or more time points. This left 4420 genes for which we fitted the models.

Pol II ChIP-seq data processing

The ChIP-seq data were processed into time series by summarising the pol-II occupancy over time for each human gene (Ensembl version 68 annotation was used for gene positions), by a series of steps as follows.

1. Each gene was divided into 200 bp bins and levels of pol-II activity were computed at each time point as the total weighted count of reads overlapping each bin, where each read was weighted by how many basepairs in the read overlap the bin, as follows. Only uniquely mapped reads were used. For any read that at least partially overlaps a bin, the number of basepairs overlapping the bin was added into the activity level of the bin. For any read spanning multiple bins, the number of basepairs overlapping each bin were added into activity of that bin. The Genomatix mapping software provides alignment scores (values between 0 and 1; with our threshold only between 0.92 and 1) for mapping reads to the genome; for any read having alignment score less than 1, the number of overlapping basepairs added to each bin was multiplied by the alignment score.
2. A noise removal was then done: a noise level was computed as the average activity level in 74 manually selected regions from Chromosome 1 that were visually determined to be inactive over the measurement time points, as follows. The regions were divided into 200 bp bins, and total weighted counts of reads overlapping each bin (each read weighted by the number of basepairs overlapping the bin) were computed in the same way as for the genes in the previous step. For each time point, the noise level was computed as the average activity level over all bins from all 74 regions. The computed noise level was subtracted from the mean of each bin in each gene, thresholding the result at zero. A list of the empty regions used is included as a Supplementary Dataset S1.
3. As the number of ChIP-seq reads collected overall for pol-II varies between time points, a robust normalisation was done. After the previous noise removal step, for each gene g at each time t we compute the mean of the remaining activity (activity level after noise removal) over bins of the gene, denoted as r_{gt}. The activity levels are weighted counts of basepairs from reads overlapping the gene; we select genes having sufficient activity, that is, at least $5 \cdot 200$ overlapping basepairs from reads over each 200 bp bin of the gene, on average over the bins. For each gene g let $T_g = \{t' \in \{5, 10, \ldots, 1280 \text{ min}\} | r_{gt'} > 5 \cdot 200\}$ denote those time points (except the first time point) where the gene has sufficient activity. For each time point we compute a normalisation factor of (3)

$$C_t = \text{Median}_g \left\{ \frac{r_{gt}}{\text{GeomMean}_{t'}\{r_{gt'}\}} \right\} .$$

   where $\text{Median}_g\{\cdot\}$ denotes median over genes and $\text{GeomMean}_{t'}\{r_{gt'}\} = (\prod_{t' \in T_g} r_{gt'})^{1/|T_g|}$ is the geometric mean over the time points having sufficient activity for gene g. The median is computed for time points after the first time point; for the first time point $t = 0$ min we set $C_t = 1$. The factor C_t normalises all the gene activity levels (weighted read counts) at a time point downwards if genes at that time point have unusually many reads, exceeding their (geometric) mean activity level, and normalises upwards if gene activity levels fall under their mean activity level.
4. Lastly, time series summaries were computed for pol-II at each gene. For each gene at each time point t, the mean activity level (weighted read-count) of pol-II over bins in the 20% section of the gene nearest to transcription end was computed, normalised by C_t. This measured pol-II level represents transcriptional activity that had successfully passed through the gene to the transcription end site; it is expected to correspond better with mRNA production rate than pol-II activity at the transcription start of the gene, since pol-II near the transcription start site can be in the active or inactive state and after activation may require a significant time for transcription to complete.
5. For a small number of genes where the active mRNA transcripts covered only part of the gene, we considered the area from the first active exon to last active exon, and summarised the gene using the 20% section nearest to the end of the area. Active transcripts were defined here as transcripts with a mean of more than 1.1 assigned counts in the BitSeq posterior expression estimates. BitSeq uses a prior that assigns 1 "pseudo-count" per transcript, so the active transcripts were only required to have minimal posterior expression that was distinguishable from the prior. A list of active transcripts is included as Supplementary Dataset S2.
6. Lastly, for mathematical convenience, for each pol-II time series we subtracted from all time points the minimum value over the time points.

Differential equation based modelling

We model the role of pol-II as a catalyst of the transcription of DNA into mRNA as a differential equation for each gene; the differential equation relates the pol-II time series $p(t)$ of the gene and the corresponding mRNA time series $m(t)$.

Let us assume the momentary pol-II activity directly represents the momentary rate of transcription, potentially with a delay, and that the mRNA decays at a constant rate. We model this as a linear differential equation

$$\frac{\mathrm{d}m(t)}{\mathrm{d}t} = \beta_0 + \beta p(t - \Delta) - \alpha m(t) \qquad \textbf{[S2]}$$

where Δ is a delay parameter between the pol-II activity and the momentary transcription rate, β_0 is a parameter representing the *baseline transcription rate from unobservable pol-II background* (baseline production level of mRNA), β is a parameter representing *transcriptional efficiency*, that is, sensitivity of the transcription rate to activity of pol-II, and α is a constant mRNA *decay* rate parameter that is related to mRNA half-life $t_{1/2}$ through $\alpha = \ln(2)/t_{1/2}$.

The momentary mRNA level $m(t)$ can be solved from the differential equation to yield the following solution:

$$m(t) = m_0 e^{\alpha(t_0 - t)} + \frac{\beta_0}{\alpha}\left(1 - e^{-\alpha t}\right) + \beta e^{-\alpha t} \int_{u=0 \text{ min}}^{t} e^{\alpha u} p(u - \Delta) \mathrm{d}u \qquad \textbf{[S3]}$$

where pol-II activity is assumed to start at $t = 0$ min ($p(t) = 0$ for $t < 0$ min), $t_0 \gg 0$ min is the time of the first observation, and m_0 is an initial mRNA abundance at t_0 which is inferred as a parameter of the model. No parametric assumptions are made about the shape of the pol-II time series function $p(t)$, and the only assumption about the mRNA level $m(t)$ is that it arises through the differential equation.

The linear differential equation **[S2]** and its linear solution operator **[S3]** are similar to those used previously in (5–7) except for the added delay. As in the previous works, the linearity of the solution operator permits exact joint Gaussian process (GP) modelling over $p(t)$ and $m(t)$.

Gaussian process inference

We model pol-II and mRNA time series, $p(t)$ and $m(t)$, in a nonparametric fashion which avoids the assumption of a specific parametric shape for the time series function; instead, we set a GP prior over the time series functions.

For each gene, GP inference of the posterior distribution over the underlying pol-II and mRNA time series can be done in closed form given fixed values of the differential equation parameters. GP inference is based on mean and covariance functions. Below we describe the GP model of pol-II and mRNA, their respective mean functions and covariance functions, and the cross-covariance function between pol-II and mRNA. GP inference of the posterior is then a standard inference equation which we provide for completeness.

The above inference provides a posterior distribution over the profiles $p(t)$ and $m(t)$ given known values for the differential equation parameters. However, these values are not known and to infer a full posterior over both time series and these parameters we carry out Markov Chain Monte Carlo (MCMC) sampling over the parameter values, as described in Section "Parameter inference by Hamiltonian Monte Carlo sampling".

GP model of pol-II. For each gene, we model the pol-II activity time series in a nonparametric fashion by applying a GP prior over the shapes of the time series. Previous similar GP models (5–7) have used a squared exponential covariance function for $p(t)$, as that allows derivation of all the shared covariances in closed form. This covariance has the limitation that it is stationary, and functions following it revert to zero away from data. These properties severely degrade its performance on our highly unevenly sampled data. To avoid this, we model $p(t)$ as an *integral* of a function having a GP prior with a squared exponential covariance: then the posterior mean of $p(t)$ tends to remain constant between observed data. That is, we model

$$p(t) = p_0 + \int_{u=0}^{t} v(t)\mathrm{d}t \qquad \textbf{[S4]}$$

where p_0 is the initial value at time $t = 0$ min, and assign a GP prior with the squared exponential covariance over $v(t)$; as a result $p(t)$ will also have a GP prior whose covariance function is an integral of the covariance function of $v(t)$. For mathematical convenience we assume $p(t) = 0$ min for $t < 0$ min, and set the initial observation time t_0 to a sufficiently large value to avoid any discontinuity resulting from assumption in pol-II or mRNA modeling.

To define the GP prior, we first define the mean function of $p(t)$. Assume that $v(t)$ is drawn from a zero-mean GP prior with a squared exponential covariance function $k_v(t,t') = C_p \cdot \exp(-(t-t')^2/l^2)$ where C_p is a magnitude parameter and l is a length scale, which has been parametrised in a non-standard manner to simplify the derivations. Then $E[p(t)] = E[p_0] + \int_{u=t_0}^{t} E[v(t)]\mathrm{d}t = E[p_0] \equiv \mu_p$ for $t \geq 0$ min.

Next we compute the corresponding covariance function for the GP prior of $p(t)$. We have

$$\begin{aligned} k_p(t,t') &\equiv E[(p(t)-\mu_p)(p(t')-\mu_p)] = \int_{s=0}^{t}\int_{s'=0}^{t'} k_v(s,s')\mathrm{d}s\mathrm{d}s' \\ &= \frac{\sqrt{\pi}C_p l}{2}\int_{s=0}^{t}\Big(\mathrm{erf}((t'-s)/l) - \mathrm{erf}(-s/l)\Big)ds \qquad \textbf{[S5]} \end{aligned}$$

The remaining integral over the erf functions can be computed using integration by parts. After straightforward manipulation, the integral becomes

$$\begin{aligned} k_p(t,t') = \frac{C_p\sqrt{\pi}l^2}{2}\Big(& t_l \mathrm{erf}\left(t_l\right) + t_l' \mathrm{erf}\left(t_l'\right) \\ & - (t_l' - t_l)\mathrm{erf}\left(t_l' - t_l\right)\Big) + \frac{C_p l^2}{2}\Big(\exp\left(-t_l^2\right) \\ & + \exp\left(-(t_l')^2\right) - \exp\left(-(t_l' - t_l)^2\right) - 1\Big). \qquad \textbf{[S6]} \end{aligned}$$

where we denoted $t_l = t/l$ and $t_l' = t'/l$ for brevity.

The right-hand side is the covariance function $k_p(t,t')$ of the integrated squared-exponential GP prior for pol-II.

GP model of mRNA. We model the mRNA abundance in a similar nonparametric fashion as the pol-II activity. Since the mRNA is related to pol-II through a differential equation, the GP prior of mRNA can be computed from the GP prior of pol-II through the differential equation. In particular, as shown in Eq. **[S3]**, the mRNA time series is an integral of the pol-II time series. Since integration is a linear operation, the expected mRNA time series is an integral of the expected pol-II time series; that is, the GP mean function of mRNA is an integral of the mean function of pol-II, so that

$$\begin{aligned} \mu_m(t) &\equiv E[m(t)] = m_0 e^{\alpha(t_0-t)} + \\ &\quad \frac{\beta_0}{\alpha}\left(1 - e^{-\alpha t}\right) + \beta e^{-\alpha t}\int_{u=0}^{t} e^{\alpha u} E[p(u-\Delta)]\mathrm{d}u \\ &= m_0 e^{-\alpha(t-t_0)} + \frac{\beta_0}{\alpha}\left(1 - e^{-\alpha t}\right) + \beta e^{-\alpha t}\int_{u=\Delta}^{t} e^{\alpha u}\mu_p \mathrm{d}u \\ &= m_0 e^{-\alpha(t-t_0)} + \frac{\beta_0}{\alpha}\left(1 - e^{-\alpha t}\right) + \frac{\beta\mu_p}{\alpha}(1 - e^{-\alpha(t-\Delta)}) \qquad \textbf{[S7]} \end{aligned}$$

where the third line follows since pol-II activity starts at $t = 0$ min. Note that the start of pol-II activity at $t = 0$ min is for mathematical convenience, and the initial observation time t_0 will be set to a sufficiently large value so that the $t - \Delta \geq 0$ min for all $t \geq t_0$ and hence observed mRNA values are integrated over active pol-II only regardless of delay Δ.

We next compute the corresponding covariance function for the GP prior of mRNA. The covariance function arises from computing the integral relating mRNA to pol-II as follows:

$$\begin{aligned} & k_m(t,t') \equiv E[(m(t) - \mu_m(t))(m(t') - \mu_m(t'))] \\ & = E\Bigg[\left(\beta e^{-\alpha t}\int_{u=0}^{t} e^{\alpha u} p(u-\Delta)\mathrm{d}u - \frac{\beta\mu_p}{\alpha}(1 - e^{-\alpha(t-\Delta)})\right) \\ & \left(\beta e^{-\alpha t'}\int_{u'=0}^{t'} e^{\alpha u'} p(u'-\Delta)\mathrm{d}u' - \frac{\beta\mu_p}{\alpha}(1 - e^{-\alpha(t'-\Delta)})\right)\Bigg] \\ & = \beta^2 E\Bigg[\left(e^{-\alpha t}\int_{u=\Delta}^{t} e^{\alpha u} p(u-\Delta)\mathrm{d}u - \frac{\mu_p}{\alpha}(1 - e^{-\alpha(t-\Delta)})\right) \\ & \left(e^{-\alpha t'}\int_{u'=\Delta}^{t'} e^{\alpha u'} p(u'-\Delta)\mathrm{d}u' - \frac{\mu_p}{\alpha}(1 - e^{-\alpha(t'-\Delta)})\right)\Bigg] \\ & \qquad \textbf{[S8]} \end{aligned}$$

where the last equality follows since pol-II activity starts at time 0 min. The computation of the integrals follows similar steps as computation of the pol-II GP covariance. The result is

$$k_m(t,t') = k_{m,1}(t,t') + k_{m,2}(t,t') + k_{m,3}(t,t') + k_{m,4}(t,t') \qquad \textbf{[S9]}$$

where we divided the covariance function into four parts. The first part is

$$\begin{aligned} k_{m,1}(t,t') = \frac{\sqrt{\pi}lC_p\beta^2}{2\alpha^2}\Bigg(\left(t_\Delta - \frac{1}{\alpha} + \frac{\exp(-\alpha t'_\Delta)}{\alpha}\right)\mathrm{erf}\left(\frac{t_\Delta}{l}\right) \\ + \left(t'_\Delta - \frac{1}{\alpha} + \frac{\exp(-\alpha t_\Delta)}{\alpha}\right)\mathrm{erf}\left(\frac{t'_\Delta}{l}\right) - (t_\Delta - t'_\Delta)\mathrm{erf}\left(\frac{t_\Delta - t'_\Delta}{l}\right)\Bigg) \qquad \textbf{[S10]} \end{aligned}$$

where $t_\Delta = \max(0 \text{ min}, t - \Delta)$. The second part is

$$\begin{aligned} k_{m,2}(t,t') = \frac{l^2 C_p\beta^2}{2\alpha^2}\Bigg(\exp\left(-\left(\frac{t_\Delta}{l}\right)^2\right) + \exp\left(-\left(\frac{t'_\Delta}{l}\right)^2\right) \\ - \exp\left(-\left(\frac{t_\Delta - t'_\Delta}{l}\right)^2\right) - 1\Bigg). \qquad \textbf{[S11]} \end{aligned}$$

The third part is

$$k_{m,3}(t,t') = -\sqrt{\pi} l C_p \beta^2 \Bigg(\frac{\alpha^{-3}}{4} \exp(\alpha^2 l^2/4 + \alpha(t_\Delta - t'_\Delta)) \big(\mathrm{erf}(\alpha l/2 + t_\Delta/l) - \mathrm{erf}\big(\alpha l/2 + (t_\Delta - t'_\Delta)/l\big)\big) + \frac{\alpha^{-3}}{4} \exp(\alpha^2 l^2/4 - \alpha t'_\Delta - \alpha t_\Delta) \big(\mathrm{erf}(\alpha l/2) - \mathrm{erf}(\alpha l/2 - t'_\Delta/l)\big) - \frac{\alpha^{-3}}{2} \exp(\alpha^2 l^2/4 - \alpha t'_\Delta) \big(\mathrm{erf}(\alpha l/2) - \mathrm{erf}(\alpha l/2 - t'_\Delta/l)\big) \Bigg) . \quad \textbf{[S12]}$$

The fourth part is

$$k_{m,4}(t,t') = -\sqrt{\pi} l C_p \beta^2 \Bigg(\frac{\alpha^{-3}}{4} \exp(\alpha^2 l^2/4 - \alpha(t_\Delta - t'_\Delta)) \big(\mathrm{erf}(\alpha l/2 + t'_\Delta/l) - \mathrm{erf}(\alpha l/2 - (t_\Delta - t'_\Delta)/l)\big) + \frac{\alpha^{-3}}{4} \exp(\alpha^2 l^2/4 - \alpha t_\Delta - \alpha t'_\Delta) \big(\mathrm{erf}(\alpha l/2) - \mathrm{erf}(\alpha l/2 - t_\Delta/l)\big) - \frac{\alpha^{-3}}{2} \exp(\alpha^2 l^2/4 - \alpha t_\Delta) \big(\mathrm{erf}(\alpha l/2) - \mathrm{erf}(\alpha l/2 - t_\Delta/l)\big) \Bigg) . \quad \textbf{[S13]}$$

GP joint model. To define the full GP prior over both pol-II and mRNA, it remains to define the cross-covariance function between pol-II and mRNA. The full GP covariance is defined by the individual covariances of pol-II and mRNA and the cross-covariance.

The cross-covariance function between (noiseless) mRNA abundance $m(t)$ at time t and (noiseless) pol-II activity $p(t')$ at time t' is computed with similar steps as the computation of the mRNA covariance function. The result is

$$\begin{aligned} k_{mp}(t,t') &= E[(m(t) - \mu_m(t))(p(t') - \mu_p(t'))] \\ &= k_{mp,1}(t,t') + k_{mp,2}(t,t') + k_{mp,3}(t,t') \end{aligned} \quad \textbf{[S14]}$$

where for convenience we separated the kernel function into a sum of three components. The first component part of the kernel can be written as

$$\begin{aligned} k_{mp,1}(t,t') = & \\ & - \frac{\sqrt{\pi}\beta^2\sqrt{C_p}l}{2\alpha^2} \exp\left(\left(\frac{\alpha l}{2}\right)^2 - \alpha t_\Delta + \alpha t' \right) \\ & \cdot \left[\mathrm{erf}\left(\frac{\alpha l}{2} + \frac{t'}{l} \right) - \mathrm{erf}\left(\frac{\alpha l}{2} + \frac{t' - t_\Delta}{l} \right) \right] \\ & - \frac{\sqrt{\pi}\beta^2\sqrt{C_p}l}{2\alpha^2} \exp\left(\left(\frac{\alpha l}{2}\right)^2 - \alpha t_\Delta \right) \\ & \cdot \left[\mathrm{erf}\left(\frac{\alpha l}{2} - \frac{t_\Delta}{l} \right) - \mathrm{erf}\left(\frac{\alpha l}{2} \right) \right] . \end{aligned} \quad \textbf{[S15]}$$

The second component can be written as

$$\begin{aligned} k_{mp,2}(t,t') = & \\ & - \frac{\beta^2\sqrt{C_p}l^2}{2\alpha} \left[\exp\left(- \left(\frac{t_\Delta - t'}{l} \right)^2 \right) - \exp\left(- \left(\frac{t_\Delta}{l} \right)^2 \right) \right. \\ & \left. + 1 - \exp\left(- \left(\frac{t'}{l} \right)^2 \right) \right] . \end{aligned} \quad \textbf{[S16]}$$

The third component can be written as

$$\begin{aligned} k_{mp,3}(t,t') = & -\frac{\sqrt{\pi}\beta^2\sqrt{C_p}l}{2\alpha} \Big[(t_\Delta - t' - 1/\alpha)\mathrm{erf}((t_\Delta - t')/l) \\ & - (t_\Delta - 1/\alpha)\mathrm{erf}(t_\Delta/l) - (t' + \exp(-\alpha t_\Delta)/\alpha)\mathrm{erf}(t'/l) \Big] . \end{aligned} \quad \textbf{[S17]}$$

Observation model. In order to fit the models of the pol-II and mRNA functions to observations, we need an observation model. It is assumed that we observe noisy values $\tilde{m}(t) = m(t) + e_m(t)$ and $\tilde{p}(t) = p(t) + e_p(t)$ where $e_m(t)$ and $e_p(t)$ are zero-mean Gaussian noise independently sampled for each time point. For simplicity we assume the noise variance of $e_p(t)$ is a constant σ_p^2 and infer it as a parameter of the model. We estimate the mRNA noise variances $\sigma_m^2(t)$ for each time point t as sums of a shared constant σ_m^2 and a fixed variance inferred by BitSeq by combining the technical quantification uncertainty from BitSeq expression estimation with an estimate of biological variance from the BitSeq differential expression model (full details are in Sec. RNA-seq data processing).

Since the noise is zero-mean, the GP prior for the noisy observations has the same means as the noiseless means, that is, $E[\tilde{m}(t)] = E[m(t)]$ and $E[\tilde{p}(t)] = E[p(t)]$. Since the noise is independently added to each observation, the covariance function of observed pol-II becomes

$$k_{\tilde{p}}(t,t') = k_p(t,t') + \delta(t,t')\sigma_p^2 \quad \textbf{[S18]}$$

where $\delta(t,t') = 1$ if $t = t'$ and zero otherwise, the covariance function of observed mRNA becomes

$$k_{\tilde{m}}(t,t') = k_m(t,t') + \delta(t,t')\sigma_m^2(t) , \quad \textbf{[S19]}$$

and the cross-covariance function between observed pol-II and mRNA is the same as the noiseless version so that

$$k_{\tilde{m}\tilde{p}}(t,t') = k_{mp}(t,t') . \quad \textbf{[S20]}$$

The GP prior over time series functions and the observation model together define a full probability model for the pol-II and mRNA data. As the observations are noisy and available only at a small set of time points, we will apply Bayesian inference to infer the underlying time series $m(t)$ and $p(t)$ from the observations.

Covariance matrix for GP inference. Given a set of time points, here the 10 time points

$$\begin{aligned} T_{obs} &= t_0 + (0, 5, 10, 20, 40, 80, 160, 320, 640, 1280) \\ &= (t_1, \ldots, t_N) \end{aligned}$$

where t_0 is the initial observation time and the numbers denote time in minutes, and the corresponding observation data consisting of $N = 10$ pol-II observations and $N = 10$ mRNA observations $\mathcal{D} = (\tilde{p}(t_1), \ldots, \tilde{p}(t_N), \tilde{m}(t_1), \ldots, \tilde{m}(t_N))$, we wish to compute the posterior distribution of GP hyperparameters, and to predict the shape of the underlying time series functions $p(t)$ and $m(t)$ given the posterior. We will especially wish to study delay between pol-II and mRNA; for mathematical convenience we set $t_0 = 300$ min and consider mRNA delay parameters $0 \leq \Delta \leq 300$ min.

For GP inference, given the hyperparameters we must compute the prior GP covariance matrix for the observations $\mathcal{D}$. We describe the matrix here in a general form which is needed later for inference of time series values at previously unseen time points.

The covariance matrix describes covariance between measurements at one set of time points (indexed by rows of the matrix) and another set of time points (indexed by columns of the matrix). Let $T_{row} = (t_{row,1}, \ldots, t_{row,N_{row}})$ be a vector of N_{row} time indices for rows of the matrix, and let $T_{col} = (t_{col,1}, \ldots, t_{col,N_{col}})$ be the vector of N_{row} time indices for columns of the matrix.

The resulting covariance matrix $K(T_{row}, T_{col})$ has the block structure

$$K(T_{row}, T_{col}) = \begin{bmatrix} K_{\tilde{p}} & K_{\tilde{p}\tilde{m}} \\ K_{\tilde{m}\tilde{p}} & K_{\tilde{m}} \end{bmatrix} \quad \textbf{[S21]}$$

where each block is a $N_{row} \times N_{col}$ matrix of covariance function values between the time points $t \in T_{row}$ and the time points $t' \in T_{col}$, so that $K_{\tilde{p}}$ is composed of values $k_{\tilde{p}}(t,t')$, $K_{\tilde{m}}$ is composed of values $k_{\tilde{m}}(t,t')$, $K_{\tilde{m}\tilde{p}}$ is composed of the cross-covariance values $k_{\tilde{m}\tilde{p}}(t,t')$, and $K_{\tilde{p}\tilde{m}}$ is composed of the cross-covariance values $k_{\tilde{m}\tilde{p}}(t',t)$. The covariance matrix of observed data is then simply $K_{obs} = K(T_{obs}, T_{obs})$.

Marginal likelihood function. The analytical tractability of the GP model allows us to marginalise out the latent functions $p(t)$ and $m(t)$ to compute a marginal likelihood that only depends on the model parameters. The marginal probability density of the observations $\mathcal{D}$ is Gaussian and the marginal log-likelihood is

$$\log P(\mathcal{D}) \;=\; (1/2)(-d\log(2\pi) - \log(|K_{obs}|) - u^\top K_{obs}^{-1} u) \quad \textbf{[S22]}$$

where P denotes the marginal probability density, $d = 20$ is the total number of pol-II and mRNA observations and u is the column vector of observations with their expected values subtracted, $u = \big[\tilde{p}(t_1) - E[\tilde{p}(t_1)], \ldots, \tilde{p}(t_N) - E[\tilde{p}(t_N)], \tilde{m}(t_1) - E[\tilde{m}(t_1)], \ldots, \tilde{m}(t_N) - E[\tilde{m}(t_N)]\big]^\top$.

Posterior prediction. The analytical tractability of the GP model also allows us to obtain the full posterior distribution over the latent functions in closed form given the parameters. Given the observed data, we can thus compute the mean and covariance of the underlying time series function values at each time point, as expectations over the posterior distribution of the underlying functions. For N^* new time points $T^* = (t_1^*, \ldots, t_{N^*}^*)$ the posterior mean is

$$E[[\tilde{p}(t_1^*), \ldots, \tilde{p}(t_{N^*}^*), \tilde{m}(t_1^*), \ldots, \tilde{m}(t_{N^*}^*)]|\mathcal{D}] = u^*_{\text{prior}} + K(T^*, T_{obs}) K_{obs}^{-1} u \quad \textbf{[S23]}$$

where

$$u^*_{\text{prior}} = \big[E[\tilde{p}(t_1^*)], \ldots, E[\tilde{p}(t_{N^*}^*)], E[\tilde{m}(t_1^*)], \ldots, E[\tilde{m}(t_{N^*}^*)]\big]^\top \quad \textbf{[S24]}$$

is the vector of prior means computed at the new time points, and the posterior covariance matrix is

$$Cov[(\tilde{p}(t_1^*), \ldots, \tilde{p}(t_{N^*}^*), \tilde{m}(t_1^*), \ldots, \tilde{m}(t_{N^*}^*))|\mathcal{D}] = K(T^*, T^*) - K(T^*, T_{obs}) K_{obs}^{-1} K(T^*, T_{obs})^\top . \quad \textbf{[S25]}$$

The log-likelihood and predictions of function values described here are computed given fixed values of hyperparameters of the GP prior and the observation model. We will compute a posterior distribution for the hyperparameters, given suitable prior distributions for each. This will allow summarisation of underlying pol-II and mRNA functions and GP parameters over the posterior distribution of the hyperparameters. We next describe the prior distributions of hyperparameters and then describe the sampling based inference of hyperparameter posterior distributions.

Parameter prior distributions. All parameters except the delay Δ have approximately uniform bounded logistic-normal priors. These priors were used because of convenience: they allow easy Hamiltonian Monte Carlo sampling that requires very little tuning (see below for details).

The density of the logistic-normal prior $\text{logit-normal}(\mu, \sigma^2, a, b)$ with location parameter μ and scale parameter σ^2 for variable θ bounded to the interval $[a, b]$ is

$$p(\theta|\mu, \sigma^2, a, b) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\text{logit}((\theta - a)/(b - a)) - \mu)^2}{2\sigma^2}\right) \cdot \frac{b - a}{(\theta - a)(b - \theta)}, \quad \textbf{[S26]}$$

where $\text{logit}(x) = \log(x/(1 - x))$. We use $\mu = 0$ and $\sigma = 2$ which lead to an approximately uniform distribution on the interval $[a, b]$. The interval bounds a, b for all variables are presented in Table S1.

For the delay Δ we use a prior with $\mu = -2, \sigma = 2$ to reflect our prior belief that the delays should in most cases be small. For β and l we set the priors with respect to β^2 and $2/l^2$ respectively, because these are more convenient as model parameters.

Table S1: Bounds for bounded logistic-normal priors of differential equation parameters in the GP inference of pol-II and mRNA time series. Each parameter is bounded to an interval $[a, b]$, we list the values of the lower bound a and upper bound b. Here $\hat{\sigma}^2_{Pol2}$ is the empirical variance of the pol-II time series after preprocessing.

Parameter	Lower bound a	Upper bound b
$2/l^2$	$(1280\ \text{min})^{-2}$ $\approx 6.1 \cdot 10^{-7}\ \text{min}^{-2}$	$(5\ \text{min})^{-2}$ $= 4 \cdot 10^{-2}\ \text{min}^{-2}$
C_p	$2 \cdot 10^{-4}\ \hat{\sigma}^2_{Pol2}$	$\hat{\sigma}^2_{Pol2}$
σ_p^2	$0.05\hat{\sigma}^2_{Pol2}$	$\hat{\sigma}^2_{Pol2}$
α	$1 \cdot 10^{-6}\ \text{min}^{-1}$	$\log(2)\ \text{min}^{-1} \approx 0.69\ \text{min}^{-1}$
β^2	$1 \cdot 10^{-6}\ \text{min}^{-2}$	$1\ \text{min}^{-2}$
Δ	0 min	299 min
β_0	$0\ \text{min}^{-1}$	$1\ \text{min}^{-1}$
m_0	0	2
μ_p	0	1

Parameter inference by Hamiltonian Monte Carlo sampling. Given the data and the priors for the parameters, we apply fully Bayesian inference with Hamiltonian Monte Carlo (HMC) sampling (8) to obtain samples from the posterior distribution of the parameters. HMC is a MCMC algorithm that uses gradients of the target distribution to simulate a Hamiltonian dynamical system with an energy function based on the target distribution. This allows taking long steps while maintaining a high acceptance rate in the sampling.

In order to apply HMC more easily, we transform all parameters to an unbounded space using the logistic transformation. The logistic-normal priors correspond to normal priors on the transformed variables, which effectively prevent the sampler from wandering off to the saturated region of the transformation near the bounds of the intervals.

We run 4 parallel chains starting from different random initial states to allow convergence checking. We use the HMC implementation from NETLAB toolkit in Matlab with momentum persistence and number of leap frog steps $\tau = 20$ which were found to work well in all cases. The step length ϵ is tuned separately for every model (see below). After tuning, each chain is run for 10000 iterations. The samples are then thinned by a factor of 10, and the first half of the samples are discarded, leaving 500 samples from each chain, 2000 in all. Convergence is monitored using the potential scale reduction factor $\sqrt{\hat{R}}$ (9). $\sqrt{\hat{R}}$ is computed separately for each variable, and if any of them is greater than 1.2, the result is discarded and a new sample obtained in a similar manner. The 9 genes that did not converge after 10 iterations of this process were removed from further analysis. In most cases these had severely multimodal delay distributions that were difficult to sample from and would have made further analysis difficult.

Tuning

The applied logistic transformation and priors together allow using the same global step length ϵ for all variables, or using the identity matrix as the mass matrix in the HMC formulation. The step length ϵ was determined by trying different alternatives in the set $\{10^{-5}, 10^{-4}, 10^{-3}, 0.003, 0.005, 0.01, 0.03, 0.05, 0.07, 0.1, 0.3, 0.5, 1\}$ in increasing order, running the sampler for 100 steps and using the largest value with at least 80% acceptance rate. This target rate is higher than usual in random walk MCMC because HMC acceptance rate should be nearly 100% even with very long steps if the Hamiltonian system is simulated accurately.

Summarisation of inference results

The inference results are summarised using the median of the posterior samples. This is a convenient statistic because it is invariant to transformations of the parameter space, such as those used during the sampling.

Validation of the GP modelling results. In order to validate the GP model, we implemented inference for the same ODE using a smoothing spline fit for pol-II. A comparison of the results for the subset of genes that yielded reliable results with the spline approach is presented in Fig. S10.

Filtering of results

Reliable posterior samples were obtained for models of 4373 genes. 4304 of these had multiple-exon transcripts, and could thus be used for intron analyses. These genes were further filtered to remove bad fits by only keeping genes that satisfy the following:

1. The global maximum $t_{\max}$ of $p(t)$ posterior mean $\bar{p}(t)$ in the interval $t \in [0\text{ min}, 1280\text{ min}]$ occurs in the interval $t_{\max} \in (1\text{ min}, 160\text{ min})$. This condition ensures the profile has a peak in the densely sampled region which is necessary for accurate estimation of the delay.
2. The posterior median delay $\hat{\Delta} < 120$ min. Because of the increasingly sparse sampling, longer delay estimates were considered unreliable. The specific cut-off was determined by visual inspection of the fits to rule out implausible ones.
3. The posterior mean $\bar{p}(t)$ of $p(t)$ does not change too much just before $t = 0$ min. This condition is necessary to avoid cases where a long delay pushes distinctive features of $m(t)$ to $p(t), t < 0$ min, which conflicts with the assumption that the system is at a steady state until $t = 0$ min. Quantitatively, we define an index

$$D = D_- - D_+ = D_{[-30\text{ min},0\text{ min}]} - D_{[0\text{ min},10\text{ min}]} \quad \textbf{[S27]}$$

where

$$D_I = \left(\max_{t\in I}[\bar{p}(t)] - \min_{t\in I}[\bar{p}(t)] \right) / \max_{t\in[-30\text{ min},1280\text{ min}]}[\bar{p}(t)], \quad \textbf{[S28]}$$

and only include genes with

$$D < 0.05. \quad \textbf{[S29]}$$

Intuitively, D_I looks at the magnitude of change in $\bar{p}(t)$ in the interval I relative to the global magnitude of change in $\bar{p}(t)$. The final statistic D looks for genes that have small changes in $[-30\text{ min}, 0\text{ min}]$, but forgives genes with early large changes in $[0\text{ min}, 10\text{ min}]$ because these would often spill over to $t < 0$ min because of the properties of the GP model. The cut-off 0.05 represents 5% change in magnitude, which seems reasonably small. The main conclusions of the work are robust to different cutoffs, as demonstrated in Fig. S5 below.

After these filtering steps, there were 1814 genes left for the analysis.

Main results under an additional filter of setting a maximum for posterior inter-quartile range are presented in Fig. S11.

Synthetic data generation

The synthetic data were generated by fitting a GP with the MLP covariance (10) to the Pol-II measurements of the gene TIPARP (ENSG00000163659), and numerically solving the mRNA level using Eq. **[S3]** with the GP posterior mean as $p(t)$. The parameters used were: $\Delta \in \{0, 10, 20, 30\}$ min, $t_{1/2} = \log(2)/\alpha \in \{2, 4, 8, 16, 32, 64\}$ min, $\beta_0 = 0.005$, $\beta = 0.03$, $m_0 = 0.008/\alpha$. The parameter values were chosen empirically to get profiles that approximately fitted the actual mRNA observations while looking reasonable and informative across the entire range of parameter values.

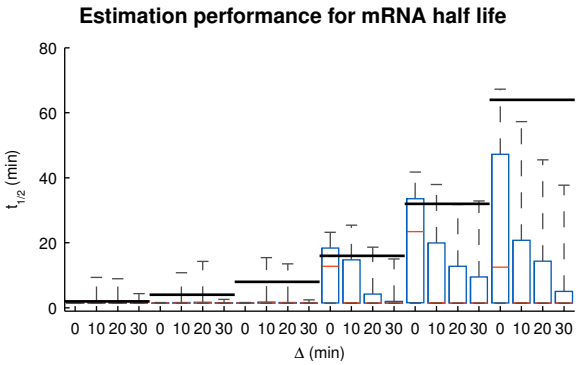

Fig. S1: Boxplots of parameter posterior distributions illustrating parameter estimation performance on synthetic data for the mRNA half life $t_{1/2} = \log(2)/\alpha$. The strong black lines indicate the ground truth used in data generation. The box extends from 25th to 75th percentile of the posterior distribution while the whiskers extend from 9th to 91st percentile. The model often underestimates the half lives, especially in the presence of a significant delay.

Supplementary Results

In this section we provide supplementary Figs. S1–S8 discussed in the main paper as well as Figs. S9–S11 discussed in the Supplementary Methods.

SI References

1. **Glaus, P, Honkela, A, Rattray, M (2012) Identifying differentially expressed transcripts from RNA-seq data with biological variation. *Bioinformatics* 28:1721–1728.**
2. **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.**
3. **Anders, S, Huber, W (2010) Differential expression analysis for sequence count data. *Genome Biol* 11:R106.**
4. **Robinson, MD, McCarthy, DJ, Smyth, GK (2010) edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. *Bioinformatics* 26:139–140.**
5. **Lawrence, ND, Sanguinetti, G, Rattray, M (2007) in *Advances in Neural Information Processing Systems*, eds Schölkopf, B, Platt, JC, Hofmann, T (MIT Press, Cambridge, MA) Vol. 19, pp 785–792.**
6. **Gao, P, Honkela, A, Rattray, M, Lawrence, ND (2008) Gaussian process modelling of latent chemical species: applications to inferring transcription factor activities. *Bioinformatics* 24:i70–i75.**
7. **Honkela, A et al. (2010) Model-based method for transcription factor target identification with limited data. *Proc Natl Acad Sci U S A* 107:7793–7798.**
8. **Duane, S, Kennedy, AD, Pendleton, BJ, Roweth, D (1987) Hybrid Monte Carlo. *Physics Letters B* 195:216–222.**
9. **Gelman, A, Rubin, DB (1992) Inference from iterative simulation using multiple sequences. *Statistical Science* 7:457–472.**
10. **Rasmussen, CE, Williams, C (2006) *Gaussian Processes for Machine Learning* (MIT Press).**
11. **Hah, N et al. (2011) A rapid, extensive, and transient transcriptional response to estrogen signaling in breast cancer cells. *Cell* 145:622–634.**

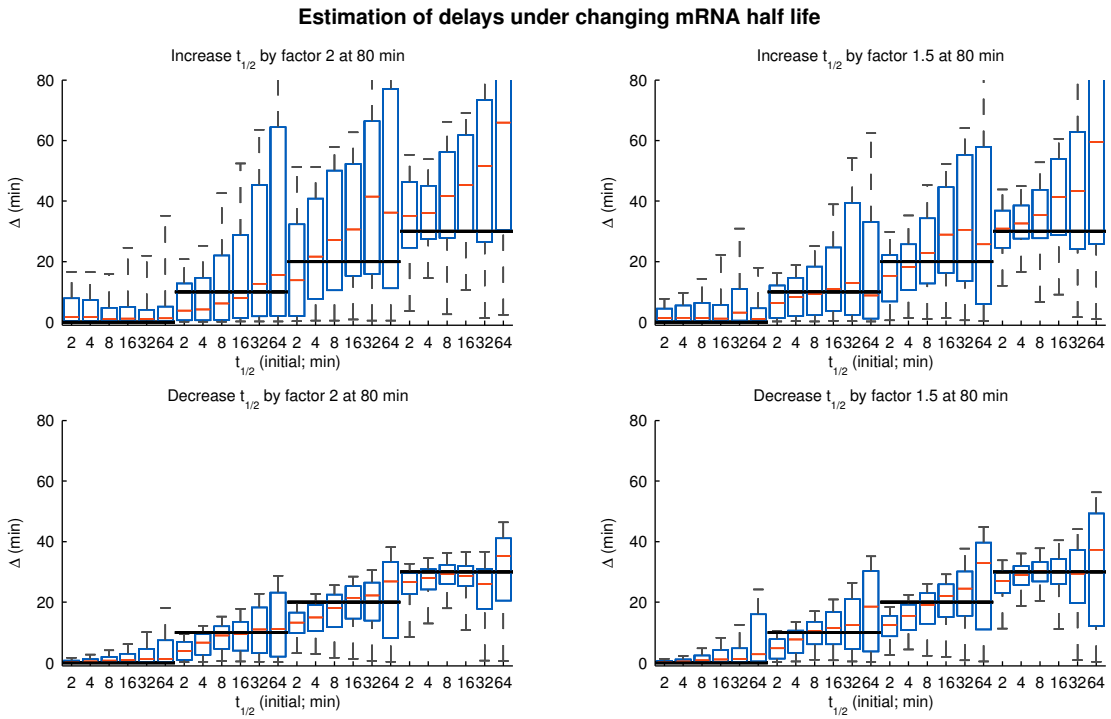

Fig. S2: Boxplots of parameter posterior distributions illustrating parameter estimation performance on synthetic data for the delay parameter Δ. The strong black lines indicate the ground truth used in data generation. The box extends from 25th to 75th percentile of the posterior distribution while the whiskers extend from 9th to 91st percentile. This is a counterpart of Fig. 3 in a situation where the simulated mRNA half-life $t_{1/2}$ changes during the time course, something our model cannot capture. The simulated changes are point changes up or down with a factor of 1.5 or 2 at 80 min. The results show that delay estimates remain accurate and reliable, with the true value always in the high posterior density region, and demonstrate the conservativeness of the estimates with no sign of serious overestimation of small delays.

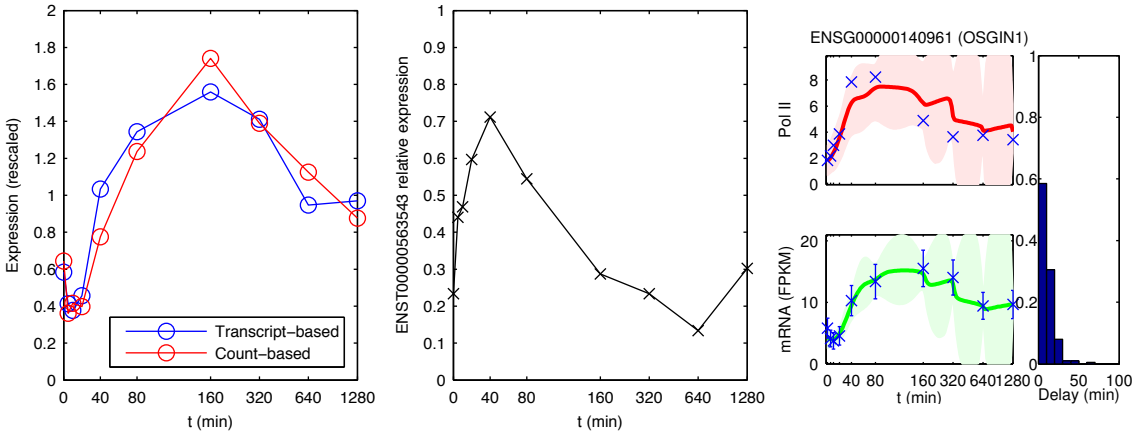

Fig. S3: An illustration of how proper summarisation of the RNA-seq data is important for ruling out confounding effects from changing splicing patterns during the experiment. Left: Gene expression time series of the gene OSGIN1 from RNA-seq data using either transcript-based RNA-seq data summarisation as in the paper or using a simpler summarisation of counts of reads aligned uniquely to the mRNA transcripts of OSGIN1. Centre: Proportion of transcript ENST00000563543 out of all OSGIN1 transcripts. At 567 bp, this mRNA transcript is much shorter than the other major transcripts whose mRNAs are around 2 kb. Right: The model fit for OSGIN1 shows no evidence of significant delay, while the count-based profile in the left figure would suggest a longer delay.

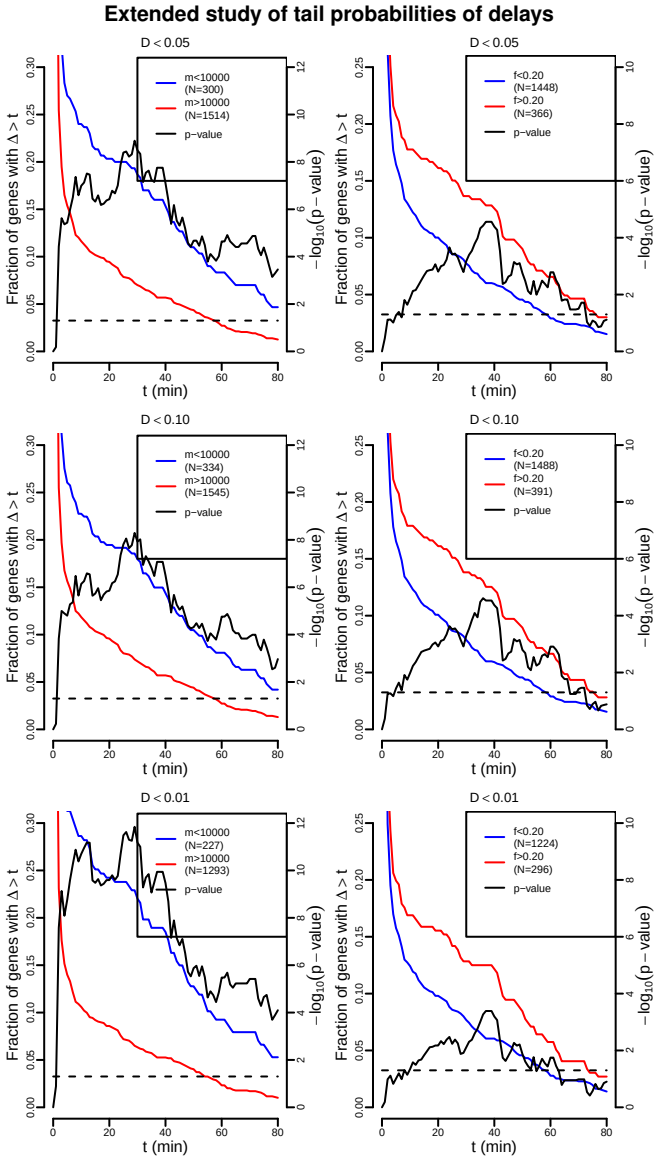

Fig. S4: Alternative versions of Fig. 5 of the main paper: tail probabilities for delays for different cut-offs for D in Eq. **[S29]**. Top: $D < 0.05$ (value used for main results), middle: $D < 0.1$, bottom: $D < 0.01$. Left: genes whose longest pre-mRNA transcript is short (m is the length from transcription start to end). Right: genes with relatively long final introns (f is the ratio of the length of the final intron of the longest annotated transcript of the gene divided by the length of that transcript pre-mRNA). The fraction of genes with long delays Δ is shown by the red and blue lines (left axis). In both subplots, the black curve denotes the p-values of Fisher's exact test conducted separately at each point (right axis) with the dashed line denoting $p < 0.05$ significance threshold. The general shapes of the curves are the same in every case.

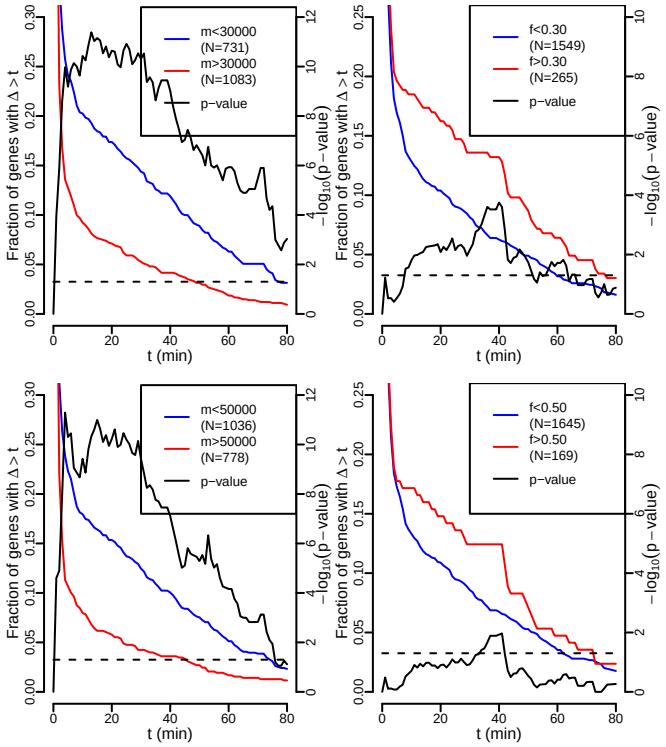

Fig. S5: Alternative versions of Fig. 5 of the main paper: different cut-offs for f and m. Left: genes whose longest pre-mRNA transcript is short (m is the length from transcription start to end). Right: genes with relatively long final introns (f is the ratio of the length of the final intron of the longest annotated transcript of the gene divided by the length of that transcript pre-mRNA). The fraction of genes with long delays Δ is shown by the red and blue lines (left axis). In both subplots, the black curve denotes the p-values of Fisher's exact test conducted separately at each point (right axis) with the dashed line denoting $p < 0.05$ significance threshold. The general shapes of the curves are the same in every case.

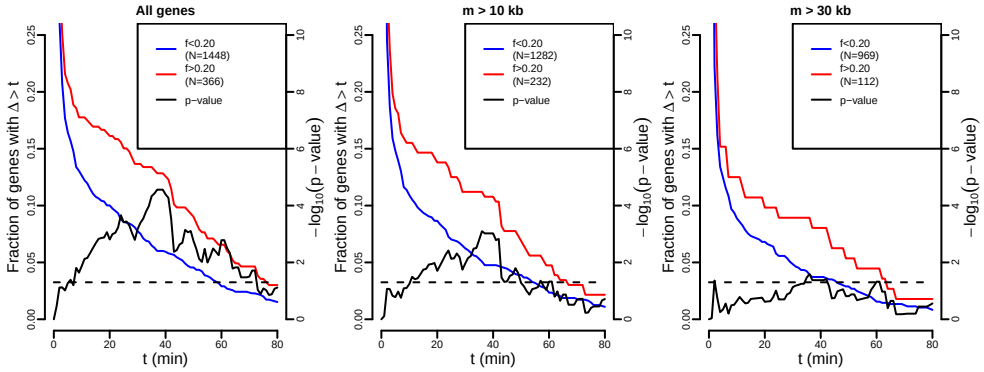

Fig. S6: Alternative versions of Fig. 5 (right) of the main paper: explore the dependence on f for genes with a lower bound on mRNA length m. The fraction of genes with long delays Δ is shown by the red and blue lines (left axis). The black curve denotes the p-values of Fisher's exact test conducted separately at each point (right axis) with the dashed line denoting $p < 0.05$ significance threshold. The general shapes of the curves are the same in every case.

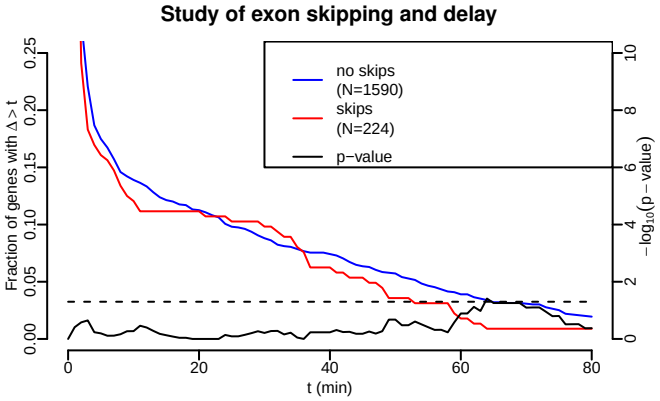

Fig. S7: Proportion of long posterior median delays for genes with/without annotated exon skipping. The fraction of genes with long delays Δ is shown by the red (no skipped exons) and blue (skipped exons) lines (left axis). The black curve denotes the p-values of Fisher's exact test conducted separately at each point (right axis) with the dashed line denoting $p < 0.05$ significance threshold. There is no clear difference between the two groups.

Table S2: Fraction of reads mapping to mRNA transcripts alone (junction reads), pre-mRNA transcripts alone and both across all time points.

t	mRNA	pre-mRNA	both
0 min	0.035	0.287	0.678
5 min	0.033	0.308	0.659
10 min	0.036	0.270	0.694
20 min	0.033	0.321	0.647
40 min	0.033	0.312	0.654
80 min	0.031	0.372	0.597
160 min	0.033	0.331	0.636
320 min	0.033	0.327	0.640
640 min	0.034	0.330	0.636
1280 min	0.032	0.339	0.629

Table S3: Fraction of reads assigned by BitSeq on average to mRNA transcripts and pre-mRNA transcripts, as well as fraction predicted for pre-mRNA when distributing 'both' category reads from Table S2 uniformly according to average transcript lengths. Only multi-exon genes are considered here, because the division is not meaningful for others. The results demonstrate that BitSeq can split the RNA-seq data to mRNA and pre-mRNA fractions in a meaningful manner.

t	mRNA	pre-mRNA	pre-mRNA pred.
0 min	0.62	0.38	0.37
5 min	0.59	0.41	0.39
10 min	0.64	0.36	0.36
20 min	0.58	0.42	0.40
40 min	0.59	0.41	0.40
80 min	0.53	0.47	0.45
160 min	0.57	0.43	0.41
320 min	0.57	0.43	0.41
640 min	0.57	0.43	0.41
1280 min	0.55	0.45	0.42

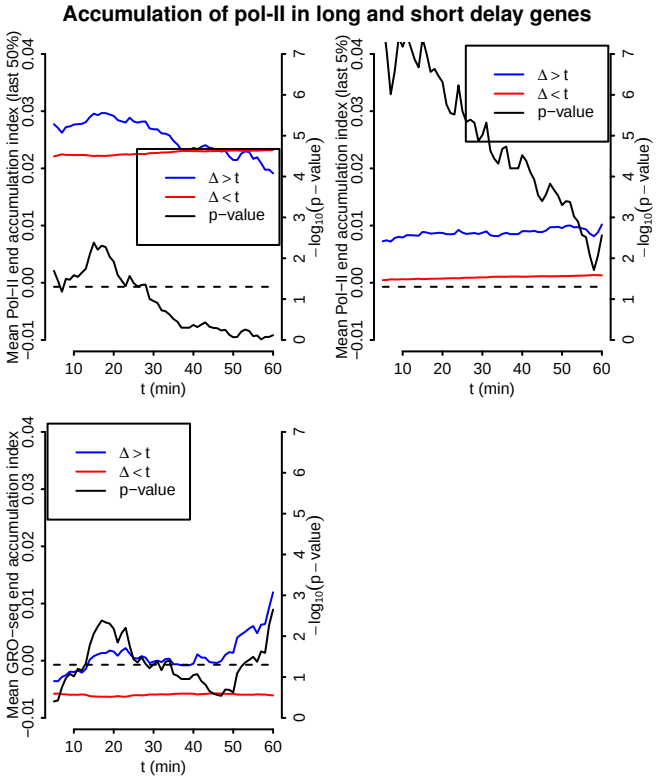

Fig. S8: Alternative version of Fig. 6 (right) of the main paper: computing the index based on pol-II ChIP-seq and GRO-seq instead of intronic RNA-seq reads. The top plots show differences in the mean pol-II accumulation index in long delay genes (blue) and short delay genes (red) as a function of the cut-off used to distinguish the two groups (left axis). Positive values indicate increased pol-II accumulation at the 3' end (top left: last 50% of the gene body, top right: last 5% of the gene body) over time. The black line shows the p-values of Wilcoxon's rank sum test between the two groups at each cut-off (right axis). The bottom plot is the same as top right, except for GRO-seq data of (11), with the index is defined as the difference between the only late (160 min) time point and the average of the early (0-40 min) time points. In contrast to the pre-mRNA figure in the main paper, both long and short delay genes show a clear tendency towards accumulation of pol-II towards the end of the gene, but there is no clear difference between the two groups for the last 50% (top left), while there is a very consistent pattern of more pol-II accumulation very close to 3' end (top right) for long delay genes, and the level is essentially independent of the estimated delay. GRO-seq data in the last 5% (bottom) behave similarly as pol-II ChIP-seq (top right).

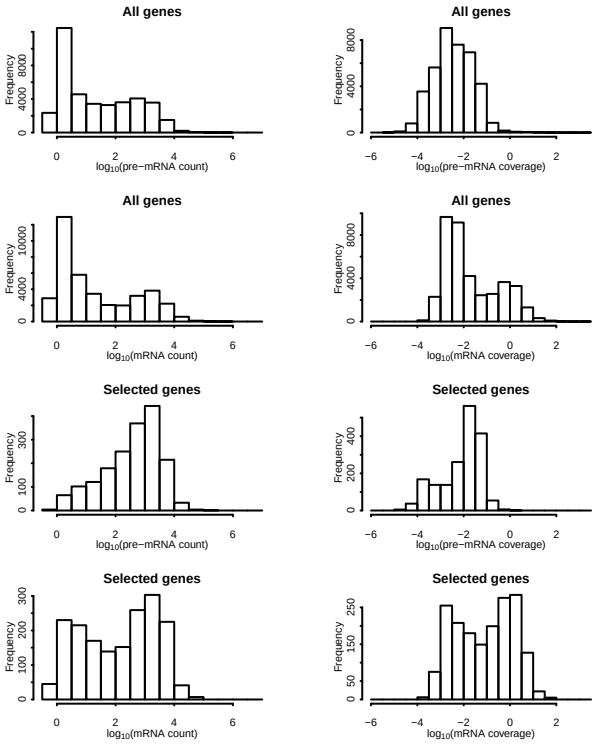

Fig. S9: Distributions of per-gene mRNA and pre-mRNA counts and coverages based on BitSeq expression estimates. Top two rows show broad distributions for all genes, while the bottom two rows show distributions biased toward higher values for the selected 1786 genes. The results show that the mRNA coverages are mostly clearly higher than for pre-mRNA, again demonstrating a sensible split between pre-mRNA and mRNA.

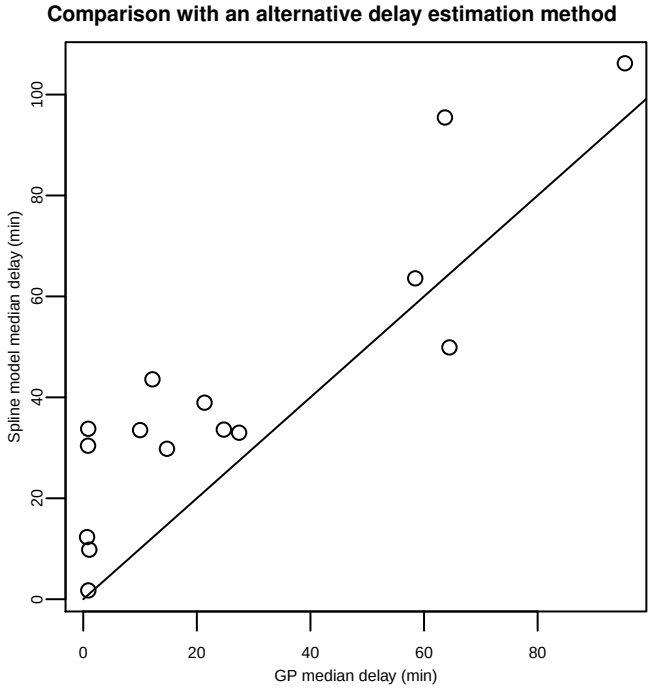

Fig. S10: Comparison of estimated posterior median delays from the GP model with an alternative spline-based model for $n = 15$ genes with reliable estimates. In this model we used cubic smoothing splines to fit continuous curves to pol-II measurements. To account for the uneven sampling the times were transformed as $t' = \log(t/\text{min}+5)$. The regularisation strength was shared over all genes and optimised by leave-one-out cross validation over all internal time points. The time transformation was also found to work much better than untransformed time in the cross validation. The smoothed pol-II curves were used as input to Eq. **[S2]** which was solved numerically to obtain predictions for $m(t)$. Assigning a Gaussian noise model to $m(t)$ similar to the GP model and using similar priors for all shared parameters, we run MCMC to obtain posteriors over the parameters. We were only able to obtain reliable parameter estimates for a small subset of genes for which the method had a good fit (measured through expected relative residual variance) for both pol-II and mRNA. The other estimates were unreliable presumably because the method estimated the pol-II profiles independently but then ignored the uncertainty related to this estimation, which further highlights the benefits of the GP approach.

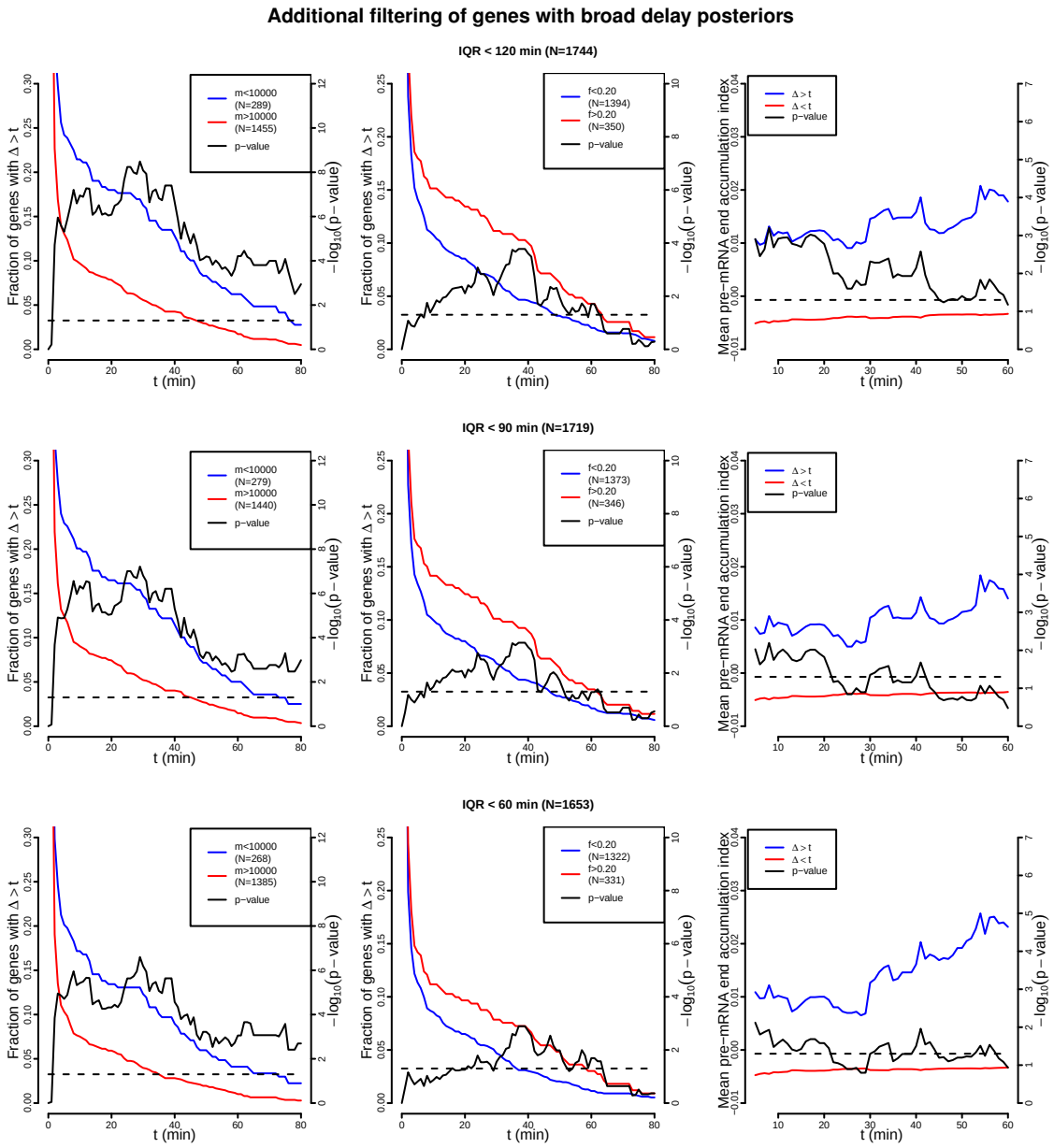

Fig. S11: Alternative versions of Figs. 5 and 6 under more stringent filtering of delay posterior interquartile range (IQR).