Molecular Systems Biology 7; Article number 529; doi:10.1038/msb.2011.62
Citation: ***Molecular Systems Biology* 7: 529**

www.molecularsystemsbiology.com

molecular systems biology

Coupled pre-mRNA and mRNA dynamics unveil operational strategies underlying transcriptional responses to stimuli

Amit Zeisel[1,6], Wolfgang J Köstler[2,6,7], Natali Molotski[3], Jonathan M Tsai[2], Rita Krauthgamer[4], Jasmine Jacob-Hirsch[5], Gideon Rechavi[5], Yoav Soen[1,3], Steffen Jung[4], Yosef Yarden[2,*] and Eytan Domany[1,*]

[1] Department of Physics of Complex Systems, Weizmann Institute of Science, Rehovot, Israel, [2] Department of Biological Regulation, Weizmann Institute of Science, Rehovot, Israel, [3] Department of Biological Chemistry, Weizmann Institute of Science, Rehovot, Israel, [4] Department of Immunology, Weizmann Institute of Science, Rehovot, Israel and [5] Sheba Cancer Research Center, The Chaim Sheba Medical Center and Sackler School of Medicine, Tel Aviv University, Tel Aviv, Israel
[6] These authors contributed equally to this work
[7] Present address: Department of Medicine 1 & Comprehensive Cancer Center, Medical University of Vienna, Vienna, Austria
* Corresponding authors. Y Yarden, Department of Biological Regulation, Weizmann Institute of Science, Rehovot 76100, Israel. Tel.: + 972 8934 4502; Fax: + 972 8934 2488 E-mail: yosef.yarden@weizmann.ac.il or E Domany, Department of Physics of Complex Systems, Weizmann Institute of Science, Rehovot 76100, Israel. Tel.: + 972 8934 3964; Fax: + 972 8934 4109; E-mail: eytan.domany@weizmann.ac.il

Received 28.3.11; accepted 17.7.11

Transcriptional responses to extracellular stimuli involve tuning the rates of transcript production and degradation. Here, we show that the time-dependent profiles of these rates can be inferred from simultaneous measurements of precursor mRNA (pre-mRNA) and mature mRNA profiles. Transcriptome-wide measurements demonstrate that genes with similar mRNA profiles often exhibit marked differences in the amplitude and onset of their production rate. The latter is characterized by a large dynamic range, with a group of genes exhibiting an unexpectedly strong transient production overshoot, thereby accelerating their induction and, when combined with time-dependent degradation, shaping transient responses with precise timing and amplitude.
Molecular Systems Biology **7**: 529; published online 13 September 2011; doi:10.1038/msb.2011.62
Subject Categories: computational methods; RNA
Keywords: half-life; operational strategy; pre-mRNA; transcriptional response

Introduction

A major component of cellular response to changing conditions is a shift of the transcriptome to a new state, which is more adequate for facing the new environment. In eukaryotes, this shift is governed by a highly dynamic interplay between epigenetic, co-transcriptional and post-transcriptional processes, which determine the temporal concentration profiles of RNAs by controlling their production and degradation (Orphanides and Reinberg, 2002; Garneau *et al*, 2007). The most commonly used mathematical description of the transcriptional response is expressed by the following differential equation (Gorini and Maas, 1957):

$$\frac{dM}{dt} = \beta - \alpha M \qquad (1)$$

Here, M is the mRNA concentration, dM/dt is its rate of change with respect to time t, β denotes the rate of transcript production and α is the transcript degradation coefficient. Response of a gene to stimulus is commonly described as an abrupt shift of its β and/or α to new values, which then remain constant (see Figure 1A, B, D and E). mRNA concentration then approaches its new asymptotic value, β/α, with kinetics determined solely by the mRNA degradation coefficient α (Alon, 2007). Thus, a small steady-state value of α implies slow dynamics (a long mRNA half-life, $T_{1/2}=\ln 2/\alpha$) and also supports an economically favorable low production rate (Shalem *et al*, 2008; Elkon *et al*, 2010).

Notably, many organisms across the phylogenetic tree exhibit rapid rise times of long-lived mRNAs, in contradiction to this simple model. In bacteria, accelerated production can be achieved by time-delayed negative autoregulation (Rosenfeld *et al*, 2002) and in yeast through transcriptional control by an incoherent feed-forward loop (Mangan and Alon, 2003; Mangan *et al*, 2003). In mammalian systems, however, the operational strategies which govern transcript production and degradation profiles remain less well characterized.

Important case studies for dynamics of transcript production and degradation are exposure to serum, pathogen components such as lipopolysaccharide (LPS), or growth factors like the epidermal growth factor (EGF). These stimuli initiate well-characterized signaling cascades that culminate in orchestrated transcriptional responses involving *primary* and *secondary response genes* (PRGs and SRGs; Cochran *et al*, 1983; Lau and Nathans, 1987; Iyer *et al*, 1999; Amit *et al*, 2007; Medzhitov and Horng, 2009). PRGs include *immediate-early genes* (IEGs), whose mRNA expression often peaks during the first hour and *delayed early genes* (DEGs), which mostly peak

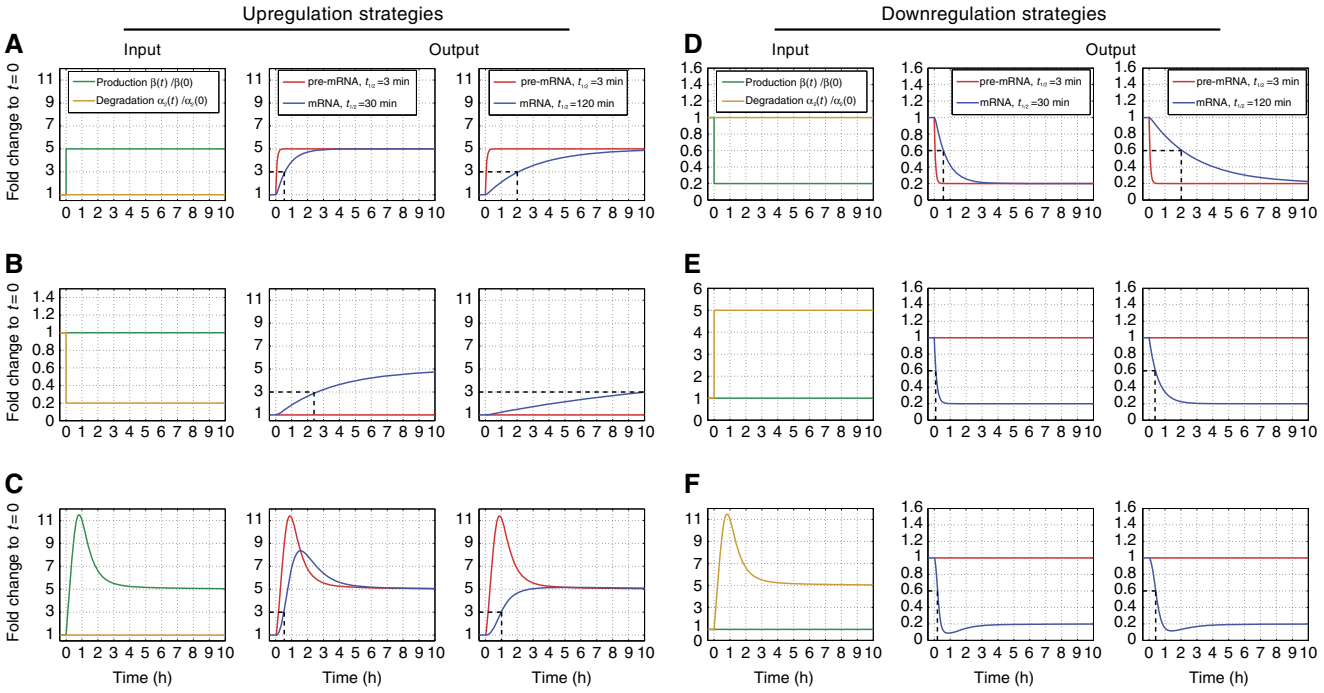


Figure 1 Pre-mRNA and mRNA response profiles for stimulus-induced production and degradation changes. Different time-dependent production rates (green) and degradation coefficients (gold) are shown in the left column. Resulting pre-mRNA (red) and mRNA (blue) profiles, obtained by numerically solving Equations (4) and (5), are presented for mRNAs with initial half-lives of 30 and 120 min in the middle and right columns, respectively. All strategies yield a 5-fold change in the steady-state mRNA level, pre-mRNA profiles closely mimic production, while mRNA FC profiles are delayed and may be qualitatively different. (**A**) Five-fold step increase of production rate at time t=0. A rapid increase of pre-mRNA levels is followed by slower mRNA response. Dashed black line indicates the rise time to half of the aimed level, determined by the mRNA half-life. (**B**) Signal-induced 5-fold step decrease of degradation rate. Pre-mRNA levels reflect unchanged production; mRNA accumulates now more slowly compared with (A). (**C**) Signal-induced transient overshoot in production rate accelerates mRNA induction, particularly for mRNAs with long initial half-lives. (**D**) Five-fold step decrease in production rate at t=0. (**E**) Degradation driven downregulation: response time is shorter due to reduction of the mRNA half-life. (**F**) A transient overshoot in degradation accelerates response time of downregulated genes.

during the second hour and mainly comprise delayed PRGs (Amit *et al*, 2007; Tullai *et al*, 2007; Hao and Baltimore, 2009). While PRGs are rapidly induced by pre-existing transcription factors, without *de novo* protein synthesis (Herschman, 1991; Byun *et al*, 2009; Hargreaves *et al*, 2009; Ramirez-Carrozzi *et al*, 2009; Wang *et al*, 2009), induction of SRGs depends upon newly synthesized activators. Correct identification of transcripts as PRG or SRG is of central importance in order to understand the networks that govern the response to stimulus. According to currently accepted views, IEGs primarily comprise transcriptional regulators and are mostly encoded by short, intron-poor and short-lived transcripts. Their rapid induction is facilitated by permissive chromatin structures and by swift attenuation of pre-existing negative regulators (Hargreaves *et al*, 2009; Avraham *et al*, 2010). Expression of IEGs is controlled by activation-dependent feedback (e.g., receptor endocytosis), by negative autoregulation and by DEGs (Sassone-Corsi *et al*, 1988; Carballo *et al*, 1998; Amit *et al*, 2007). The latter regulate IEGs by transcriptional repression, destabilization of IEG transcripts or by attenuating the signaling pathways that lead to IEG induction. Regulation of DEGs remains less understood. The lag in their expression has been attributed to their distinct promoter properties, and to delays in transcription initiation and elongation (Tullai *et al*, 2007; Ramirez-Carrozzi *et al*, 2009). Intriguingly, many DEGs encode for long-lived mRNAs, yet they are rapidly induced (Tullai *et al*, 2007).

The aim of our study is to uncover and quantify the dynamics of transcriptional responses to stimuli and to elucidate the operational strategies that govern them. To this end, we introduced a simple method that allows, for the first time, genome-wide simultaneous measurement of pre-mRNA and mRNA expression. When combined with a new mathematical model for transcription dynamics, production and degradation profiles (i.e., time-dependent functions) can be reliably inferred from these measured pre-mRNA and mRNA profiles. Our transcriptome-wide study reveals that transcript production rates, reflected by pre-mRNA profiles, exhibit a high dynamic range, and identifies a subset of genes that exhibit novel relationships between transcript production and mRNA abundance profiles. In particular, we identify genes that exhibit pre-mRNA *fold changes* (FCs) that exceed by an unexpectedly high margin those of the associated mRNA. Indeed, such temporally confined *production overshoot* is used to solve the 'conflict' between long mRNA half-life and the need for rapid response. Taken together, our findings reveal complex dynamics of both pre-mRNA production and mRNA degradation rates, which shape the expression profiles of mRNAs in response to extracellular stimuli.

Results

Inference of production and degradation profiles requires measurement of pre-mRNA and mRNA dynamics

Cellular response to stimuli involves time-dependent variation of the production rate β and the degradation coefficient α of Equation (1). Exact inference of *two* unknown functions ($\alpha(t)$ and $\beta(t)$) from the time-dependent mRNA abundance $M(t)$ alone is impossible. Furthermore, the expression profiles of transcripts with typical mRNA half-lives track very poorly the profile $\beta(t)$ of RNA production (Barenco *et al*, 2009; Hao and Baltimore, 2009). Thus, independent assessments of production or degradation rates (Fan *et al*, 2002; Shalem *et al*, 2008; Hao and Baltimore, 2009) are necessary.

Direct measurements of $\alpha(t)$ and $\beta(t)$ involve interference with the system: prototypical methods induce transcription arrest (e.g., by actinomycin D, ActD) followed by measurements of mRNA decay to obtain mRNA half-lives. More recently introduced methods involve biosynthetic tagging by averaging the incorporation of labeled nucleotides (e.g., 4-thiouracil, 4-sU) over a certain time interval to measure newly synthesized RNA (Miller *et al*, 2011; Rabani *et al*, 2011; Schwanhausser *et al*, 2011). When combined with measurements of mRNA abundance, either method can be used to infer production and degradation rates. Transcription arrest may alter transcript stability and thereby the measurement may affect the properties one wishes to measure. Methods based on incorporation of labeled nucleotides are more promising, and the underlying assumptions, that cellular uptake and incorporation of tagged nucleotides is constant over stimulus, and that labeled transcripts are spliced, exported from the nucleus and degraded at the same rate as their unlabeled counterparts, may be valid. By contrast, our method involving simultaneous and direct measurements of mRNA and pre-mRNA abundance profiles, $M(t)$ and $P(t)$, is not only much simpler, but also free of any 'interference' with either transcription or mRNA degradation. Moreover, since no time interval for labeling is required, our method allows for tracking of pre-mRNA and mRNA profiles at high temporal resolution.

Mathematical modeling of the transcription process

We extended the 'minimal model' of Equation (1) and describe production of a particular transcript in terms of $P(t)$ and $M(t)$. Dynamics of these variables are described by two coupled linear differential equations of the form

$$\frac{dP}{dt} = \beta(t) - \alpha_1 P(t) \tag{2}$$

$$\frac{dM}{dt} = \alpha_1 P(t) - \alpha_2(t) M(t) \tag{3}$$

Here, β is the (time-dependent) production rate of pre-mRNA, α_1 denotes the conversion (splicing) coefficient of pre-mRNA to mRNA and $\alpha_2(t)$ represents the degradation coefficient of mRNA. This simple model relies on the following main assumptions:

(i) Time-dependent production rate: in principle, the production rate of a particular transcript may depend on the activity of various proteins (e.g., transcription factors, Pol II and so on), and may occur monotonically or in transcriptional bursts (Suter *et al*, 2011). We assume that all these can be absorbed in an effective (gene-specific) time-dependent production rate $\beta(t)$ (Larson *et al*, 2011; Suter *et al*, 2011). We do not introduce explicit dependence of β on either P or M, because Pol II most likely elongates across the distal transcription units irrespective either of the number of polymerases recruited or of previous rounds of transcription (Darzacq *et al*, 2007; Hager *et al*, 2009; Singh and Padgett, 2009; Wada *et al*, 2009; Larson *et al*, 2011; Suter *et al*, 2011).

(ii) For most genes, the measured data can be explained and fitted well without incorporating in the model processes involving storage or transport of mRNA to specific cellular compartments (e.g., export to the cytoplasm). The main approximation introduced by this assumption is that even though mRNA is degraded only in the cytoplasm, in our model degradation is proportional to $M(t)$, the total mRNA (rather than to the cytoplasmic fraction). Since the vast majority of the mRNA transcripts we study do localize to the cytoplasm (data not shown) and export of newly synthesized mRNA to the cytoplasm occurs much faster than mRNA degradation, using $\alpha_2 M$ as the degradation rate is a good approximation.

(iii) Degradation of pre-mRNA in the nucleus, beyond its conversion to mRNA, is minimal. Hence, all pre-mRNA lost (in Equation (2)) is converted and reappears as mRNA (in Equation (3)).

(iv) The conversion coefficient α_1 is independent of time and is not affected by the stimulus.

As described below, assumptions (iii) and (iv) were confirmed experimentally. Finally, our model conforms with the fact that pre-mRNA splicing temporally coincides with other processing events such as capping, RNA editing and poly(A)-tail addition (Hirose and Manley, 1998; Orphanides and Reinberg, 2002; Proudfoot *et al*, 2002).

In both quantitative PCR (qPCR) and microarray measurements, the output signal for different nucleotide sequences is characterized by different amplification and hybridization efficiencies. Hence, these measurements yield reliably only relative concentrations, for example, FC, measured, for each transcript, with respect to its concentration in some reference condition. Throughout this study, we used the pre-stimulus state, t=0, as our reference, and all relative concentrations shown are the FC with respect to this reference. Hence, we used a trivial transformation of variables to cast Equations (2) and (3) in a form that presents the dynamics of normalized FC variables, $\hat{X}(t)=X(t)/X(0)$, where X stands for any of the quantities P, M, β, α:

$$\frac{d\hat{P}}{dt} = \alpha_1[\hat{\beta}(t) - \hat{P}(t)] \tag{4}$$

$$\frac{d\hat{M}}{dt} = \alpha_2(0)[\hat{P}(t) - \hat{\alpha}_2(t)\hat{M}(t)] \tag{5}$$

These are the actual dynamic equations we used to infer for each transcript its normalized, time-dependent production and degradation rates $\hat{\beta}(t)$ and $\hat{\alpha}_2(t)$. These rates were inferred (see *Materials and methods*) from measurements of the corresponding $\hat{P}(t)$ and $\hat{M}(t)$. Note that for fast conversion, that is, $\alpha_1 \gg 1$, Equation (4) can be rewritten (using first order Taylor expansion) as $\hat{\beta}(t) \approx \hat{P}(t + 1/\alpha_1)$, demonstrating that the production FC profile is given, to good approximation, by the time-shifted pre-mRNA FC.

Pre- and post-stimulus values of conversion coefficients $\alpha_1(t)$ were determined by direct measurements (see below in subsection *Conversion with a constant coefficient dominates pre-mRNA outflux*) while degradation coefficients $\alpha_2(0)$ were derived from the transcript FC profiles by optimization of the fit to the data (see Materials and methods).

The predicted transcript responses to different production and degradation profiles

To demonstrate the time-dependent FC of pre-mRNA and mRNA generated by our model, we solved numerically Equations (4) and (5) for different time-varying forms of transcript production $\hat{\beta}(t)$ and degradation $\hat{\alpha}_2(t)$. We find that the temporal pre-mRNA profile closely resembles the time-dependent transcript production rate $\hat{\beta}(t)$ for both upregulated and downregulated transcripts (Figure 1).

In contrast, the profile of mRNA is affected also by degradation and exhibits much slower kinetics, particularly for mRNAs with long half-lives (Elkon *et al*, 2010). In terms of end point steady-state levels, a 5-fold increase in production (Figure 1A) is equivalent to a 5-fold reduction in degradation (Figure 1B). However, the response time, that is, the time until half the end point change is reached, is much longer in the latter case. Our simulations of upregulation also demonstrate that a simple step to new production or degradation rates cannot generate the rapid dynamics of transcriptional responses that were in fact observed across a series of experimental conditions (Iyer *et al*, 1999; Amit *et al*, 2007, 2009), especially for medium to long-lived mRNAs. In contrast, as seen in Figure 1C, marked acceleration of the response of even long-lived mRNAs *can* be achieved by a strategy of *production overshoot* (defined as a transient increase of the pre-mRNA FC to values that exceed at least twice the maximal mRNA FC). Thus, unlike the response time to simple step changes in β and/or α_2, which is governed by α_2 alone (Alon, 2007; Shalem *et al*, 2008; Elkon *et al*, 2010), our results emphasize the importance of short pulses of production to achieve rapid transcriptional induction.

Simulations of responses to different strategies that lead to downregulation are shown in Figure 1D–F. In order to accelerate downregulation, cells have to transiently increase degradation rates; otherwise even complete arrest of production will result in decay at a rate not faster than the initial half-life of the transcript.

Another point to note is the insensitivity of the mRNA profiles to the precise value of the conversion coefficient α_1. The reason is that typically conversion times are much shorter than mRNA half-lives (Supplementary Figure S1A).

Measuring pre-mRNA and mRNA expression using intronic and exonic probe sets

As a model system for studying strategies of mammalian transcriptional responses to extracellular signals, we used the human mammary epithelial MCF10A cell line. Within 8 h of stimulation with EGF, MCF10A cells develop a migratory and invasive phenotype that requires *de novo* transcription of pro-migratory genes (Amit *et al*, 2007; Katz *et al*, 2007). Importantly, under our experimental conditions MCF10A cells remain viable for over 10 days, but do not proliferate, thus precluding the confounding effect of mRNA dilution due to cell division (see Materials and methods).

We extracted total RNA from biological triplicates at 7 time points following stimulation, hybridized these samples to 21 Affymetrix GeneChip Human Exon 1.0 ST Arrays, and measured, for the first time, transcriptome-wide dynamics of mRNA *and* pre-mRNA. The key to this advance was using signals from exonic and intronic probe sets (PS), both present on these arrays (Figure 2A). Owing to the instantaneous degradation of introns that were spliced out of pre-mRNA (Singh and Padgett, 2009), the signal of intronic PS reflects the amount of the respective pre-mRNA (validated for example for the *VCL* gene by comparing Figures 2D and 3C). Exons, on the other hand, are common to both pre-mRNA and mature mRNA. Because mature mRNA abundance vastly exceeds that of pre-mRNA (the ratio of their steady-state levels equals the ratio of their half-lives), the signals of exonic PS are dominated by mRNA.

We classified each PS as interrogating exons or introns, by combining annotation-based criteria with constraints on signal quality and intensity (see Supplementary information). Properly weighted intronic readings were used to assess, at each time point, the gene-level pre-mRNA expression, while gene-level mRNA expression levels were computed by combining signals from the gene's exonic PS. Of note, our strategy of using signals from intronic PS to measure changes in pre-mRNA expression may be applicable to most multicellular organisms: Analysis of the number of intron-containing genes and the size of their introns revealed that the majority of genes, from *C. elegans* to human, contain introns that are large enough to be interrogated by one or more PS (Supplementary Figure S2). The main limitation in this methodology may reside in (i) the paucity of intronic PS in existing microarray platforms and (ii) difficulty to detect gene-level intronic FC above noise level in less abundant transcripts. Analysis of three biological replicates of each time point permitted us to reliably detect gene-level exonic FC for about 8000 genes. In nearly half of those, a sufficient number of intronic PS (see Materials and methods) were present and exhibited a signal clearly exceeding noise level, thus allowing definition of gene-level intronic FC values (see Supplementary information and Supplementary Figure S3).

Genome-wide time-dependent pre-mRNA and mRNA transcriptional responses to EGF stimulation

By measuring the FC of pre-mRNAs (introns) and mRNAs (exons) at 7 time points, we identified 441 transcriptionally

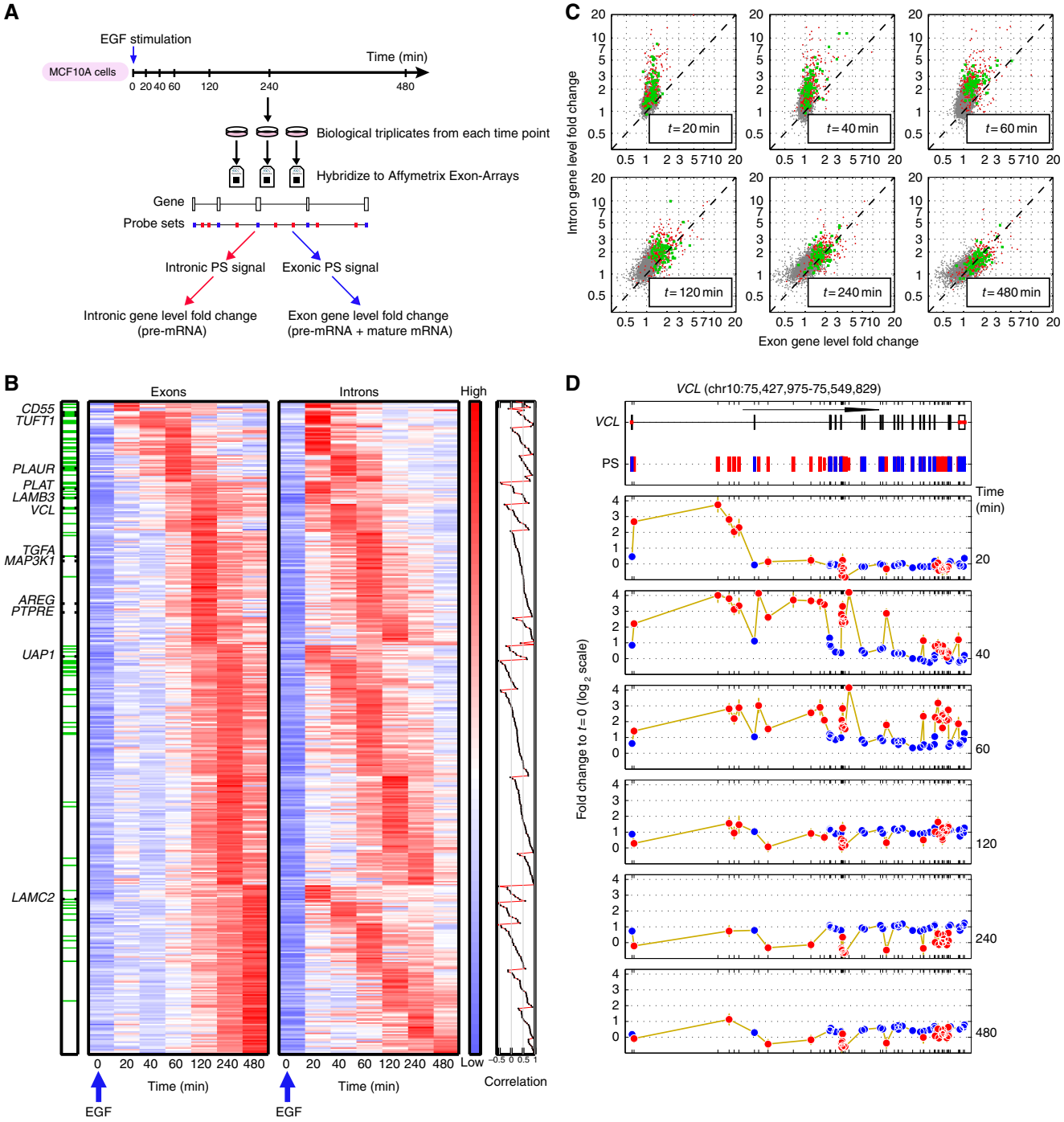


Figure 2 EGF induces distinct transcriptional dynamics of pre-mRNAs and mRNAs in MCF10A cells. (**A**) Experimental outline: MCF10A mammary epithelial cells were stimulated with EGF for the indicated intervals and RNA was hybridized to Affymetrix Exon Arrays. Annotation and signal intensity-based filtering was performed to define probe sets (PS) interrogating exons (blue marks) and introns (red). Fold changes (FCs) of intronic and exonic PS were weighed separately to define the pre-mRNA and mRNA FC for each gene. (**B**) Pre-mRNA and mRNA profiles of 441 transcriptionally induced genes. The left (right) heatmap displays mRNA (pre-mRNA) expression, as defined by the gene-level exon (intron) FC profiles; genes are grouped according to the peak time of their mRNA expression and within each such group ordered according to the peak time of their pre-mRNA FC. Finally, within each subgroup the transcripts were ordered according to the correlation between the pre-mRNA and mRNA profiles (rightmost panel). The order of genes is the same in both heatmaps. Note that the log-transformed expression values of each row were centered and normalized (separately for pre-mRNA and mRNA, owing to their different dynamic range). Hence, the log-FC values at $t=0$ are not uniformly 0. Green lines on the bar on the left indicate genes with production overshoot (i.e., the maximal FC of intronic PS exceeded that of exonic PS by >2-fold). Genes chosen for detailed validation (Figure 3) are indicated. (**C**) Scatter plots comparing gene-level intron (pre-mRNA) and exon (mRNA) FC during EGF stimulation. Genes with mRNA FC$\leqslant$1.5 over the whole time course (grey dots), transcriptionally induced genes with production overshoot (green dots) or without (red dots) are shown. Note the similarity of mRNA FC of overshooting and non-overshooting genes. (**D**) Space-time description of probe-level FC of an overshooting gene. Upper panel: Genomic organization of the vinculin (*VCL*) gene (122 kb). Arrow length corresponds to 50 kb. Positions of exonic (blue) and intronic (red) PS are indicated. Lower panels: FC (log2 scale) of each PS with respect to its baseline value is shown for the time course outlined in (A). Only PS with present calls in all replicates of the respective time points compared are shown. Error bars (in gold) represent standard deviations. Note that intronic and exonic PS display different behavior and dynamic range.

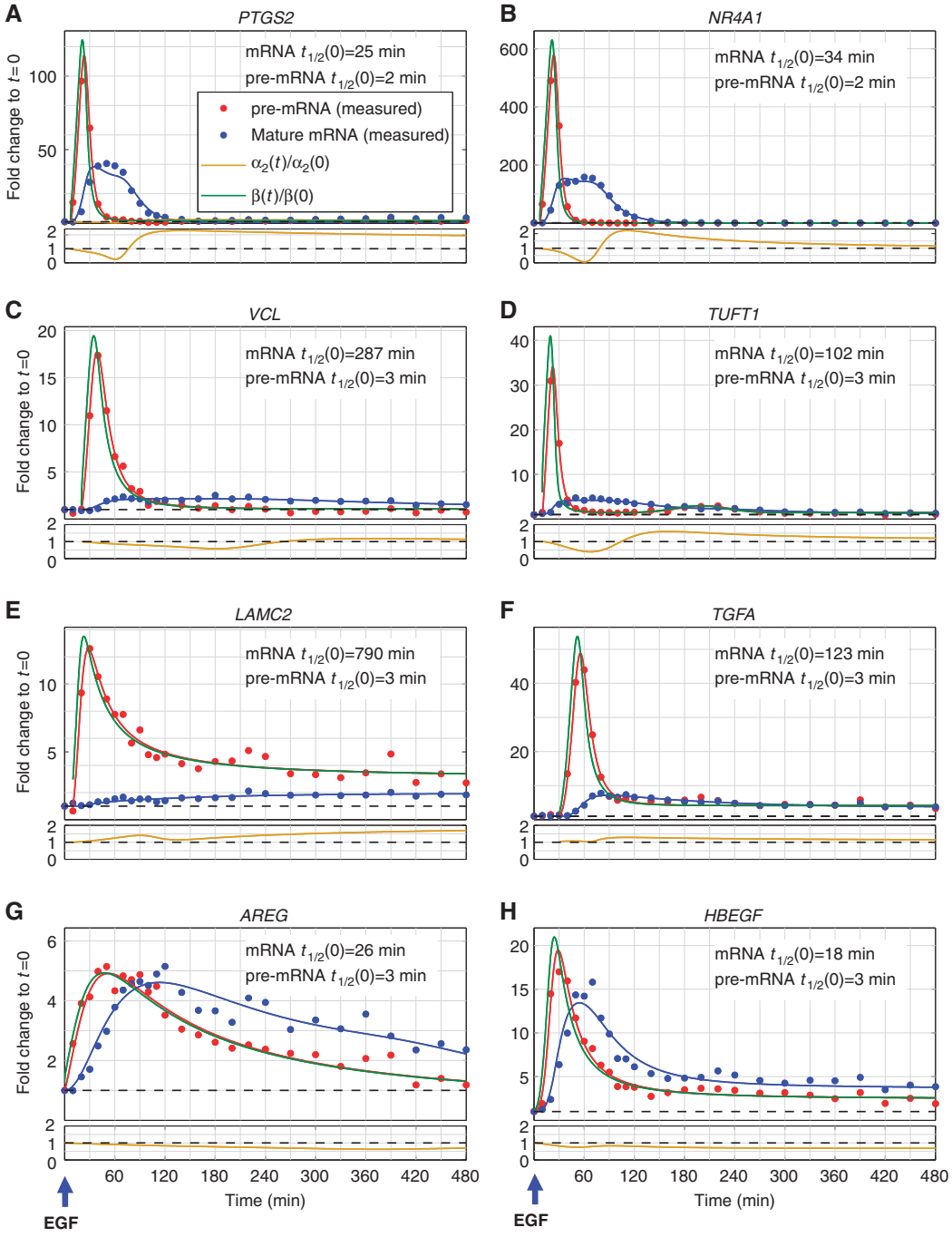


Figure 3 Genes with distinct mRNA expression profiles exhibit production overshoot. Real-time quantitative PCR (qPCR) measurements of expression FC profiles of selected genes with production overshoot (**A**–**F**) or no production overshoot (**G**, **H**), in a high-resolution time course following stimulation of MCF10A cells with EGF. Dots represent averages of technical triplicates, solid lines represent best fit (see Supplementary information) of pre-mRNA (red) and mature mRNA (blue). Production profiles (green lines) and degradation (gold) were inferred from these measurements. Half-lives ($t_{1/2}$) of pre-mRNAs were measured as described in the text, whereas those of mRNA were estimated from fit to the data (see Materials and methods).

induced genes (with maximum FC exceeding 2.1 for pre-mRNA and 1.5 for mRNA, see Supplementary information for choice of cutoffs). Figure 2B displays the time-dependent FC profiles of these genes. Genes were first grouped according to the peak time of their mRNA FC; members of each group were then internally ordered according to the peak time of their pre-mRNA. Finally, each subgroup of transcripts that shared both mRNA and pre-mRNA peak times was sorted by the correlation between the two profiles (see right bar on Figure 2B and Supplementary information). As expected, the onset and peak expression time of introns typically preceded that of exons (Figure 2B; Supplementary Figure S1B–E). Surprisingly,

most genes exhibited markedly different temporal profiles of mRNA and pre-mRNA expression (Figure 2B, right bar). For instance, the mRNA of *LAMC2* peaked 8 h after EGF stimulation, while its pre-mRNA had reached its maximum FC already after 20 to 30 min (Figures 2B and 3E). Strikingly, genes with similar peak times of mRNA expression exhibited clearly distinct pre-mRNA dynamics. Intuitively, genes with similar pre-mRNA profiles would be more likely to share common *cis*-regulatory elements compared with genes exhibiting similar mRNA profiles. Our analyses of genes sharing either similar pre-mRNA or mRNA profiles, however, did not reveal a significant enrichment in known transcription factor-binding DNA sequence motifs (data not shown). Importantly, by 20 min after EGF stimulus, pre-mRNA levels of most upregulated mRNAs had already increased (see heatmaps in Figure 2B), suggesting that the initial regulation of these genes occurs via the primary transcriptional response, whereas the amplitude and duration of induction of these genes may be differentially shaped by newly synthesized transcriptional and post-transcriptional regulators.

Complementary time-dependent amplitude information revealed that most of the induced genes exhibited much higher and narrower peaks of pre-mRNA FC compared with mRNA FC (Figures 2C and 3). Whereas the early FC of many pre-mRNAs tremendously exceeded that of their respective mRNAs, their FCs at later time points were comparable. In particular, the peak pre-mRNA FC of 18% (79/441) of the induced genes exceeded the peak FC of their corresponding mRNAs by >2-fold (green lines on the bar in Figure 2B and green dots in Figure 2C). We refer to these as genes exhibiting *production overshoot*. The role of a brief pulse of production which significantly exceeds the eventual mRNA FC is to accelerate the rise of mRNA abundance and thereby shorten response time. Production overshoot is the strategy of choice to overcome the 'dynamic barrier' imposed by long mRNA life times on the 'classical' step rise in production described above (Figure 1A). Production overshoot is not synonymous (see Figure 2C) to having either higher long-time mRNA FC or to a transient peak of mRNA abundance.

Genes exhibiting production overshoot are mostly PRGs

Production overshoot was typically found in genes exhibiting significantly earlier onset and peak of production, as well as earlier mRNA peak time compared with genes without overshooting introns. For example, within the first 40 min of EGF stimulation pre-mRNA levels had increased by >2-fold in 96% of the overshooting versus 45% of the non-overshooting genes (Supplementary Figure S1B–E). Such rapid kinetic characteristics identify genes with production overshoot as PRGs. To examine this notion, we considered a data set of 98 genes identified as PRGs in glioblastoma cells stimulated with the platelet-derived growth factor in the presence of cycloheximide (Tullai *et al*, 2007). Thirteen (13%) of these PRGs were also transcriptionally induced by EGF in our MCF10A cells. Nine out of these thirteen PRGs (69%) exhibited production overshoot in MCF10A cells indicating highly significant overlap ($P<4.0E{-}06$, hypergeometric test). Functional annotation analysis (Huang da *et al*, 2009) of the overshooting genes revealed their significant enrichment by functional categories associated with cell adhesion and motility (Supplementary File 1).

A typical example of production overshoot is shown in Figure 2D for the vinculin (*VCL*) gene, which encodes for a tension sensor localized to focal adhesions. Here, EGF stimulation leads to a transient, 16-FC in expression of intronic PS, reached 40–60 min after stimulation and rapidly fading thereafter. In contrast, exonic PS display delayed and much more subtle changes, exhibiting a peak FC of 2 between 120 and 240 min followed by a slow decrease. Note that in large genes such as *VCL* (122 kb) or *ITGA2* (105 kb, Supplementary Figure S4), the space-time dependence of intronic PS reflects the propagation of the initial wave of polymerases sweeping along the gene (Singh and Padgett, 2009; Wada *et al*, 2009; Figure 2D).

Inferring transcript production and degradation dynamics from pre-mRNA and mRNA profiles

To obtain time-dependent transcript production and mRNA degradation rates, we measured pre-mRNA and mRNA expression profiles by qPCR (see Materials and methods and Supplementary Figure S5) in a frequently sampled time course experiment. We selected 12 upregulated genes (indicated in Figure 2B), 9 with production overshoot and, for comparison, 3 without production overshoot. Additionally, we profiled *MYC* as well as three dramatically induced IEGs: *HBEGF*, *NR4A1* and *PTGS2* (IEGs would otherwise have been underrepresented in our analyses due to lack of a sufficient number of intronic PS, see Materials and methods). Results for downregulated genes are described later (Figure 4).

Overshooting pre-mRNA levels occurred in genes with diverse mRNA profiles, such as exhibiting transient or sustained induction, high or low level fold change, and early or late peaks (Figure 3; Supplementary Figure S5). Genes, for which exon arrays did not reveal an overshoot in intron profiles, did not exhibit overshoot in qPCR measurements either (Figure 3G and H; Supplementary Figure S5), validating the array results. Transcript production and degradation dynamics (Figure 3, $\hat{\beta}(t)$ and $\hat{\alpha}_2(t)$ are shown in green and gold curves, respectively) were inferred from measured pre-mRNA and mRNA time courses (as described in Materials and methods).

Importantly, these functions were obtained without any interference with the transcriptional response.

In an independent experiment, we measured the values of the pre-mRNA conversion coefficient α_1, in the pre-stimulus steady state as well as following stimulation, using transcription arrest during short temporal intervals (see below, Supplementary Figure S6A). These values were found to be highly similar across all transcripts analyzed. Since the relevant time scale of conversion is a few minutes, the problems associated with the severe disruption of the cells (caused by transcription arrest) have not yet taken effect and the results are reliable. By contrast, the pre-stimulus mRNA degradation coefficients $\alpha_2(0)$ could reliably be determined without transcription arrest, by fitting our data (see Materials

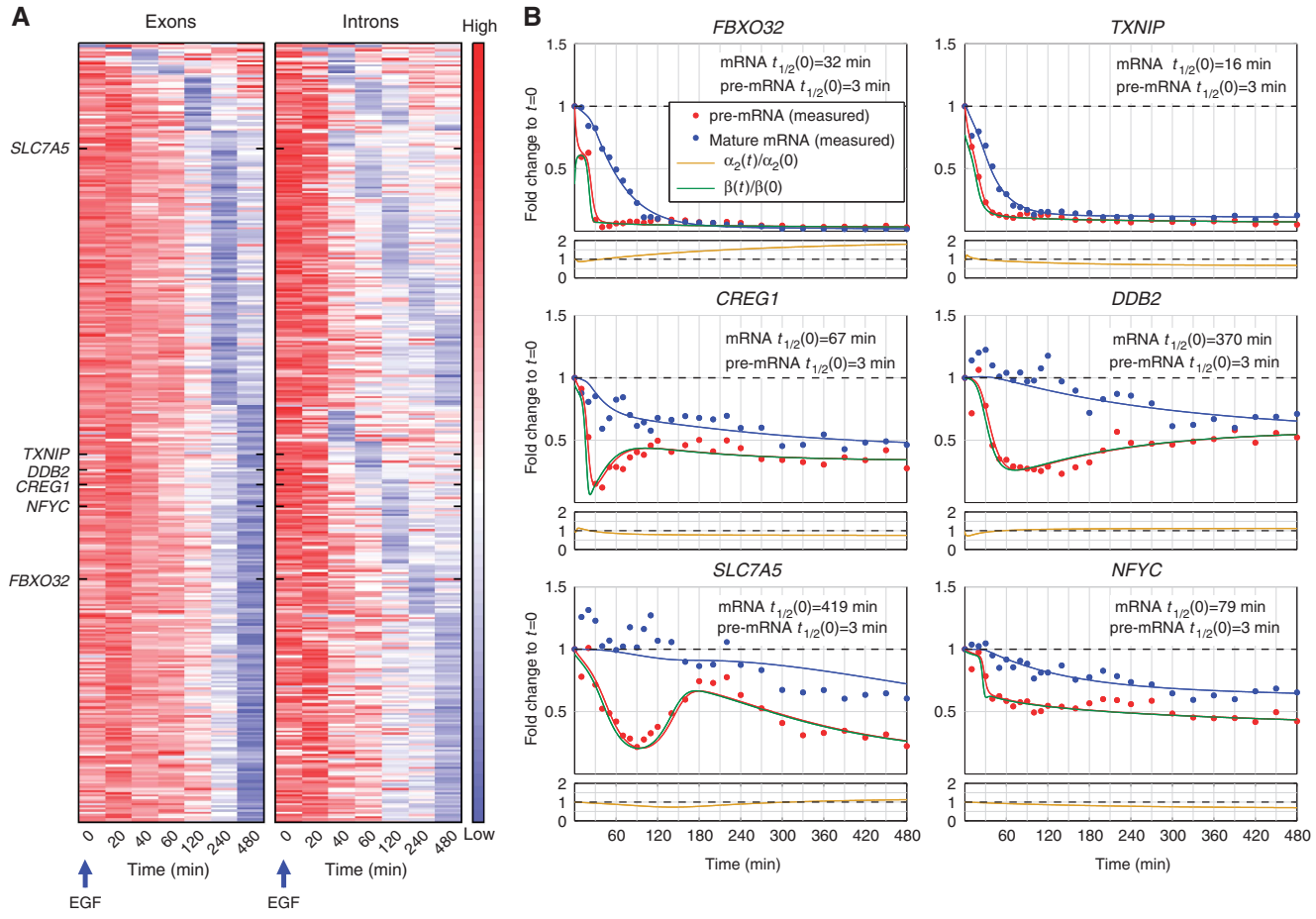


Figure 4 EGF driven downregulation of pre-mRNAs and mRNAs in MCF10A cells. (**A**) Pre-mRNA and mRNA profiles of 364 transcriptionally repressed genes. The left heatmap displays mRNA expression as defined by the gene-level exon FC profiles; genes are grouped first according to the time of their lowest mRNA FC, and within each such group, they are ordered according to the time of their lowest pre-mRNA FC, as reflected by intronic FC (right heatmap). The order of the genes is the same in both heatmaps. Names of genes that were chosen for detailed validation are indicated. Note that mRNA profiles are very poor indicators of shutdown of pre-mRNA production. Heatmaps were normalized as in Figure 2B. (**B**) qPCR measurements of expression FC profiles of selected downregulated genes in a high-resolution time course following stimulation of MCF10A cells with EGF. Dots represent averages of technical triplicates, solid lines represent best fit (see Materials and methods) of pre-mRNA (red) and mature mRNA (blue). Production profiles (green lines), mRNA half-lives and degradation (gold) were inferred from these measurements as described in Materials and methods.

and methods), and exhibited wide variation between transcripts. Analysis of the statistical error of the inferred profiles (described in Supplementary information) shows that the observed temporal variations are statistically significant.

Profiles shown in Figure 3 reveal that the overshoot in pre-mRNA levels reflects an overshoot in the production rate. In contrast to pre-mRNA levels, mRNA levels often peaked while production had already decreased much below its peak level, or even returned to its initial level (e.g., see *NR4A1* and *VCL* in Figure 3B and C). Some of the genes also exhibited a second, albeit much smaller, peak of production (e.g., *TUFT1* in Figure 3D and *CD55* in Supplementary Figure S5).

mRNA degradation coefficients of several genes also exhibited non-monotonic behavior, including stabilization at long times (e.g., *AREG* and *HBEGF*; Figure 3G and H), demonstrating that cells delicately balance degradation with time-varying production to establish the desired temporal mRNA profiles. The inferred EGF-induced changes of mRNA stability at long times were qualitatively confirmed by experiments employing transcription arrest (Supplementary Figure S6B). The inferred (slightly <2-fold) stabilization in *AREG* and *HBEGF* (see Figure 3G and H) was confirmed, while the actual measured degradation times were significantly longer than the ones inferred without transcription arrest (pre-stimulus for *AREG* 69 versus 26 min, for *HBEGF* 28 versus 18 min), indicating the extent to which such methods are comparable.

Downregulated genes

An important aspect of transcriptional response involves downregulation of many transcripts. Due to the typically long mRNA half-lives, the timing of transcript production shutdown is hard to determine from mRNA data. We identified 364 downregulated genes characterized by exon FC <0.7 and/or intron FC <0.5. Time-dependent mRNA and pre-mRNA temporal profiles of these genes emphasize again the mismatch between mRNA and production profiles (Figure 4A).

Transcription dynamics of six genes, measured at high temporal resolution (Figure 4B), revealed non-trivial production dynamics, mostly involving pronounced early shut-down followed by partial restoration of production. Importantly, in our experimental system, mRNA downregulation always involved a decrease of production (versus relying on mRNA degradation only, e.g., Figure 1F).

Conversion with a constant coefficient dominates pre-mRNA outflux

Inference of production and degradation profiles from our model was derived assuming that the pre-mRNA to mRNA conversion coefficient α_1 does not vary with time. An alternative scenario, of conversion slowing down (due to either saturation of the pre-mRNA processing machinery (Patel *et al*, 2002; Pessa *et al*, 2006; Singh and Padgett, 2009) or to prolonged nuclear retention of incompletely processed transcripts (Prasanth *et al*, 2005)), could lead to pile-up of pre-mRNA and interfere with correct estimation of production rates. To exclude this possibility, we directly calculated the pre-mRNA conversion coefficients (α_1) from measurements of pre-mRNA decay following transcription arrest. This was done for both unstimulated MCF10A cells and EGF-stimulated cells, yielding very similar decay rates of pre-mRNA under the two conditions (Figure 5A; Supplementary Figure S6A), confirming our modeling assumptions: indeed, α_1 remains unchanged over a large range of pre-mRNA concentrations.

Another explicit assumption of our model is that the half-life of pre-mRNA is mainly determined by conversion to mRNA, rather than by degradation. Conceivably, a significant level of pre-mRNA degradation by the nuclear RNAi machinery, that changes in the course of the stimulus (Bousquet-Antonelli *et al*, 2000; Hargreaves *et al*, 2009; Guang *et al*, 2010), could also contribute to the increased pre-mRNA levels, which we

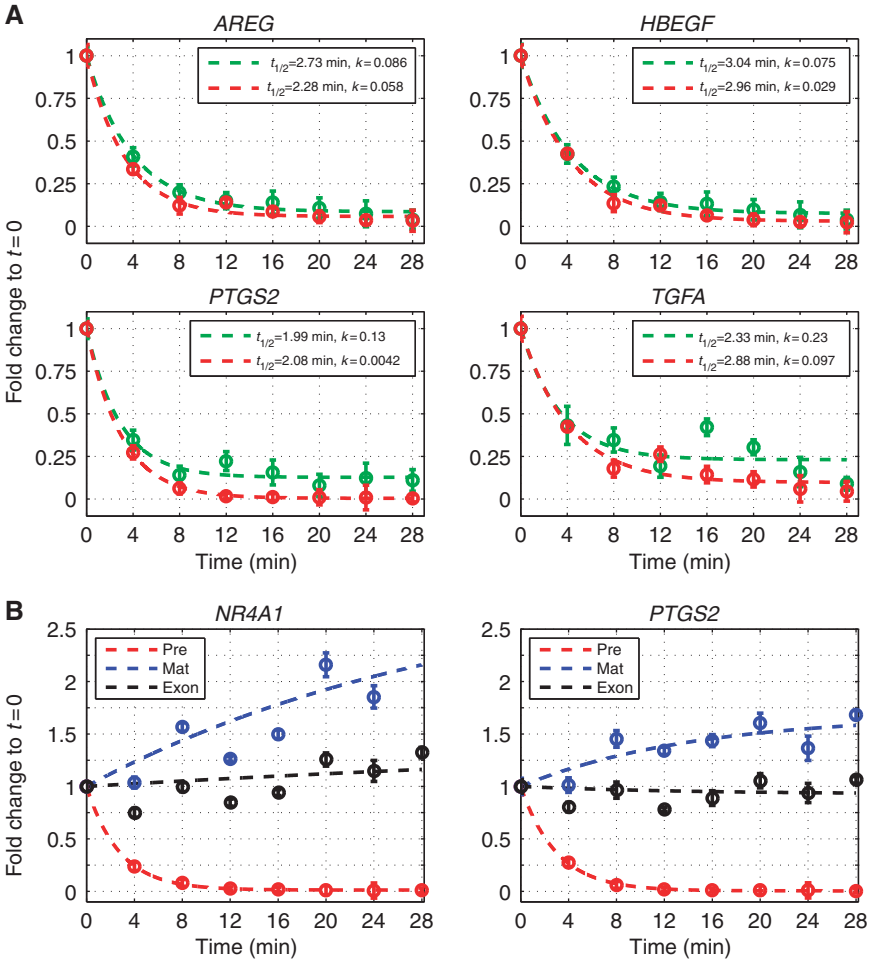


Figure 5 Validation of model assumptions. (**A**) Pre-mRNA decay in unstimulated (green curves) and EGF-stimulated (red curves) MCF10A cells: transcription was arrested by applying ActD (10 μM) to mock-stimulated MCF10A cells or to cells near the peak of pre-mRNA induction after pre-incubation with EGF (20 ng/ml) for 20 min (*AREG*, *HBEGF*, *PTGS2*) or 40 min (*TGFA*), respectively. RNA was isolated at the indicated time points and qPCR was used to measure pre-mRNA decay. Note that the pre-mRNA conversion coefficient is unaffected by EGF stimulation. (**B**) Dynamics of pre-, exon and mature mRNA after transcription arrest at the peak of pre-mRNA induction: MCF10A cells were stimulated with EGF for 20 min before applying ActD as described in (A). Pre-mRNA (red), exonic (black) and mature mRNA (blue) were measured by qPCR. Note that pre-mRNA outflux is predominantly determined by conversion to mRNA. All symbols and error bars denote mean values and standard deviations of technical triplicates from one out of three independently repeated experiments. RNA half-lives ($t_{1/2}$) and asymptotic constants (k) were calculated and curves were fit as described in Materials and methods.

attributed to sharply increased production. Because for most genes the ratio of pre-mRNA to mRNA is very low—even close to the peak of pre-mRNA—assessing the quantitative impact of pre-mRNA degradative processes is quite challenging. To accomplish this, we analyzed two highly induced genes (*PTGS2* and *NR4A1*), both exhibiting overshooting pre-mRNA. We arrested transcription near the point of maximum pre-mRNA accumulation (at which the pre-mRNA to mRNA ratio was high enough), and measured pre-mRNA, mRNA and exon FCs by qPCR. Since the copy number of exons is preserved by pre-mRNA conversion, but not preserved when pre-mRNA is degraded, loss of exons would be indicative of non-negligible pre-mRNA degradation. We found that conversion was the predominant process of pre-mRNA depletion; the rapid decay of pre-mRNA was accompanied by nearly constant exon abundance and an increase of mRNA (Figure 5B).

Production overshoot accelerates the induction of mRNAs

Our results demonstrate that the production dynamics of many induced genes do not exhibit a simple step increase (Figure 1A); rather, production overshoot is the strategy of choice (Figure 1C) and is most likely employed in order to accelerate mRNA response. Therefore, we compared the temporal profiles of mRNAs of induced transcripts with and without pre-mRNA overshoot. To eliminate the confounding effect of mRNA half-life, we first grouped genes into sets of similar half-lives, and then compared the dynamics of transcripts within each set. mRNA half-life was estimated for all induced genes using the fitting procedure as described in Materials and methods.

Beyond revealing the expected faster response and earlier peak times for short-lived versus long-lived mRNAs (Shalem *et al*, 2008; Hao and Baltimore, 2009; Elkon *et al*, 2010), our analyses demonstrated clearly shorter response times for genes exhibiting production overshoot across the entire range of mRNA half-lives (Figure 6). Similar results were obtained, when information on mRNA half-lives from another data source, which employed biosynthetic labeling (genomic run-on) methods to calculate mRNA half-lives in different cell lines, was used (Friedel *et al*, 2009; Supplementary Figure S7).

Production overshoot is a generic operational strategy enabling accelerated response

To evaluate the generality of production overshoot in transcriptional responses, we studied two additional very disparate cell types subjected to different types of stimuli, namely LPS-stimulated murine bone marrow-derived primary dendritic cells (DCs), and human embryonic stem cells (hESCs) exposed to retinoic acid (RA). Binding of LPS to Toll-like receptor-4 at the plasma membrane instigates signaling cascades that culminate in the activation of transcription factors such as NFκB, which induce an inflammatory response and maturation of DCs (Medzhitov and Horng, 2009; Supplementary Figure S8A). By contrast, RA diffuses through the cell membrane and forms a transcription regulatory complex with the RA receptor, which promotes hESC differentiation toward ectodermal (i.e., neuronal) fates within several days (Supplementary Figure S8B; Boyer *et al*, 2005).

In both systems, the temporal profiles of pre-mRNAs and mRNAs during the very initial phase of the transcriptional response to stimulation revealed transcripts exhibiting production overshoot, along with several known changes in mRNA stability accompanying transcriptional induction (Hao and Baltimore, 2009; Figure 7; Supplementary Figure S9). The observed occurrence of production overshoot in the three very different stimulated systems described herein demonstrates that this operational strategy is a general characteristic of mammalian transcriptional responses to extracellular cues.

Discussion

Transcriptional responses of cells to external signals involve orchestrated changes in transcript production and degradation rates. These changes are often assumed to be simple shifts of production and degradation to new constant values. By combining mathematical modeling with measured temporal profiles of pre-mRNA and mRNA abundance in response to extracellular stimuli, we obtained, with unprecedented resolution, the time-dependent behavior of the processes that control transcript induction, that is, production and degradation. We discovered and quantified a most prominent feature of the pre-mRNA profiles of many genes, reflecting a transient pulse of production of previously unanticipated high dynamic range. Thus, production FC can exhibit a large overshoot over eventual mRNA FC. Moreover, genes with similar mRNA peak times exhibit a wide variation in production peak times, suggesting that the expression of such genes may be governed by different regulatory elements. Most EGF-induced genes initiate their production within the first hour after stimulation (Supplementary Figure S1B and C, blue curves). Two recently published studies addressed related issues. The first used global run-on and sequencing (Gro-seq) in a breast cancer cell line after estradiol stimulation (Hah *et al*, 2011), while the second used pulse labeling by 4sU in LPS-stimulated DCs. In agreement with our results, the first study reported that a large fraction of the transcriptional response was executed very rapidly, while the ensuing change of mRNA abundance was delayed by intervals that varied between 1 and 3 h. In contrast, the second study reported that changes in total mRNA lagged behind the corresponding changes in newly synthesized RNA by a fairly uniform interval of 15–30 min (Rabani *et al*, 2011).

Beyond production, the temporal profiles of mRNA induction are shaped also by degradation (Barenco *et al*, 2009; Hao and Baltimore, 2009). Our quantitative assessment properly weighs the relative contributions of production and degradation to the dynamics of transcriptional responses.

Production overshoot is instrumental, together with time-dependent degradation, in shaping precisely transient expression profiles, to bring a transcript to the right level at the right time and for the right duration. We found that most genes exhibiting production overshoot are *bona fide* PRGs and are enriched by executors of the phenotypic response to stimulation. These genes encode for relatively long-lived mRNAs,

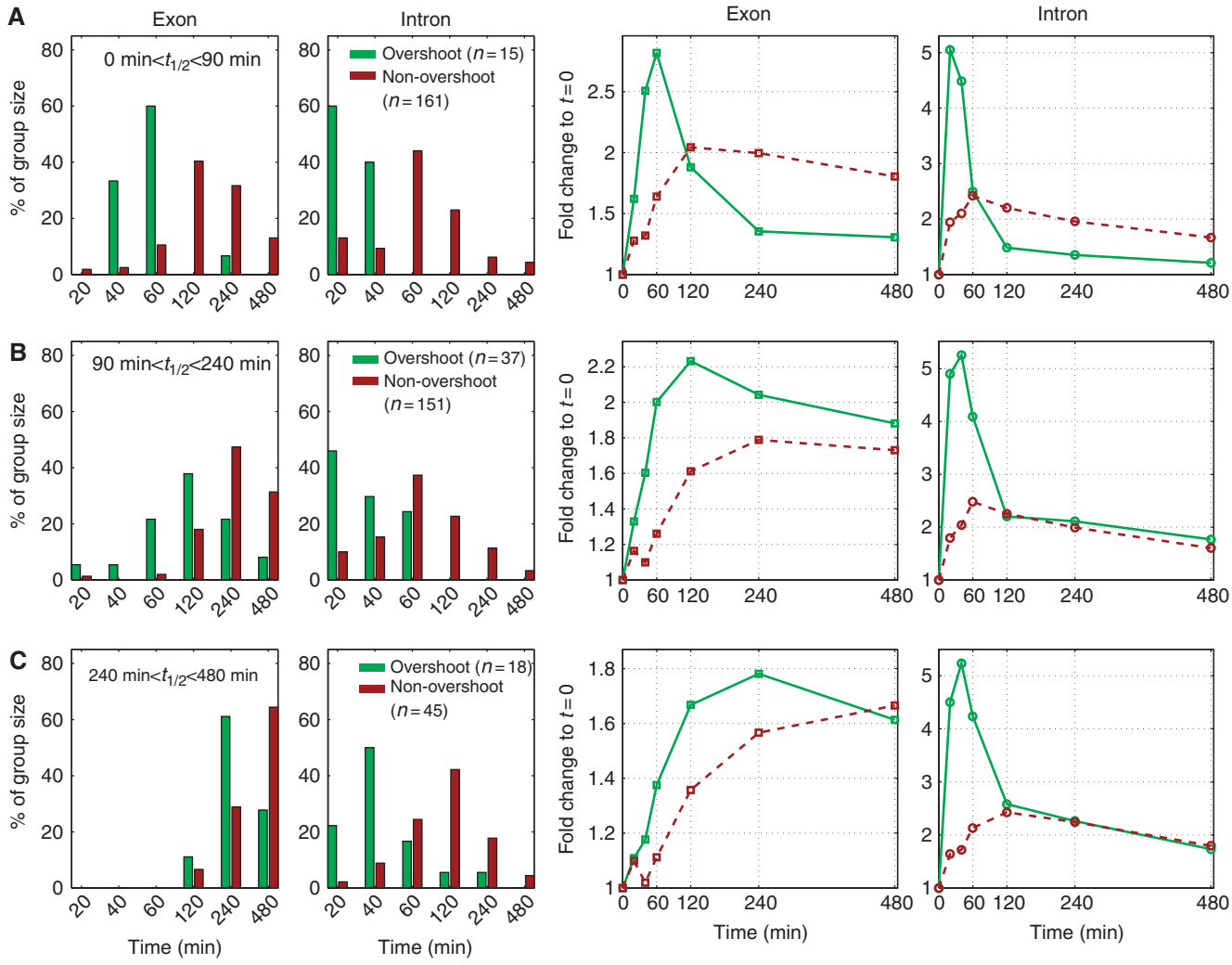


Figure 6 Pre-mRNA production overshoot accelerates the response time of mRNAs. Genes that were transcriptionally induced by EGF stimulation of MCF10A cells (as described in legend to Figure 2) were grouped according to their mRNA half-lives (extracted from our measured expression profiles, as described in Materials and methods). Results are shown for three representative groups with half-lives between (**A**) 0–90 min (**B**) 90–240 min and (**C**) 240–480 min, respectively. The two columns on the left depict the percentage of genes, whose exons and introns peaked at the indicated time points, green for overshooting and brown for non-overshooting genes. The two columns on the right present the mean temporal expression FC profiles for exons and introns, for the overshooting (green) and non-overshooting (brown) genes. Note that within each group of genes with similar half-lives we observe earlier mRNA peak times for the genes with pre-mRNA production overshoot.

whose levels may be maintained at an economically favorable low production rate in the absence of stimuli.

Whether the overshoot is a digital all-or-none phenomenon, tuned only by the fraction of cells responding to stimulation (Podtschaske *et al*, 2007) or by the duration of production (Lahav *et al*, 2004; Suter *et al*, 2011), is an open question. Alternatively, it may comprise a graded transcriptional response in individual cells, demonstrated, for example, by independent NFκB binding to adjacent regulatory sites (Giorgetti *et al*, 2010). The underlying molecular mechanisms may also include cooperative action of transcription factors, regulation of the number of polymerases traveling across the gene, and regulation of polymerase processivity (Baugh *et al*, 2009; Wada *et al*, 2009). The very rapid offset kinetics of overshooting pre-mRNA production, which often precedes significant changes in mRNA abundance, suggests that if cells use feedback to induce this decrease, its mechanism likely relies on sensing the levels of pre-mRNA or nuclear mRNA, rather than of cytoplasmic mRNA or protein. An attractive and likely alternative to feedback is a mechanism of 'prewired control'—production is designed to have a transient pulse-like profile. In different cellular model systems, we found that different genes exhibit production overshoot at different times, suggesting that the molecular mechanisms governing production overshoot may be gene and context dependent, and will require additional studies. Our findings are an essential prerequisite for such studies.

We believe that our demonstration of how similar mRNA profiles can be generated by very different production profiles constitutes an important conceptual advance. The insights gained by our modeling approach and experiments provide a consistent framework toward quantitative elucidation of operational and molecular strategies used by cells to regulate transcriptional responses to extracellular signals.

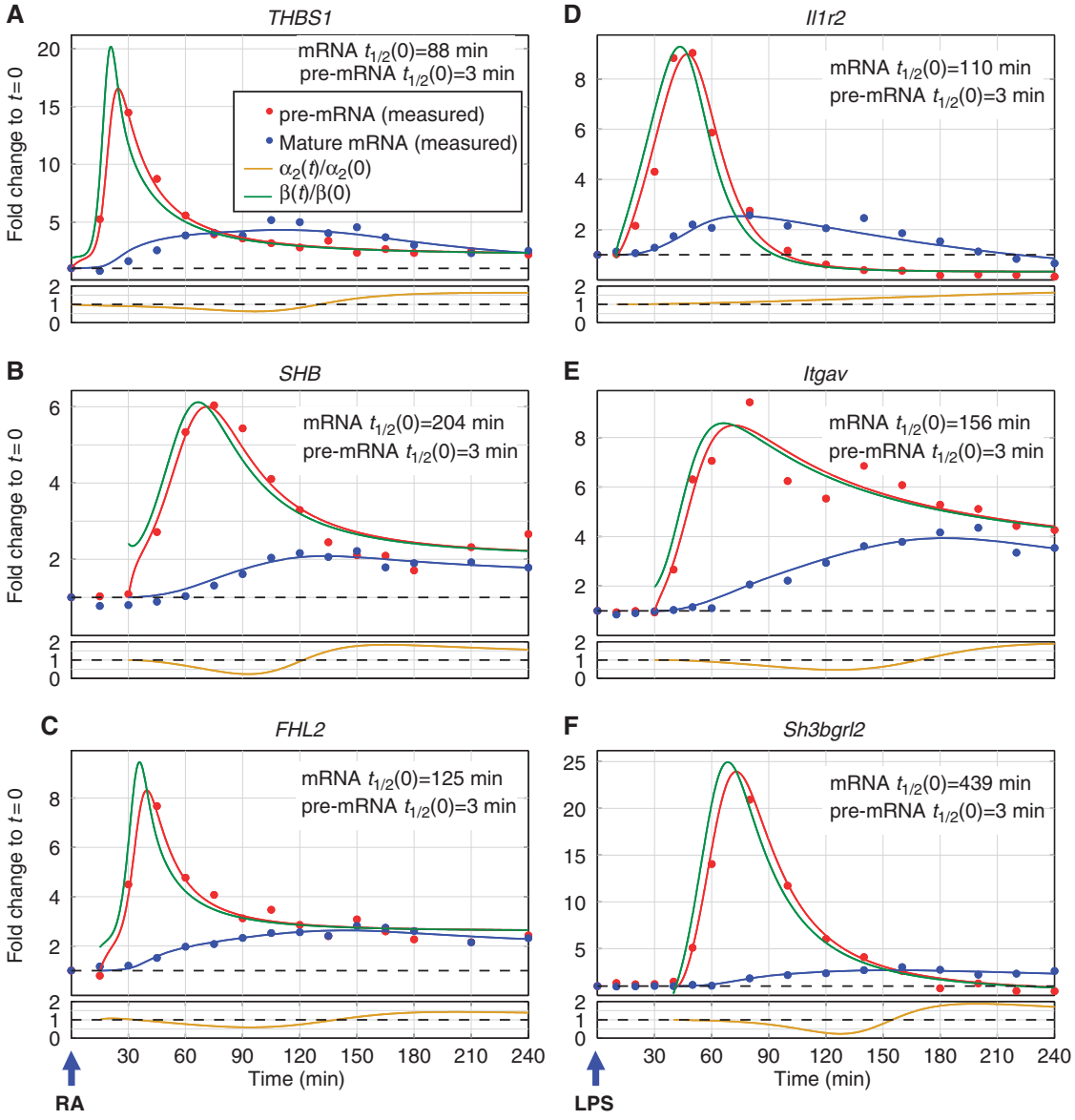


Figure 7 Pre-mRNA production overshoot is a general operational strategy in mammalian transcriptional networks. The plots in (**A**–**F**) show qPCR measurements of pre-mRNA (red) and mature mRNA (blue) expression profiles for selected genes exhibiting production overshoot. Dots represent averages of technical triplicate measurements and solid lines represent best fit (see Supplementary information). Production and degradation profiles were inferred from these measurements as described in Materials and methods, and their time-dependent profiles are indicated by green and gold curves, respectively. (A–C) Production overshoot in retinoic acid (RA)-stimulated human embryonic stem cells. (D–F) Production overshoot in lipopolysaccharide (LPS)-stimulated primary murine dendritic cells.

Materials and methods

Cell culture and stimulation

MCF10A cells were cultured as described in Katz *et al* (2007) and stimulated with EGF (20 ng/ml) for the indicated intervals. Bone marrow-derived murine DCs from C57BL/6 mice were prepared as previously described (Amit *et al*, 2009) and stimulated with LPS (100 ng/ml). H9 human ESCs were cultured as described in Supplementary information and stimulated with all-trans RA (1 μM). For RNA decay experiments, ActD was used as indicated to arrest transcription.

RNA isolation and microarray hybridization

RNA was isolated using the *PerfectPure RNA Cultured Cell Kit* (5 Prime, Hamburg, Germany) including DNAse 1 digestion and rRNA depleted. Samples were processed as recommended by the microarray manufacturer and hybridized to *Affymetrix GeneChip Human Exon 1.0 ST* arrays. Microarray data are deposited in Gene Expression Omnibus (accession number GSE24391).

Real-time qPCR

For qPCR of pre-mRNA and mRNA, respectively, forward primers were positioned in the second intron and exon, respectively, and a shared reverse primer was positioned in the third constitutive exon. For amplification of exon mRNA, primer pairs were positioned in the third exon. All qPCRs were performed using *Power SYBR Green PCR Master Mix* (Applied Biosystems, Foster City, CA) on a *7900HT Fast Real Time PCR System* platform (Applied Biosystems) along with non-template controls, melt curve analysis and cDNA dilution series. Detailed

methods and curve fitting are described in Supplementary information and primer sequences are listed in Supplementary File 1.

Microarray data analyses

Affymetrix Expression Console (parameters: Annotation confidence—full, Summarization method—iterPLIER include DABG, Background—PM-GCBG, Normalization method—none) was used, followed by normalization of all arrays together using a Lowess multi-array algorithm and signal-dependent noise estimation, as described in Zeisel *et al* (2010). Annotation and signal-based information was used to define exonic and intronic PS. Intronic and exonic PS were used to calculate the gene-level FC of pre-mRNAs and mRNAs. A detailed description of microarray data processing is given in Supplementary information and Supplementary Figures S3 and S10.

Inference of transcript production and degradation profiles

Pre-mRNA, mRNA and exon FC were measured for selected genes using qPCR at up to 27 time points. For all measurements, the average of three technical replicates was plotted versus time. In order to infer the production profile, we used Equation (4). A particular 5-parameter functional form $\hat{P}^{\text{fit}}(t)$ (see Supplementary information) was used and the parameters were determined by best fit to the data. The value of the pre-mRNA conversion coefficient was determined for each transcript (see Figure 5A and Supplementary Figure S6A). The time derivative of the fitted function was taken (analytically) and Equation (4) was inverted to yield an analytic form of the production profile,

$$\hat{\beta}(t) = \frac{1}{\alpha_1}\frac{d\hat{P}^{\text{fit}}}{dt} + \hat{P}^{\text{fit}}$$

Next, for each gene, the pre-stimulus mRNA degradation rate $\alpha_2(0)$ and degradation profile $\hat{\alpha}_2(t)$ were inferred in a nested iterative procedure. A particular 7-parameter form was assumed for this function (see Supplementary information). The main iterative procedure is the following: start with an initial guess $\hat{\alpha}_2(t)=1$, optimize (as described below) the pre-stimulus coefficient to get $\alpha_2^{(1)}(0)$; using this value, optimize the degradation profile to get a new $\hat{\alpha}_2(t)$, optimize again $\alpha_2(0)$ to obtain the next optimal $\alpha_2(0)$ and iterate until convergence.

We describe here the optimization procedure for the case of fixed $\alpha_2(0)$. We took an initial guess $\hat{\alpha}_2^{(0)}(t)$, substituted it for $\hat{\alpha}_2(t)$ in the right hand side of Equation (5), together with the analytic fitted function $\hat{P}^{\text{fit}}(t)$. Next, we integrated Equation (5) numerically to yield an approximate $\hat{M}^{(0)}(t)$. The least squares deviation of this function from the measured data was calculated, and new values for the parameters were set to define $\hat{\alpha}_2^{(1)}(t)$. The process was iterated until convergence to a function $\hat{\alpha}_2(t)$ that gave the best fit to the measured $\hat{M}(t)$. A similar iterative process was used to optimize the pre-stimulus degradation coefficient for a given degradation profile ($\hat{\alpha}_2(t)$).

The error of the inferred functions, estimated on the basis of the qPCR measurement noise, is explained in Supplementary information and presented, for a single transcript, *NR4A1*, in Supplementary Figure S11.

Functions used for fit

The functions used to fit the pre-mRNA and degradation FC profiles, $\hat{P}^{\text{fit}}(t)$ and $\hat{\alpha}_2(t)$ are described in detail in Supplementary information.

Supplementary information

Supplementary information is available at the *Molecular Systems Biology* website (www.nature.com/msb).

Acknowledgements

We thank Naama Barkai, Uri Alon, Tzachi Pilpel and Moshe Oren for fruitful discussions. This work was supported by the Leir Charitable Foundation, a Weizmann-Mario Negri collaborative research grant, the MD Moross Institute for Cancer Research, Arresto Biosciences, the European Commission, the Dr Miriam and Sheldon Adelson Medical Research Foundation, the Kekst Family Institute for Medical Genetics, the Kirk Center for Childhood Cancer and Immunological Disorders, the Women's Health Research Center funded by the Bennett-Pritzker Endowment Fund, the Marvelle Koffler Program for Breast Cancer Research, the German Research Foundation (DIP), National Cancer Institute (CA72981), and the Flight Attendants Medical Research Institute (FAMRI). ED is the incumbent of the Henri J Leir Professorial Chair, YY is the incumbent of the Harold and Zelda Goldenberg Professorial Chair, GR holds the Djerassi Chair in Oncology at the Tel-Aviv University and WJK is supported by a PhD Track for Specialist MDs fellowship of The Linda and Michael Jacobs Charitable Trust.

Author contributions: AZ, WJK, YY and ED designed and analyzed experiments and conceived mathematical modeling; AZ, WJK, NM, JMT, RK and JJ-H performed experiments; SJ, YS, GR, YY and ED provided funding and experimental platforms. AZ, WJK, YS, YY and ED wrote the manuscript.

Conflict of interest

The authors declare that they have no conflict of interest.

References

Alon U (2007) *An Introduction to Systems Biology—Designs Principles of Biological Circuits*. Boca Raton, FL: Chapman and Hall/CRC

Amit I, Citri A, Shay T, Lu Y, Katz M, Zhang F, Tarcic G, Siwak D, Lahad J, Jacob-Hirsch J, Amariglio N, Vaisman N, Segal E, Rechavi G, Alon U, Mills GB, Domany E, Yarden Y (2007) A module of negative feedback regulators defines growth factor signaling. *Nat Genet* **39:** 503–512

Amit I, Garber M, Chevrier N, Leite AP, Donner Y, Eisenhaure T, Guttman M, Grenier JK, Li W, Zuk O, Schubert LA, Birditt B, Shay T, Goren A, Zhang X, Smith Z, Deering R, McDonald RC, Cabili M, Bernstein BE *et al* (2009) Unbiased reconstruction of a mammalian transcriptional network mediating pathogen responses. *Science* **326:** 257–263

Avraham R, Sas-Chen A, Manor O, Steinfeld I, Shalgi R, Tarcic G, Bossel N, Zeisel A, Amit I, Zwang Y, Enerly E, Russnes HG, Biagioni F, Mottolese M, Strano S, Blandino G, Borresen-Dale AL, Pilpel Y, Yakhini Z, Segal E *et al* (2010) EGF decreases the abundance of microRNAs that restrain oncogenic transcription factors. *Sci Signal* **3:** ra43

Barenco M, Brewer D, Papouli E, Tomescu D, Callard R, Stark J, Hubank M (2009) Dissection of a complex transcriptional response using genome-wide transcriptional modelling. *Mol Syst Biol* **5:** 327

Baugh LR, Demodena J, Sternberg PW (2009) RNA Pol II accumulates at promoters of growth genes during developmental arrest. *Science* **324:** 92–94

Bousquet-Antonelli C, Presutti C, Tollervey D (2000) Identification of a regulated pathway for nuclear pre-mRNA turnover. *Cell* **102:** 765–775

Boyer LA, Lee TI, Cole MF, Johnstone SE, Levine SS, Zucker JP, Guenther MG, Kumar RM, Murray HL, Jenner RG, Gifford DK, Melton DA, Jaenisch R, Young RA (2005) Core transcriptional regulatory circuitry in human embryonic stem cells. *Cell* **122:** 947–956

Byun JS, Wong MM, Cui W, Idelman G, Li Q, De Siervi A, Bilke S, Haggerty CM, Player A, Wang YH, Thirman MJ, Kaberlein JJ, Petrovas C, Koup RA, Longo D, Ozato K, Gardner K (2009) Dynamic bookmarking of primary response genes by p300 and RNA polymerase II complexes. *Proc Natl Acad Sci USA* **106:** 19286–19291

Carballo E, Lai WS, Blackshear PJ (1998) Feedback inhibition of macrophage tumor necrosis factor-alpha production by tristetraprolin. *Science* **281:** 1001–1005

Cochran BH, Reffel AC, Stiles CD (1983) Molecular cloning of gene sequences regulated by platelet-derived growth factor. *Cell* **33:** 939–947

Darzacq X, Shav-Tal Y, de Turris V, Brody Y, Shenoy SM, Phair RD, Singer RH (2007) *In vivo* dynamics of RNA polymerase II transcription. *Nat Struct Mol Biol* **14:** 796–806

Elkon R, Zlotorynski E, Zeller KI, Agami R (2010) Major role for mRNA stability in shaping the kinetics of gene induction. *BMC Genomics* **11:** 259

Fan J, Yang X, Wang W, Wood III WH, Becker KG, Gorospe M (2002) Global analysis of stress-regulated mRNA turnover by using cDNA arrays. *Proc Natl Acad Sci USA* **99:** 10611–10616

Friedel CC, Dolken L, Ruzsics Z, Koszinowski UH, Zimmer R (2009) Conserved principles of mammalian transcriptional regulation revealed by RNA half-life. *Nucleic Acids Res* **37:** e115

Garneau NL, Wilusz J, Wilusz CJ (2007) The highways and byways of mRNA decay. *Nat Rev Mol Cell Biol* **8:** 113–126

Giorgetti L, Siggers T, Tiana G, Caprara G, Notarbartolo S, Corona T, Pasparakis M, Milani P, Bulyk ML, Natoli G (2010) Noncooperative interactions between transcription factors and clustered DNA binding sites enable graded transcriptional responses to environmental inputs. *Mol Cell* **37:** 418–428

Gorini L, Maas WK (1957) The potential for the formation of a biosynthetic enzyme in Escherichia coli. *Biochim Biophys Acta* **25:** 208–209

Guang S, Bochner AF, Burkhart KB, Burton N, Pavelec DM, Kennedy S (2010) Small regulatory RNAs inhibit RNA polymerase II during the elongation phase of transcription. *Nature* **465:** 1097–1101

Hager GL, McNally JG, Misteli T (2009) Transcription dynamics. *Mol Cell* **35:** 741–753

Hah N, Danko CG, Core L, Waterfall JJ, Siepel A, Lis JT, Kraus WL (2011) A rapid, extensive, and transient transcriptional response to estrogen signaling in breast cancer cells. *Cell* **145:** 622–634

Hao S, Baltimore D (2009) The stability of mRNA influences the temporal order of the induction of genes encoding inflammatory molecules. *Nat Immunol* **10:** 281–288

Hargreaves DC, Horng T, Medzhitov R (2009) Control of inducible gene expression by signal-dependent transcriptional elongation. *Cell* **138:** 129–145

Herschman HR (1991) Primary response genes induced by growth factors and tumor promoters. *Annu Rev Biochem* **60:** 281–319

Hirose Y, Manley JL (1998) RNA polymerase II is an essential mRNA polyadenylation factor. *Nature* **395:** 93–96

Huang da W, Sherman BT, Lempicki RA (2009) Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. *Nat Protoc* **4:** 44–57

Iyer VR, Eisen MB, Ross DT, Schuler G, Moore T, Lee JC, Trent JM, Staudt LM, Hudson Jr J, Boguski MS, Lashkari D, Shalon D, Botstein D, Brown PO (1999) The transcriptional program in the response of human fibroblasts to serum. *Science* **283:** 83–87

Katz M, Amit I, Citri A, Shay T, Carvalho S, Lavi S, Milanezi F, Lyass L, Amariglio N, Jacob-Hirsch J, Ben-Chetrit N, Tarcic G, Lindzen M, Avraham R, Liao YC, Trusk P, Lyass A, Rechavi G, Spector NL, Lo SH *et al* (2007) A reciprocal tensin-3-cten switch mediates EGF-driven mammary cell migration. *Nat Cell Biol* **9:** 961–969

Lahav G, Rosenfeld N, Sigal A, Geva-Zatorsky N, Levine AJ, Elowitz MB, Alon U (2004) Dynamics of the p53-Mdm2 feedback loop in individual cells. *Nat Genet* **36:** 147–150

Larson DR, Zenklusen D, Wu B, Chao JA, Singer RH (2011) Real-time observation of transcription initiation and elongation on an endogenous yeast gene. *Science* **332:** 475–478

Lau LF, Nathans D (1987) Expression of a set of growth-related immediate early genes in BALB/c 3T3 cells: coordinate regulation with c-fos or c-myc. *Proc Natl Acad Sci USA* **84:** 1182–1186

Mangan S, Alon U (2003) Structure and function of the feed-forward loop network motif. *Proc Natl Acad Sci USA* **100:** 11980–11985

Mangan S, Zaslaver A, Alon U (2003) The coherent feedforward loop serves as a sign-sensitive delay element in transcription networks. *J Mol Biol* **334:** 197–204

Medzhitov R, Horng T (2009) Transcriptional control of the inflammatory response. *Nat Rev Immunol* **9:** 692–703

Miller C, Schwalb B, Maier K, Schulz D, Dumcke S, Zacher B, Mayer A, Sydow J, Marcinowski L, Dolken L, Martin DE, Tresch A, Cramer P (2011) Dynamic transcriptome analysis measures rates of mRNA synthesis and decay in yeast. *Mol Syst Biol* **7:** 458

Orphanides G, Reinberg D (2002) A unified theory of gene expression. *Cell* **108:** 439–451

Patel AA, McCarthy M, Steitz JA (2002) The splicing of U12-type introns can be a rate-limiting step in gene expression. *EMBO J* **21:** 3804–3815

Pessa HK, Ruokolainen A, Frilander MJ (2006) The abundance of the spliceosomal snRNPs is not limiting the splicing of U12-type introns. *RNA* **12:** 1883–1892

Podtschaske M, Benary U, Zwinger S, Hofer T, Radbruch A, Baumgrass R (2007) Digital NFATc2 activation per cell transforms graded T cell receptor activation into an all-or-none IL-2 expression. *PLoS ONE* **2:** e935

Prasanth KV, Prasanth SG, Xuan Z, Hearn S, Freier SM, Bennett CF, Zhang MQ, Spector DL (2005) Regulating gene expression through RNA nuclear retention. *Cell* **123:** 249–263

Proudfoot NJ, Furger A, Dye MJ (2002) Integrating mRNA processing with transcription. *Cell* **108:** 501–512

Rabani M, Levin JZ, Fan L, Adiconis X, Raychowdhury R, Garber M, Gnirke A, Nusbaum C, Hacohen N, Friedman N, Amit I, Regev A (2011) Metabolic labeling of RNA uncovers principles of RNA production and degradation dynamics in mammalian cells. *Nat Biotechnol* **29:** 436–442

Ramirez-Carrozzi VR, Braas D, Bhatt DM, Cheng CS, Hong C, Doty KR, Black JC, Hoffmann A, Carey M, Smale ST (2009) A unifying model for the selective regulation of inducible transcription by CpG islands and nucleosome remodeling. *Cell* **138:** 114–128

Rosenfeld N, Elowitz MB, Alon U (2002) Negative autoregulation speeds the response times of transcription networks. *J Mol Biol* **323:** 785–793

Sassone-Corsi P, Sisson JC, Verma IM (1988) Transcriptional autoregulation of the proto-oncogene fos. *Nature* **334:** 314–319

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

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

Singh J, Padgett RA (2009) Rates of *in situ* transcription and splicing in large human genes. *Nat Struct Mol Biol* **16:** 1128–1133

Suter DM, Molina N, Gatfield D, Schneider K, Schibler U, Naef F (2011) Mammalian genes are transcribed with widely different bursting kinetics. *Science* **332:** 472–474

Tullai JW, Schaffer ME, Mullenbrock S, Sholder G, Kasif S, Cooper GM (2007) Immediate-early and delayed primary response genes are distinct in function and genomic architecture. *J Biol Chem* **282:** 23981–23995

Wada Y, Ohta Y, Xu M, Tsutsumi S, Minami T, Inoue K, Komura D, Kitakami J, Oshida N, Papantonis A, Izumi A, Kobayashi M, Meguro H, Kanki Y, Mimura I, Yamamoto K, Mataki C, Hamakubo T, Shirahige K, Aburatani H *et al* (2009) A wave of nascent transcription on activated human genes. *Proc Natl Acad Sci USA* **106:** 18357–18361

Wang Z, Zang C, Cui K, Schones DE, Barski A, Peng W, Zhao K (2009) Genome-wide mapping of HATs and HDACs reveals distinct functions in active and inactive genes. *Cell* **138:** 1019–1031

Zeisel A, Amir A, Kostler WJ, Domany E (2010) Intensity dependent estimation of noise in microarrays improves detection of differentially expressed genes. *BMC Bioinformatics* **11:** 400

Supplementary Material for: Coupled pre-mRNA and mRNA dynamics unveil operational strategies underlying transcriptional responses to stimuli

Amit Zeisel[1,6], Wolfgang J Kostler[2,6,7], Natali Molotski[3], Jonathan M Tsai[2], Rita Krauthgamer[4], Jasmine Jacob-Hirsch[5], Gideon Rechavi[5], Yoav Soen[1,3], Steffen Jung[4], Yosef Yarden[2,*], Eytan Domany[1,*]

August 18, 2011

[1] Department of Physics of Complex Systems, Weizmann Institute of Science, Rehovot, Israel.

[2] Department of Biological Regulation, Weizmann Institute of Science,Rehovot, Israel.

[3] Department of Biological Chemistry, Weizmann Institute of Science, Rehovot, Israel.

[4] Department of Immunology, Weizmann Institute of Science, Rehovot, Israel.

[5] Sheba Cancer Research Center, The Chaim Sheba Medical Center and Sackler School of Medicine, Tel Aviv University, Tel Aviv, Israel.

[6] These authors contributed equally to this work.

[7] Present address: Department of Medicine 1 & Comprehensive Cancer Center, Medical University of Vienna, Vienna, Austria.

[*] Corresponding authors.

1 Supplementary Information

1.1 Cell culture

The non-transformed human mammary epithelial MCF10A cell line was obtained from ATCC (Manassas, Virginia, USA) and maintained in full medium composed of 1:1 DMEM:F12(HAM), 10 μg/ml insulin, 100IU/0.1mg per ml Penicillin/Streptomycin (all from Biological Industries, Beit Haemek, Israel), 0.5 μg/ml hydrocortisone (Calbiochem, La Jolla, CA, USA), 5% heat-inactivated horse serum (ATCC), 0.1 μg/ml cholera toxin and 10 ng/ml EGF (all from Sigma-Aldrich, Rehovot, Israel). For experiments, subconfluent, low passage MCF10A cells were harvested by careful trypsinization, washed, and plated for 24 hours in full medium without EGF. Cells were then starved for 24 hours in DMEM:F12(HAM) containing 0.9% horse serum and antibiotics, followed by stimulation with EGF (20 ng/ml final concentration) for the indicated intervals. All time course experiments involving MCF10A cells were repeated thrice in independent experiments. For isolation of naive murine dendritic cells (DCs), bone marrow mononuclear cells were harvested from femora and tibiae of 6 week-old female C57BL/6 mice and enriched by lysis of erythrocytes. Mononuclear cells were grown

on non-adherent plastic Petri dishes for 10 days in RPMI medium (GIBCO, Carlsbad, CA) supplemented with 10% FBS (HyClone), 1% L-glutamine, 1% MEM non-essential amino acids, 1% sodium pyruvate, 1% penicillin/streptomycin (all from Biological Industries), 2-beta-mercaptoethanol (50μM, Merck), and GM-CSF (15 ng/ml; Preprotech, Rocky Hill, NJ, USA). Floating cells were transferred to new petri dishes on day 6, resulting in a primary dendritic cell culture (>80% purity, Supplementary Figure S8). For stimulation, LPS (from Sigma-Aldrich) was added to a final concentration of 100 ng/ml for the indicated intervals. Data shown represent the results of technical triplicate measurements from a single experiment. Maturation of CD11c+ DCs was verified by flow cytometric analysis staining for CD40, 0 and 24 hours after LPS stimulation (Supplementary Figure S8A). All animals were maintained under specific pathogenfree conditions and were handled according to protocols approved by the Weizmann Institute Animal Care Committee in compliance with international guidelines. Human embryonic stem cells (hESCs; line H9) were kindly provided by Prof. J. Itskovitz (Technion Israel Institute of Technology). Cells were grown on feeder layers of irradiated mouse embryonic fibroblasts and maintained in DMEM:F12(HAM) (Biological Industries) hESC media containing 15% Knockout Serum Replacement, 1% Penicillin-Streptomycin-Glutamine, 1% MEM-Non Essential Amino Acids, 0.2% 2-mercaptoethanol (all from Gibco) and 8ng/ml bFGF (Preprotech). For experiments, hESCs (passage 51) were harvested using 0.1% collagenase (Gibco), transferred to dishes coated with (1:20X) GF-reduced Matrigel (BD Biosciences) in hESC media supplemented 1:1 with conditioned medium (i.e., medium which had been used for 24h to grow irradiated mouse embryonic fibroblasts). Following 2 days in culture, the medium was replaced with hESC medium lacking bFGF, with or without all-trans retinoic acid (RA; Sigma-Aldrich, final concentration 1μM) for the indicated intervals. Data shown represent the results of technical triplicate measurements from one experiment that was repeated independently.

1.2 RNA isolation and microarray hybridization

Total RNA was isolated using the PerfectPure RNA Cultured Cell Kit (5 Prime, Hamburg, Germany) including DNAse 1 digestion according to the manufacturer's instructions. RNA integrity, content and purity were assessed by agarose gel electrophoresis and spectrophotometrically (Nanodrop ND-1000, Wilmington, Delaware), respectively. For oligonucleotide microarray hybridization, 1.0 μg total RNA from MCF10A cells stimulated with EGF for 0, 20, 40, 60, 120, 240 or 480 minutes (3 separate biological replicates of each time point) were rRNA depleted (RiboMinus Human/Mouse Transcriptome Isolation Kit, Invitrogen) and RNA integrity confirmed using a Bioanalyzer (2100 Bioanalyzer, Agilent, Palo Alto, CA). Subsequent steps of sense strand cDNA synthesis using the fragmentation and biotin labeling were performed using reagents provided by the array manufacturer and samples were hybridized to Affymetrix GeneChip Human Exon 1.0 ST arrays (Affymetrix, Santa Clara, CA, USA). Microarray data are deposited in Gene Expression Omnibus (accession number GSE24391).

1.3 Real-time quantitative PCR (qPCR)

Sequence information from the UCSC genome browser (http://genome.ucsc.edu) was used to set up reactions for pre- and mature mRNA amplification employing forward primers in the second intron and exon (or further downstream if mandated by sequence constraints), respectively, and a shared reverse primer positioned in the

third constitutive exon of expressed transcripts. This primer design ensures similar reaction conditions and detection of pre-mRNA production close to the time of its onset, whilst enabling preferential amplification of processively elongating transcripts (Singh & Padgett, 2009; Wada et al, 2009). For estimation of the pre- to mature RNA ratio, additional pairs amplifying the third constitutive exon (or a further downstream exon) were designed. The general primer design strategy is shown in Supplementary Figure S5A. Primer3 (http://fokker.wi.mit.edu/primer3/input.htm) was used for designing primers. Primer pairs were checked for their specificity in silico (http://genome.ucsc.edu) and primer sequences are listed in Supplementary Table 1. For qPCR, 2μg RNA (for stimulation experiments) or 10μl of isolated RNA (for RNA decay experiments) were reverse transcribed using random primers and the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems, Foster City, CA). qPCR reactions were run in a 2-step cycling protocol (initial denaturation at 95C for 10 minutes, then 40 cycles of 95C for 15 seconds and 60C for 1 minute, followed by acquisition of dissociation curves) using Power SYBR Green PCR Master Mix (Applied Biosystems) on a 384-well 7900HT Fast Real Time PCR System platform (Applied Biosystems). Every reaction included all time points for all analyzed transcript species of the same gene on the same plate, along with non-template controls, melt curve analysis to ensure the uniqueness of amplification products, and cDNA dilution series to calculate the efficacy of primers. Primer efficacy was calculated and only primer pairs with an efficacy >0.9 were used. No RT controls were performed to exclude DNA contamination. Human glyceraldehyde-3-phosphate dehydrogenase (GAPDH, MCF10A cells), mouse GAPDH (DCs) or TATA box binding protein (TBP, hESCs) were used as endogenous controls for normalization in stimulation experiments. All data shown are mean fold changes (and additional standard deviations in Supplementary Figures S5B, S8, and S9) from technical triplicates calculated using the comparative Ct method (Pfaffl, 2001; Schmittgen & Livak, 2008).

1.4 RNA decay experiments

For short-term pre-mRNA decay experiments (including those also measuring exons and mature mRNA), Actinomycin D (ActD, from Sigma-Aldrich) was added to unstimulated or EGF-stimulated cells at time 0. To achieve rapid ActD uptake and maximum Pol II inhibition we used a final concentration of $10\mu g/ml$ ActD (Perry & Kelley, 1970). Samples were harvested by direct lysis every 4 minutes up to 28 minutes. For long-term mature mRNA decay measurements, unstimulated or EGF-stimulated cells were pre-incubated for 30 minutes with Actinomycin D (5μg/ml final concentration) to ensure near complete conversion of pre-existing pre-mRNA and lysed 0, 30, 60, 90, 120, 240, or 480 minutes later. Equal efficacy of RNA extraction was ensured by spiking lysates with B. subtilis RNAs (Affymetrix GeneChip Eukaryotic Poly-A RNA Control Kit) and subsequent qPCR for prokaryotic Dap transcripts. Extracted RNA was additionally spiked with 0.1μg of RNA from Saccharomyces cervisiae (kind gift of Naama Barkai) and equal loading in subsequent qPCR experiments confirmed by amplification of the yeast *Act1* mRNA. All RNA decay experiments were performed thrice and data from technical triplicates of one representative experiment are shown.

Data analysis

1.5 Microarray preprocessing

Affymetrix GeneChip Human Exon 1.0 ST Arrays were scanned to produce CEL files for the 21 arrays used. To extract the data at the probeset (PS) level, Affymetrix Expression Console software was used with the following parameters: Annotation confidence - full, Summarization method - iterPLIER include DABG, Background - PM-GCBG, Normalization method - none. A Lowess multi-array algorithm (Ballman et al, 2004) was used for normalization of all arrays together and intensity dependent noise estimation was performed as previously described (Zeisel et al, 2010). Out of 1.3 million probe sets (PS) of the array, 666,916 PS, which could be assigned to known genes and in which all probes were uniquely aligned with perfect match (crosshyb_type= 1), were used for further analyses. To control for the signal quality, we used FDR of 5% (separately for each array) on the Affymetrix PS p-values, and a signal threshold. The signal of a PS at a specific time point is called "present", if the following two conditions are satisfied: (i) all three PS p-values from biological repeats of a time point pass the 5% FDR, and (ii) the average signal of the three repeats is higher than 3 (log2 scale). Only present PS were considered for further analyses. The typical number of PS that passed all these criteria was 270,000. The distribution functions of measured signal intensity from intronic and exonic PS that passed the PS-level quality filter described above are shown in Supplementary Figure S3A. Intronic PS are expressed at lower, but still detectable levels.

1.6 Definition of PS as interrogating exonic or intronic transcript regions

The UCSC gene model, which contains 66803 full-length transcripts, was used as a reference gene model (downloaded from http://genome.ucsc.edu/). At the first step, PS which interrogate intronic regions in all known transcripts were assigned as constitutive intronic, while the rest were classified as putative exonic. Since introns' abundance is low, their signal is naturally weaker. We used this idea to purify the list of exonic PS: For each time point the signal distribution of constitutive introns was estimated and then each putative exonic PS was tested with the null hypotheses of its signal coming from the signal distribution of intronic PS. Only exonic PS for which the signal was found to be significantly higher than the introns' signal distribution at two or more adjacent time points were considered as exonic PS, while the rest were omitted from further analysis.

1.7 Determining gene level exon/intron fold change from the genes PS

After classifying each PS as exonic or intronic, we integrated their fold change (FC) to calculate two numbers, which summarized the gene-level exonic or intronic FC. The FC of each PS was calculated with respect to its value at $t = 0$ using the average expression of the triplicate measurements at each time point. Next, we defined two alternative exon gene level FC values: the 50th and top 90th percentiles from the FC values of exonic PS over the gene. In the same manner, we defined the two alternative intron gene level FC values, as the 50th and 90th percentiles from intronic PS. Note that these gene-level intronic (exonic) FC values were calculated only for genes with at least 4 "present" PS for introns (exons). If the number of present intronic PS was less than 10, we used the PS with the highest FC value (instead of the top 90%). We decided to use the FC based on 50th percentile for exonic gene-level FC and the 90th percentile for the intronic FC, for the following reasons. First - heterogeneity of the intronic signals. The FC of different intronic PS of the same

gene many times exhibit large biologically driven heterogeneity. The most obvious reason is that during the initial phase of transcription, while the first wave of elongating Pol II moves down the gene at a speed of approximately 4kb per minute, the introns behind the front may have already attained a high FC while those downstream have not. The times for which this is the case depend on the genomic size of the transcribed gene and may range from less than a minute to more than one hour (Singh & Padgett, 2009; Tennyson et al, 1995; Wada et al, 2009). Thus, even for average-sized genes of 27kb (Venter et al, 2001), the last intron will likely be induced with a lag of more than 6 minutes. As can be seen in Figure 2D, for early times (less than 40 minutes after stimulation) using the median (50th percentile) FC to represent the intron gene level FC would imply that transcription of the gene has not started yet, which is incorrect. We are keen on catching rapid changes in production as early as possible. Thus, because pre-RNA FC reflects production which, in turn, is measured by the introns, we use the intron FC 90th percentile to define the gene-level FC. As opposed to introns, the different exonic PS of the same gene tend to have a more homogeneous behavior, because the exonic signal is dominated by the pool of preexisting mRNA rather than the exons of the currently transcribed pre-mRNA. Therefore the exonic gene-wide FC can be represented by the less fluctuative 50% percentile (we did check the effect of using 90th percentile also for exons and it did not alter the picture significantly, Supplementary Figure S10). We decided to use the 50th percentile for exons also for another reason: the set of PS that we identified as exonic may be contaminated by intronic PS that were misclassified as exonic (e.g., putative exonic PS which have a high signal, but whose dynamics resemble those of intronic PS). Using for the exons FC the 90th percentile could be sensitive to this non-exonic contamination. Overall, 7968 genes were identified as expressed on the basis of their exonic signals (i.e. they passed our filter for expressed genes, as described above). For 3422 of these genes, the intronic PS also passed the filters of reliable FC detection. We assessed the extent to which two factors, the number of intronic PS (placed on the Affymetrix exon array) and the mRNA signal intensity, govern our ability to detect gene-level intronic FC. We present histograms for both groups of genes those with detected and undetected gene-level intronic FC. In Supplementary Figure S3B we show a histogram of the number of genes according to their number of intronic PS. We observe that genes with detected intron gene-level FC had significantly more intronic PS than genes of the complementary group, which had only very few intronic PS. Our ability to detect intronic FC depends, albeit more weakly, also on the intensity of the exonic (mRNA) signal, as shown in Supplementary Figure S3C; genes with undetected intronic FC have somewhat lower mRNA intensities. We analyzed the noise associated with determining the exonic and intronic gene-level FC, using a previously described method of noise estimation (Zeisel et al, 2010). The results are shown in Supplementary Figure S3D, where the intensity dependent standard deviation (std) is plotted, versus log-intensity, for the measured exonic and intronic gene-level signals, separately for all the time points measured. When the log-intensity is above a threshold of 4, the std is less than 0.18 for all time points, transcript types (exons/introns) and intensities. Hence the std of any measured FC (remember these are log(base 2)-transformed variables) is less than $0.18\sqrt{2} = 0.25$. Thus, a 2 FC is 4 std above a FC of 1.

1.8 Definitions of production overshoot and induction by EGF

We defined a gene as exhibiting production overshoot, if its maximal (over the time course) gene-level intronic FC (using the 90th percentile) exceeded that of the maximal gene-level exonic FC (using the 50th percentile) by more than 2-fold. Supplementary Figure S10 shows that this definition of production overshoot is robust

against using different choices of the exonic and intronic gene-level FC. We define a gene as induced by EGF stimulation if it has a "present" signal along the whole time course, its maximal gene-level exonic FC (using the 50th percentile) is >1.5 and its maximal gene-level intronic FC (using the 90th percentile) is >2.1. The different thresholds for exon and intron FC were set according to the natural bias introduced by taking the 90th percentile for the introns. This bias between the 50th and 90th percentiles was estimated by the differences between the means of the two intronic maximal FC value distributions, i.e. of the introns, using 50th and using 90th percentile. This difference was about 0.5 (log2 scale); thus, instead of using a threshold of 1.5 on the introns (with 50th percentile) we used the equivalent $2.1 \approx 2^{log_2(1.5)+0.5}$ FC for the intronic gene-level FC at the 90th percentile.

1.9 Estimating correlations between exons and intron profiles in Figure 2B

In the rightmost bar of Figure 2B we present the correlation between the exonic gene-level FC and the intronic gene-level FC. Since our time course is not uniformly sampled (sampling rate between 20 minutes and 4 hours), we linearly interpolated the profiles every 20 minutes to get an equally spaced sampling, and then calculated the Pearson correlation of the two 25-component vectors.

1.10 Functions used to fit the qPCR time course profiles:

All qPCR time course profiles were fitted using the averaged data from three technical replicates each for measuring pre-mRNA, exon, and mRNA FC. Fitting the pre-mRNA FC with $\hat{P}^{(fit)}(t)$: First, for each qPCR profile of measured pre-mRNA FC we determined by inspection an onset time, t_{on}, as the first time point at which the measured fold change exceeded two-fold (i.e. was clearly above the associated noise level). For each measured pre-mRNA a temporal profile of the following functional form was fitted:

$$\hat{P}^{(fit)}(t') = \left\{ \begin{array}{ll} 1 & t' = t - t_{on} \leq 0 \\ f_1(t') + f_2(t') & t' \geq 0 \end{array} \right\} \tag{1}$$

where:

$$f_1(t') = 1 + (a_3 - 1)(1 - e^{-a_4 t'}) \tag{2}$$

$$f_2(t') = a_1 \frac{(a_2 t')^{n_2}}{1 + (a_2 t')^{n_1}} \tag{3}$$

The rationale behind these functions is the following: The first function starts at FC of 1, allowing a non-zero initial slope (corresponding to a jump in production at $t = 0$) and approaching a long-time asymptotic value (of) exponentially, at a rate a_4, also determined by the fitting procedure. Such functions are the solution of eq. (2) for the case of a discontinuous jump of the production rate to a constant value. The second function was chosen to reflect a possible peak of production. For $n_2 > 1$ the function starts at 0 with a vanishing slope, increases to a peak and decreases algebraically to zero at long times (for $n_1 > n_2$). The fit parameter a_1 is the amplitude of the function and the parameter a_2 sets the time scale. The peak position, width and asymptotic decay of the function are governed by a_2, n_1 and n_2. For two genes presented in Figure 3 (*TUFT1*) and 4

(*SLC7A5*), we used an additional function $f_3(t') = k_1 e^{-k_3(t'-k_2)^2}$ in order to fit the observed second pulse in the pre-mRNA profile. Fitting the degradation coefficient profile $\hat{\alpha}_2(t)$: Again we used

$$\hat{\alpha}_2(t') = \left\{ \begin{array}{ll} 1 & t' = t - t_{on} \leq 0 \\ g_1(t') + g_2(t') & t' \geq 0 \end{array} \right\} \tag{4}$$

The same ton was used as above, and the functions g_1, g_2 had the following forms:

$$g_1(t') = 1 + (b_3 - 1)(1 - e^{-b_4 t'}) \tag{5}$$

$$g_2(t') = b_1 \frac{(b_2 t')^{m_2} - (b_2 t')^{m_3}}{1 + (b_2 t')^{n_1}} \tag{6}$$

The rationale behind choosing these functions is along the same lines as described above. The function g_1 represents a simple smooth change of the fold change from 1 to an asymptotic new value. The second function allows additional flexibility; we first tried again a form that allows a single peak (or minimum) and found that in order to obtain a good fit to $\hat{M}(t')$ a functional form that allows a decrease followed by an increase (or the opposite order) was necessary. The fitting procedure is described in the section Inference of transcript production and degradation profiles in the *Materials and Methods* of the main text, and involves iterations to fit $\alpha_2(0)$ and $\hat{\alpha}_2(t')$. The robustness of the fitted function to the measurement noise of the qPCR was estimated and presented in Supplementary Figure S11. To estimate these errorbars we sampled 10 profiles of pre-mRNA and mRNA from the distribution at each time point as estimated from the three replicates (see Supplementary Figure S3D). Next we repeated the fitting procedure (estimating $\hat{P}^{(fit)}(t'), \alpha_2(0)$ and $\hat{\alpha}_2^{(fit)}(t')$) 10 times and presented the profiles as mean ± standard-deviation.

1.11 Curve fitting in RNA decay experiments

To estimate the decay rates (half-lives) of transcript FC after transcription arrest by ActD, we assume that the transcript decay has an exponential form. Since transcription inhibition is not perfect, the asymptotic value after inhibition is not zero. Hence, a two-parameter profile was used to fit:

$$F(t) = k + (1 - k)e^{-\alpha t} \tag{7}$$

where k is the asymptotic value and α is the decay rate

$$\alpha = \frac{log(2)}{t_{1/2}} \tag{8}$$

In experiments measuring conversion coefficients, we compare the pre-mRNA decay in unstimulated cells to the decay in EGF-stimulated cells in which transcription was arrested near the peak of pre-mRNA induction. Since the starting point of pre-mRNA levels in the unstimulated cells is much lower (with respect to the stimulated case), a similar level of transcription inhibition will result in different asymptotic constants. Thus, when fitting the parameters to the decay curves we optimize both α and k, and expect to get similar values of α, but different k. In the mRNA decay experiments, we compare the decay in unstimulated cells with that observed in cells pre-stimulated (with EGF) for 4 hours. We expect less variation in the asymptotic values, since similar inhibition levels are expected and the differences in initial levels are smaller. Therefore, when fitting the parameters to the profiles we force a similar k and obtain α.

2 Modeling transcription

Our aim is to write down a simple dynamic model, which governs RNA concentration. Most previous descriptions focus on a minimal model including only the mRNA concentration. Here, we take the opportunity to refine the model by including more components. The model includes two RNA species, pre-RNA and mRNA, and can be expressed as:

$$\begin{aligned}\frac{dP(t)}{dt} &= \beta(t) - \alpha_1 P(t)\\ \frac{dM(t)}{dt} &= \alpha_1 P(t) - \alpha_2(t) M(t)\end{aligned} \tag{9}$$

with a steady state solution:

$$P_{ss} = \frac{\beta}{\alpha_1}, M_{ss} = \frac{\beta}{\alpha_2} \Rightarrow \frac{P_{ss}}{M_{ss}} = \frac{\alpha_2}{\alpha_1} = \frac{t^P_{1/2}}{t^M_{1/2}} \tag{10}$$

$P(t), M(t)$ are the pre- and mature mRNA concentrations at time t, α_1 is the conversion constant, $\alpha_2(t)$ is the time-dependent degradation coefficient, and $\beta(t)$ is the production rate. Note that the system can easily be solved analytically for fixed (independent of time) $\alpha_1, \alpha_2, \beta$. An analytic solution can be written as:

$$\begin{aligned}P(t) &= \left(P_0 - \frac{\beta}{\alpha_1}\right) e^{-\alpha_1 t} + \frac{\beta}{\alpha_1}\\ M(t) &= \left(M_0 - \frac{\alpha_1 P_0 - \beta}{\alpha_2 - \alpha_1} - \frac{\beta}{\alpha_2}\right) e^{-\alpha_2 t} + \frac{\alpha_1 P_0 - \beta}{\alpha_2 - \alpha_1} e^{-\alpha_1 t} + \frac{\beta}{\alpha_2}\end{aligned} \tag{11}$$

Since we cannot easily measure the absolute value of RNA concentrations, but only their fold change (relative abundance), it is a practical step to change the variables to the fold change with respect to $t = 0$, at which time we assume the system was at a steady state, parameterized by:

$$\begin{aligned}\beta(0) &= \alpha_1(0) P(0)\\ \frac{P(0)}{M(0)} &= \frac{\alpha_2(0)}{\alpha_1(0)}\end{aligned} \tag{12}$$

In terms of the transformed fold change variables,

$$\hat{P} = \frac{P}{P(0)}, \quad \hat{M} = \frac{M}{M(0)}, \quad \hat{\alpha} = \frac{\alpha}{\alpha(0)}, \quad \hat{\beta} = \frac{\beta}{\beta(0)}, \tag{13}$$

Equations (2-3) in the main text become:

$$\begin{aligned}\frac{d\hat{P}(t)}{dt} &= \alpha_1(0)\left[\hat{\beta}(t) - \hat{\alpha_1}\hat{P}(dt)\right]\\ \frac{d\hat{M}(t)}{dt} &= \alpha_2(0)\left[\hat{\alpha_1}\hat{P}(t) - \hat{\alpha_2}(dt)\hat{M}(t)\right]\end{aligned} \tag{14}$$

An analytical solution for the case of step changes: β, ($\beta \to a\beta$) and/or α_2, ($\alpha_2 \to \alpha_2/b$) and/or α_1 ($\alpha_1 \to c\alpha_1$) at $t = 0$ can be written in the fold change form:

$$\begin{aligned}\hat{P}(t) &= \left(1 - \frac{a}{c}\right) e^{-c\alpha_1 t} + \frac{a}{c}\\ \hat{M}(t) &= \left(1 - \frac{1 - \frac{a}{c}}{\frac{1}{bc} - \frac{\alpha_1}{\alpha_2}} - ab\right) e^{-\frac{\alpha_2}{b} t} + \frac{1 - \frac{a}{c}}{\frac{1}{bc} - \frac{\alpha_1}{\alpha_2}} e^{-c\alpha_1 t} + ab\\ &= (1 - ab) e^{-\frac{\alpha_2}{b} t} + ab + \frac{1 - \frac{a}{c}}{\frac{1}{bc} - \frac{\alpha_1}{\alpha_2}} \left(e^{-c\alpha_1 t} - e^{-\frac{\alpha_2}{b} t}\right)\end{aligned} \tag{15}$$

The response time for this case in terms of the rise time to the asymptotic value, $\hat{P} \to 1 + \frac{1}{2}\left(\frac{a}{c} - 1\right), \hat{M} \to 1 + \frac{1}{2}(ab - 1)$, for pre-mRNA is:

$$t_{1/2}^{P} = \frac{log(2)}{c\alpha_1} \tag{16}$$

while for mature mRNA it is hard to write an explicit term, but $t_{1/2}^{M}$ should satisfy:

$$\frac{1}{2} = e^{-\frac{\alpha_2}{b}t_{1/2}^{M}} + \frac{\frac{a}{c} - 1}{\left(\frac{1}{bc} - \frac{\alpha_1}{\alpha_2}\right)(ab - 1)}\left(e^{-c\alpha_1 t_{1/2}^{M}} - e^{-\frac{\alpha_2}{b}t_{1/2}^{M}}\right) \tag{17}$$

The fold change of exons, which are a transcript region shared by pre-mRNA and mature mRNA, can be written as:

$$\hat{E} = \frac{M(t) + P(t)}{P(0) + M(0)} = \frac{\hat{M}M(0) + \hat{P}P(0)}{P(0) + M(0)} = \frac{\hat{M} + \hat{P}\frac{\alpha_2(0)}{\alpha_1(0)}}{\frac{\alpha_2(0)}{\alpha_1(0)} + 1} \tag{18}$$

By additionally measuring the FC of exons, it is possible to extract the fraction $\alpha_2(0)/\alpha_1(0)$, from the above equation (assuming steady state at $t = 0$):

$$\frac{\alpha_2(0)}{\alpha_1(0)} = \frac{\hat{E} - \hat{M}}{\hat{P} - \hat{E}} \tag{19}$$

Note that in practice this ratio is hard to extract from our measurements, because of the small differences between $\hat{E}$ and $\hat{M}$, which are of the same order as the measurement errors of qPCR. However, this ratio also predicts that in the initial phase of induction, where $\hat{P} - \hat{E} > 0$, the numerator should be positive and $\hat{E} > \hat{M}$, while in later phases, when the $\hat{P}$ starts to fall, both numerator and denominator should switch sign at the same point. This means that there is a special t, where the three curves $\hat{E}, \hat{M}, \hat{P}$ intersect. Indeed, profiles shown in Supplementary Figure S5 demonstrate that this is often the case.

3 Supplementary Figures and File

Supplementary File 1: Functional annotation analysis of genes with overshooting production and primer sequences used for qPCR experiments.

4 Supplementary References

1. Ballman KV, Grill DE, Oberg AL, Therneau TM (2004) Faster cyclic loess: normalizing RNA arrays via linear models. Bioinformatics 20: 2778-2786
   Friedel CC, Dolken L, Ruzsics Z, Koszinowski UH, Zimmer R (2009) Conserved principles of mammalian transcriptional regulation revealed by RNA half-life. Nucleic Acids Res 37: e115

2. Perry RP, Kelley DE (1970) Inhibition of RNA synthesis by actinomycin D: characteristic dose-response of different RNA species. J Cell Physiol 76: 127-139

3. Pfaffl MW (2001) A new mathematical model for relative quantification in real-time RT-PCR. Nucleic Acids Res 29: e45

4. Schmittgen TD, Livak KJ (2008) Analyzing real-time PCR data by the comparative C(T) method. Nat Protoc 3: 1101-1108

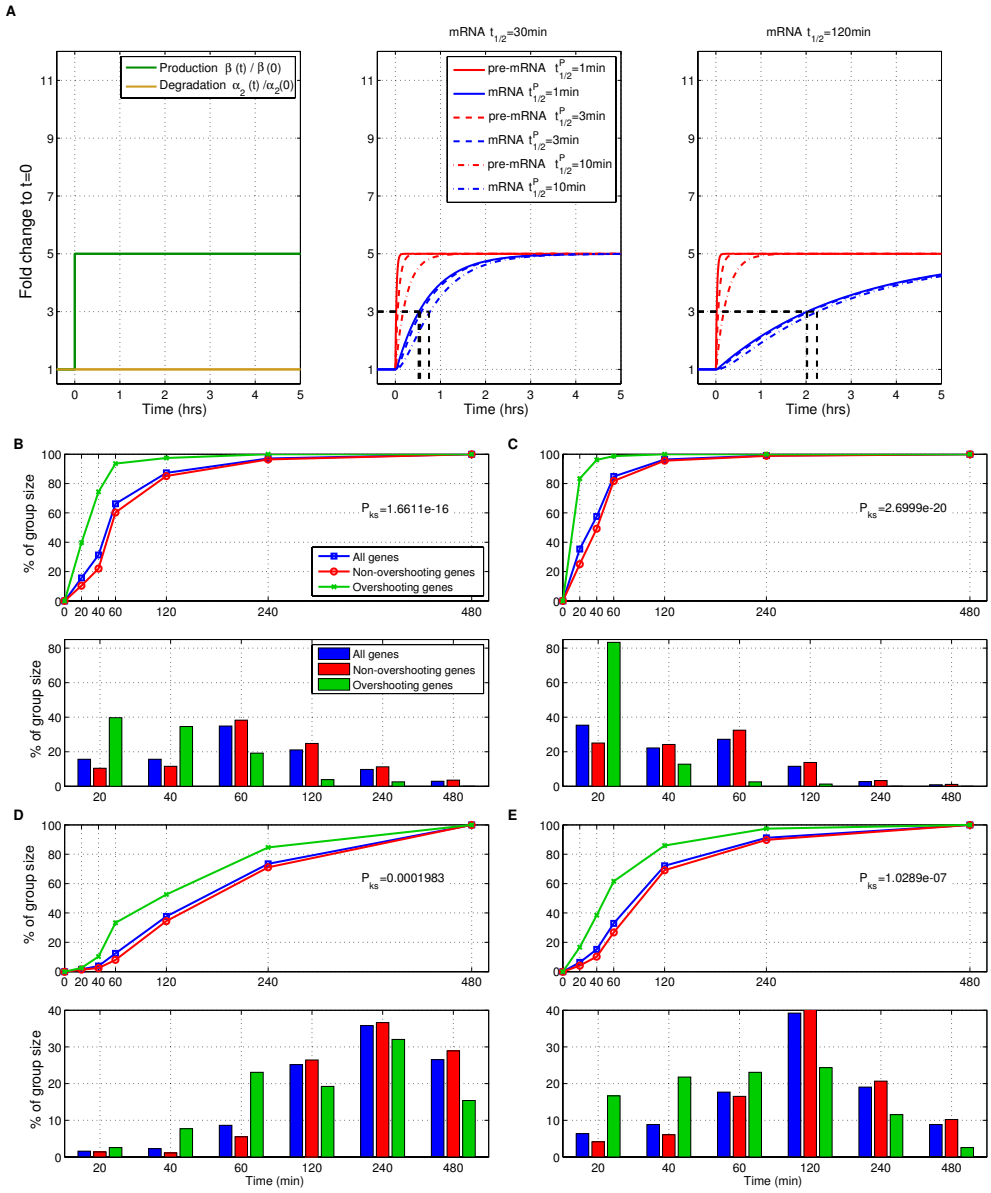
5. Singh J, Padgett RA (2009) Rates of in situ transcription and splicing in large human genes. Nat Struct Mol Biol 16: 1128-1133

6. Tennyson CN, Klamut HJ, Worton RG (1995) The human dystrophin gene requires 16 hours to be transcribed and is cotranscriptionally spliced. Nat Genet 9: 184-190

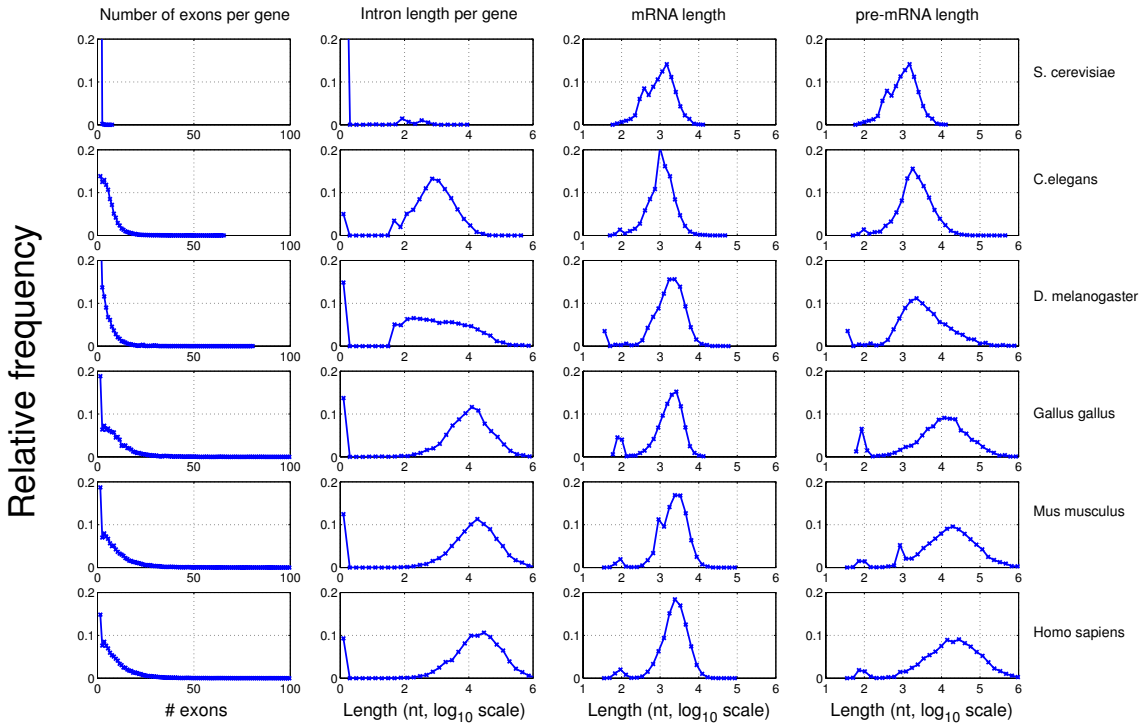
7. Venter JC, Adams MD, Myers EW, Li PW, Mural RJ, Sutton GG, Smith HO, Yandell M, Evans CA, Holt RA, Gocayne JD, Amanatides P, Ballew RM, Huson DH, Wortman JR, Zhang Q, Kodira CD, Zheng XH, Chen L, Skupski M et al (2001) The sequence of the human genome. Science 291: 1304-1351

8. Wada Y, Ohta Y, Xu M, Tsutsumi S, Minami T, Inoue K, Komura D, Kitakami J, Oshida N, Papantonis A, Izumi A, Kobayashi M, Meguro H, Kanki Y, Mimura I, Yamamoto K, Mataki C, Hamakubo T, Shirahige K, Aburatani H et al (2009) A wave of nascent transcription on activated human genes. Proc Natl Acad Sci U S A 106: 18357-18361

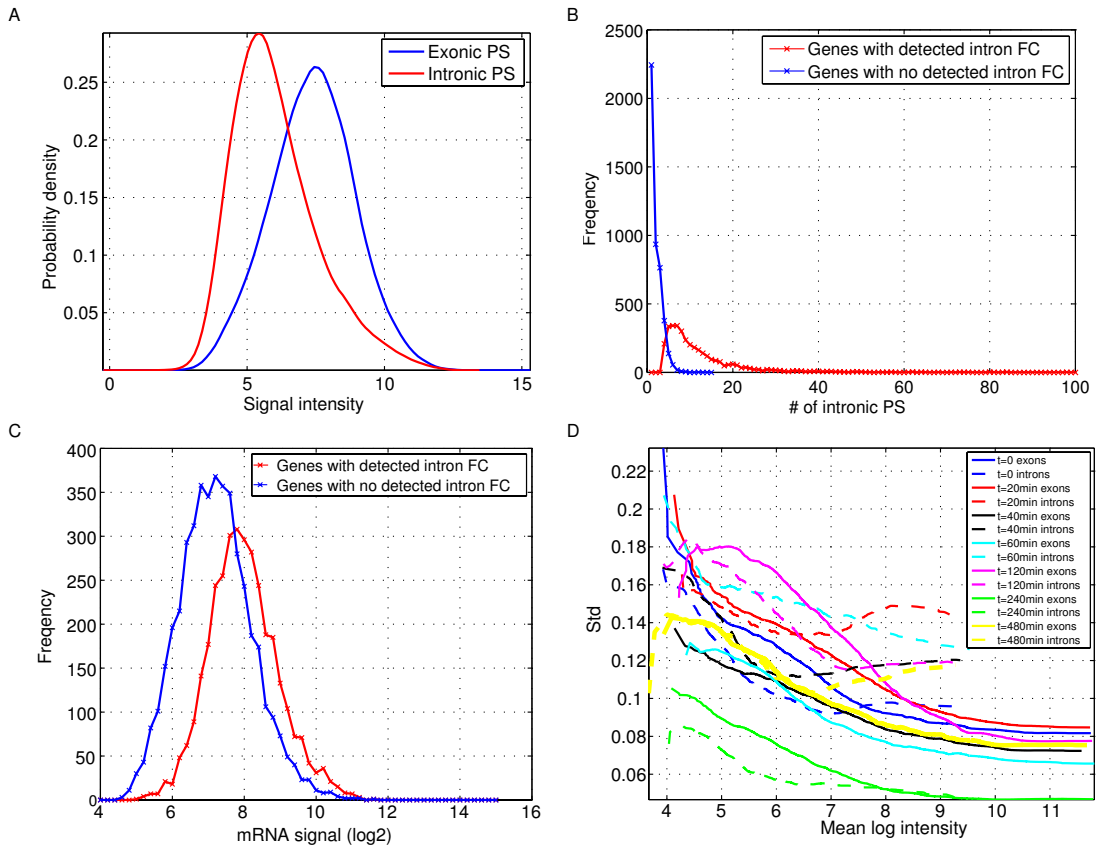
9. Zeisel A, Amir A, Kostler WJ, Domany E (2010) Intensity dependent estimation of noise in microarrays improves detection of differentially expressed genes. BMC Bioinformatics 11: 400



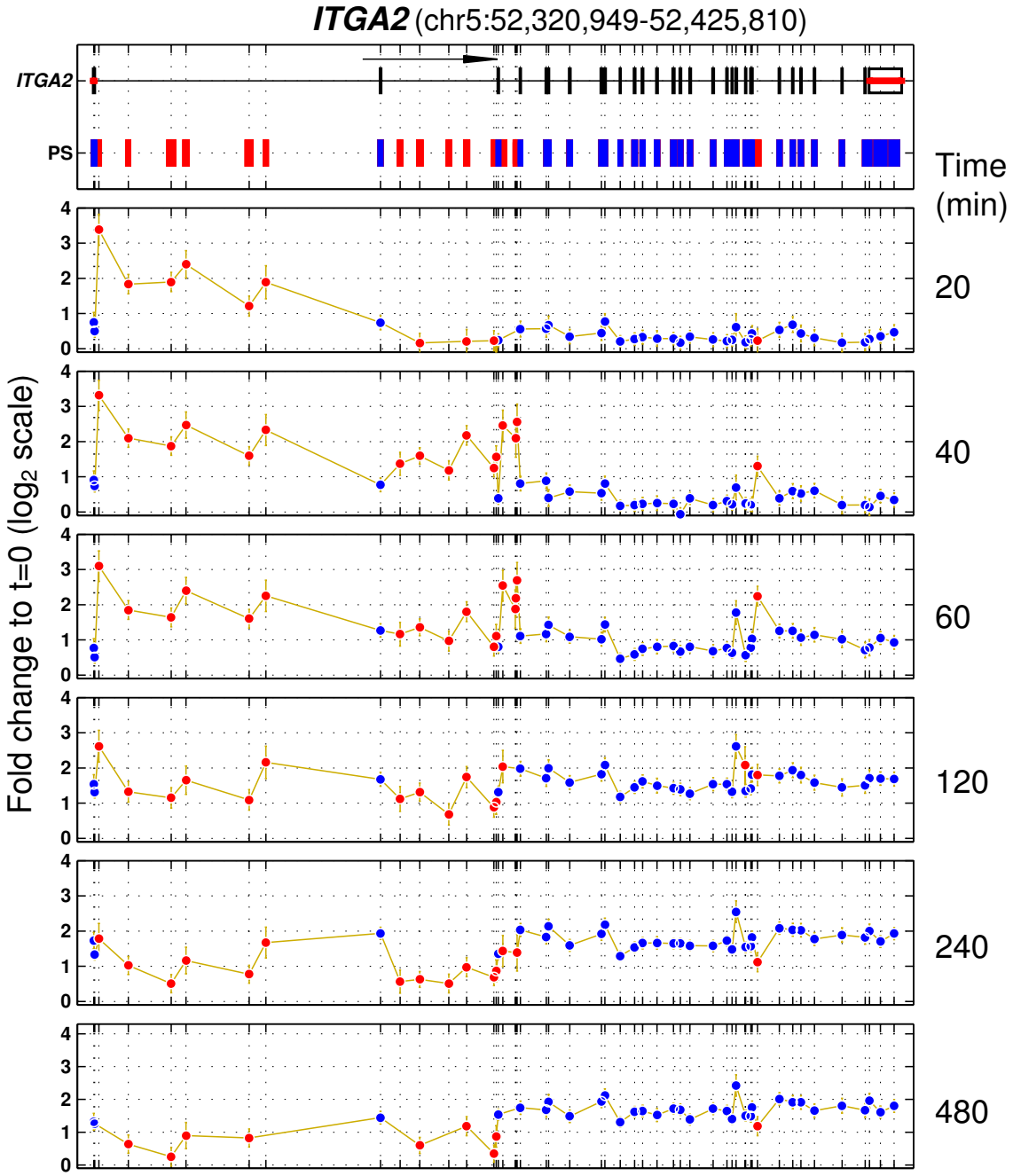
Supplementary Figure S1: Onset and peak times of gene-level intron and exon fold changes (A) pre-mRNA half-life has a small effect on mRNA dynamics: simulation results for the case of 5-fold step induction of transcript production rate, for the case of pre-mRNA half-lives of 1,3, or 10 minutes (solid, dashed, dash-dot lines, respectively), and for mRNA half-lives of 60 (middle panel) or 120 (right panel) minutes. Results show that varying pre-mRNA half-lives over a 10-fold range has only a minor effect on mRNA dynamics. (B) The cumulative distribution function by gene-level intronic FC peak times is shown for all 441 transcriptionally induced genes (blue); 79 overshooting genes (green) and for 362 non-overshooting genes (red). (C) Same as (B) for intronic onset times (defined as the time when FC > 2). Kolmogorow-Smirnov (KS) tests, comparing the red and green curves, indicated significantly earlier peak ($p = 8.10E-17$) and onset ($p = 1.19E-20$) times of gene-level intron FC for genes with overshooting introns. (D) and (E): In analogy to (B) and (C), histograms of the onset times (D) and peak times (E) of gene-level exon changes are shown. Again, KS tests indicated significantly earlier peak ($p = 1.70E-07$) and onset ($p = 1.16E-07$) times of gene-level exon FC, for genes with overshooting introns. Note that the intronic peak and onset times clearly preceded the corresponding exonic times.



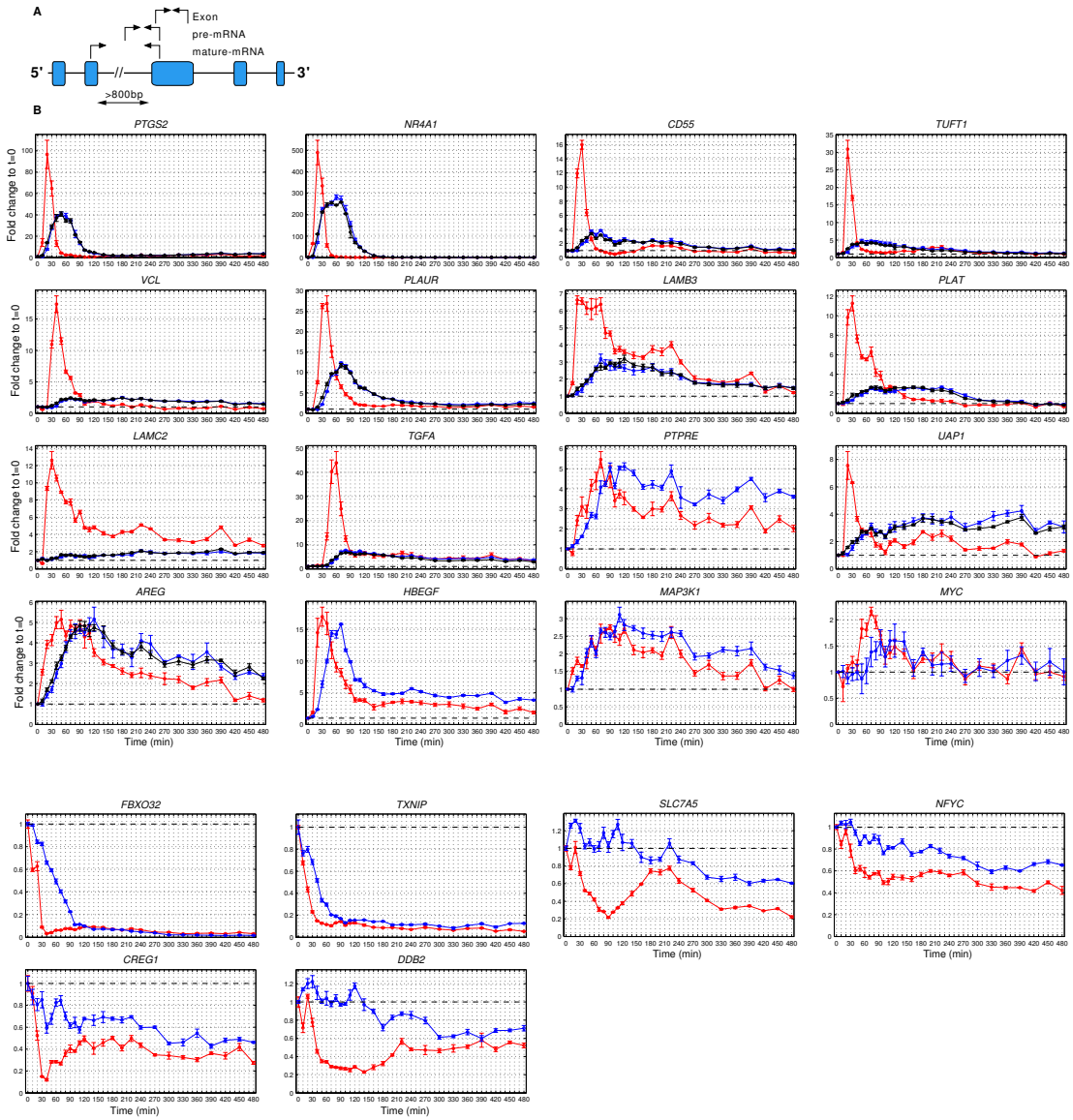
Supplementary Figure S2: Number of exons, intron length, mRNA length and pre-mRNA length of six different organisms Annotated full-length transcripts for six eukaryotic organisms were downloaded from the UCSC Genome Browser (http://genome.ucsc.edu/). For each organism, the distributions of the number (per gene) of exons, the total number of intronic nucleotides, and the lengths of mRNAs and pre-mRNAs are shown on a logarithmic scale. Note that already in simple multicellular organisms (such as C. elegans) there is a clear difference between pre-mRNA and mRNA lengths for most genes. Thus, for most multicellular organisms it is possible to analyze pre-mRNA and mRNA expression, using intronic and exonic probes, as done in the present study.



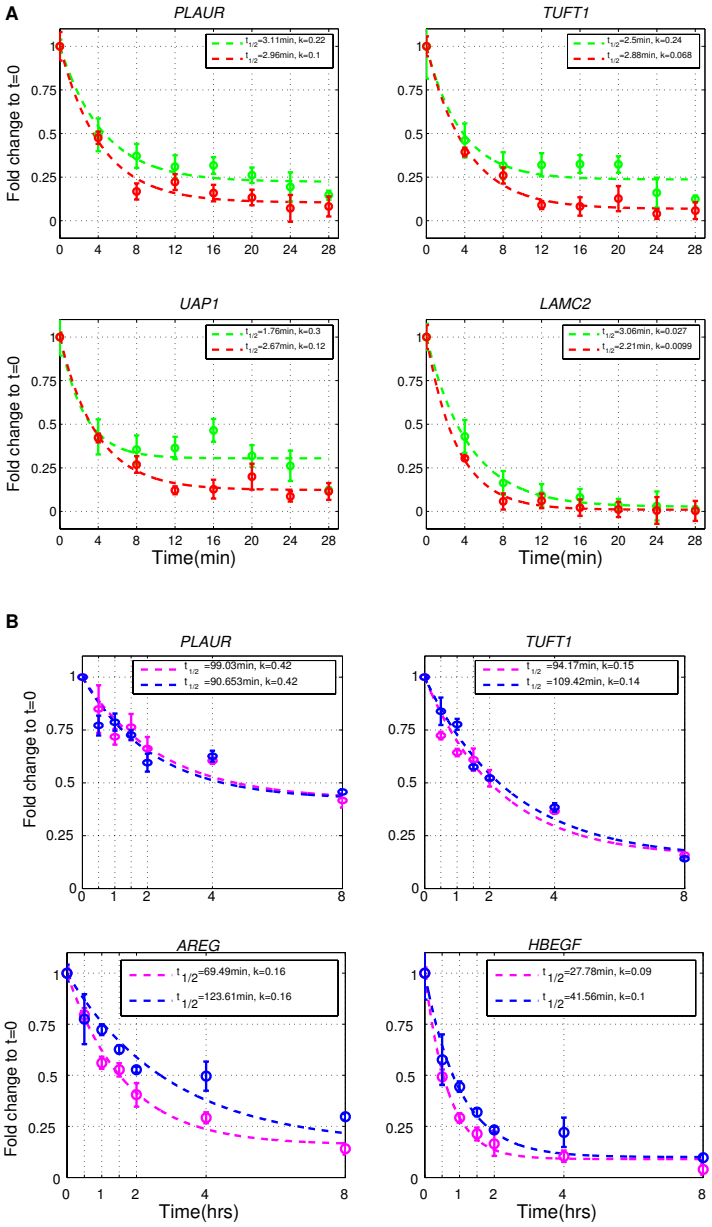
Supplementary Figure S3: Signal distributions of intronic and exonic PS (A) Distribution of measured signal intensity of all intronic and exonic probesets (PS) that passed the PS-level quality filter (described in *Supplementary Information*, Microarray preprocessing). The distribution of intronic expression levels is is shifted to lower values with respect to the exonic PS. (B) Histograms of the number of intronic PS (on the Affymetrix exon array) are shown for genes with detectable gene-level intronic FC (red) and (blue) for genes with undetected gene-level intronic FC (i.e., genes whose intronic signals did not pass the filters). Note that for each gene at least 4 intronic PS had to be expressed to define a detected gene-level intronic FC (*Supplementary Information*, Microarray preprocessing). (C) Histograms of the mRNA (gene-level exonic) signals for the same two groups of genes as in (B). Genes whose gene-level intronic FC was not detectable have lower gene-level exonic signal intensity. (D) Estimated intensity-dependent standard deviation of the gene-level intronic and exonic signals for biological triplicate measurements taken at different time-points. Intron gene-level signal is slightly more noisy than exonic gene-level signal, but the standard-deviation is less than 0.18 when the signal is not very low.



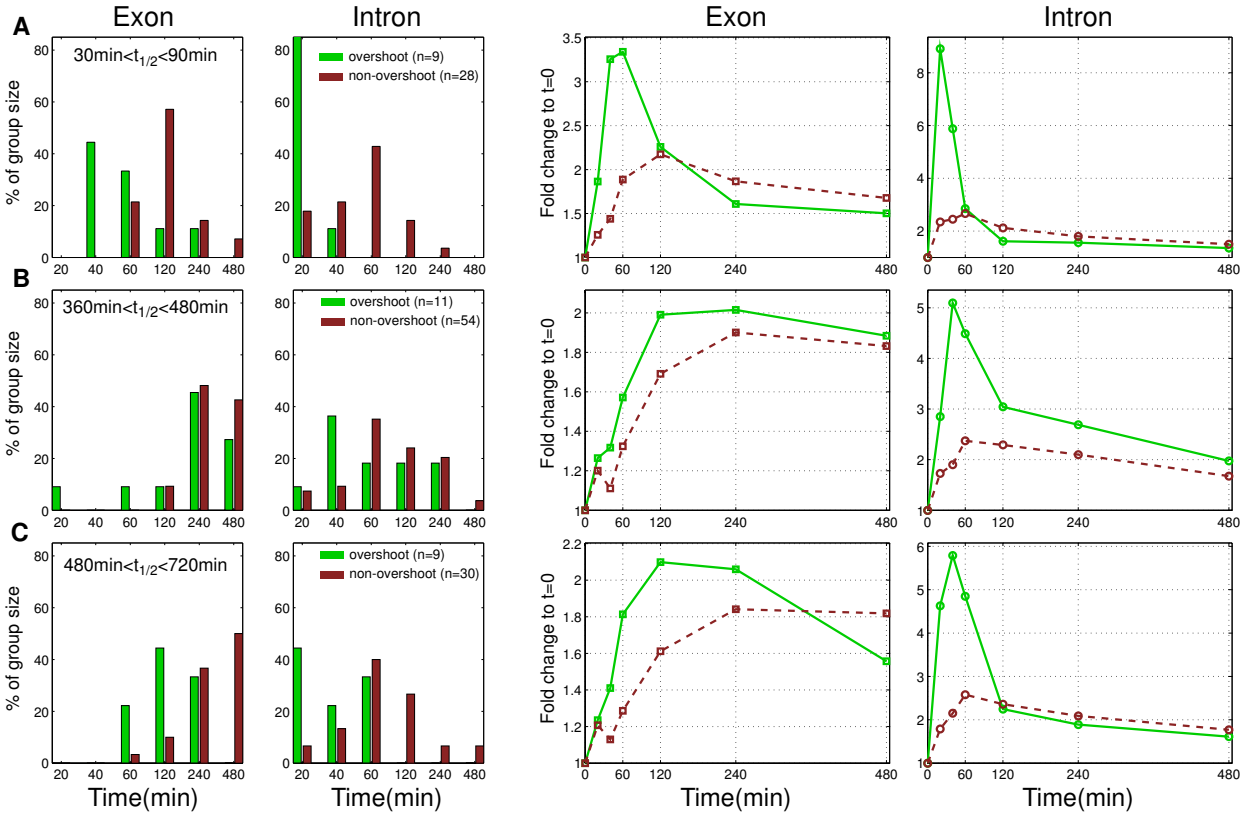
Supplementary Figure S4: Transcriptional induction of a large non-overshooting gene (*ITGA2*) Upper panel: Top - genomic organization of the *ITGA2* gene (genomic size 105kb). Vertical rectangles represent exons and are connected by intronic regions. The arrow indicates the direction of transcription. Bottom - positions of exonic (blue) and intronic (red) probesets are indicated. Lower panels: Fold change (log2 scale) of each probeset (PS) with respect to its baseline values is shown for the time course experiment outlined in Figure 2A. Error bars (in gold) represent the signal range of biological triplicates. Only PS with present calls (see *Supplementary Information*) in all replicates from the respective time points are shown. Note that intronic PS (red dots) display different dynamics from exonic PS (blue dots). However, in contrast to the *VCL* gene shown in Figure 3, the dynamic range of the FC of intronic PS did not exceed by much that of the exonic PS.



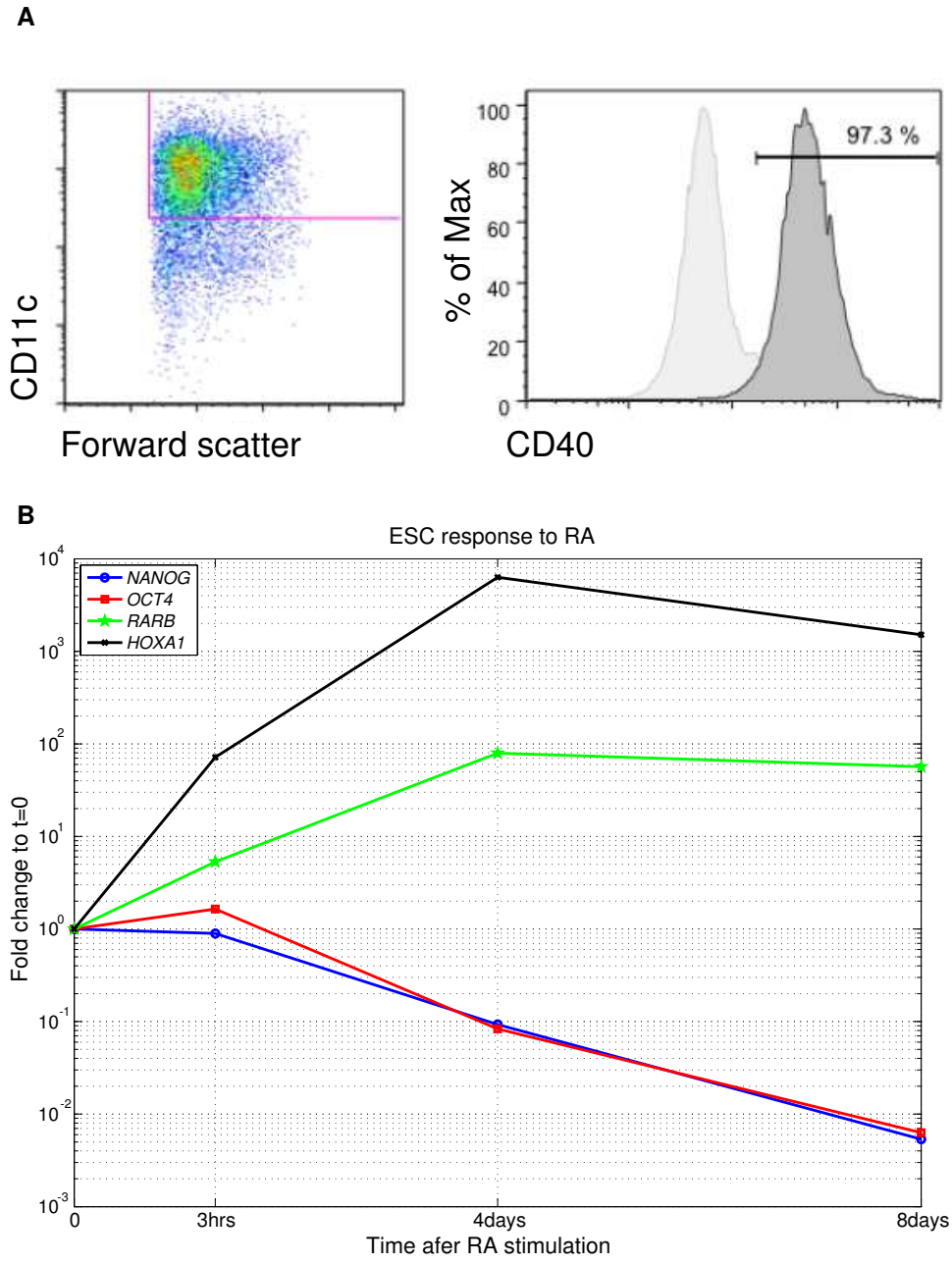
Supplementary Figure S5: Real time quantitative PCR (qPCR) measurements of exons, pre-mRNAs and mRNAs of genes induced/repressed by EGF in MCF10A cells (A) Schematic positioning of primers for qPCR validation experiments is shown. (B) Means and standard deviations from technical triplicates of qPCR reactions for exons (black), pre-mRNA (red) and mRNA (blue) of the fold change (relative to $t = 0$), as a function of time following stimulation, for *MYC, HBEGF, NR4A1, PTGS2* and the 12 EGF-induced genes that were listed on the left in Figure 2B, and 6 down-regulated genes from Figure 4.



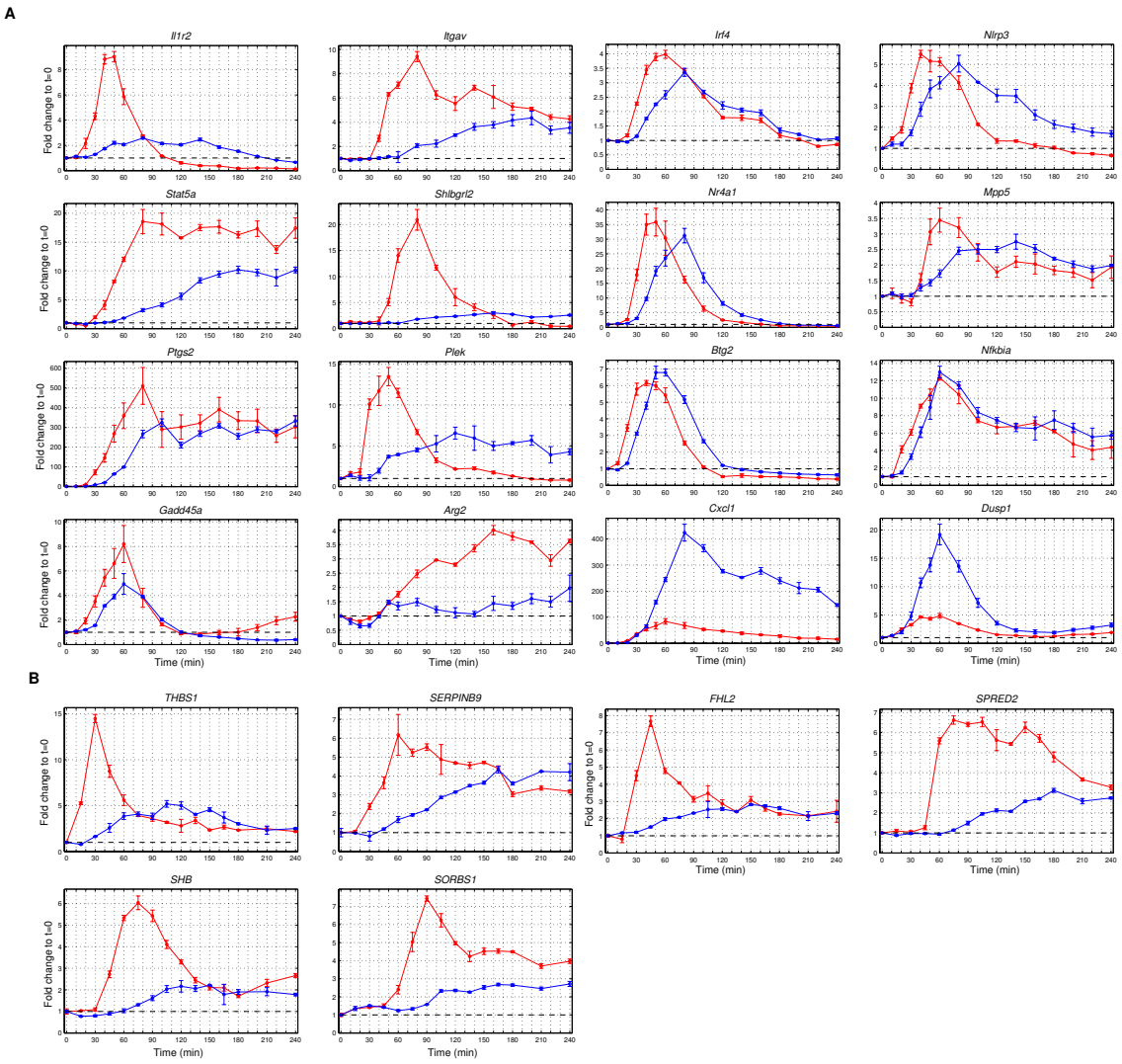
Supplementary Figure S6: Decay rates of pre-mRNA and mRNA in unstimulated and EGF-stimulated MCF10A cells RNA decay was measured as described in the legend to Figure 5, by qPCR after transcription arrest using Actinomycin D in unstimulated cells and in cells pre-stimulated with EGF. (A) Pre-mRNA decay was measured for 4 additional genes (with EGF-induced overshooting pre-mRNA dynamics) by applying transcription arrest in unstimulated MCF10A cells (green curves) and in cells stimulated (red) for 20 minutes (*PLAT, TUFT*) or for 40 minutes (*LAMC2, PLAUR*) with EGF (transcription was arrested at the peaks of pre-mRNA induction). The results demonstrate that pre-mRNA conversion is practically unaffected by EGF stimulation. (B) EGF-induced changes in mRNA decay rates: Starved MCF10A cells were stimulated for 4 hours by EGF (blue symbols) or unstimulated (magenta symbols) but left for 4 more hours in starvation medium, before addition of Actinomycin D (ActD, 5μM). RNA was extracted at the indicated time points starting 30 minutes after transcription arrest and mRNA was measured by qPCR. mRNA decay curves are shown for genes, for which inferences of mRNA production and degradation rates (see Figure 4) did (*AREG, HBEGF*) or did not (*PLAUR, TUFT1*) reveal an EGF-induced change in transcript stability at long times.The inferred changes of stability are in agreement with those measured here, while the actual degradation times obtained with transcription arrest are longer.



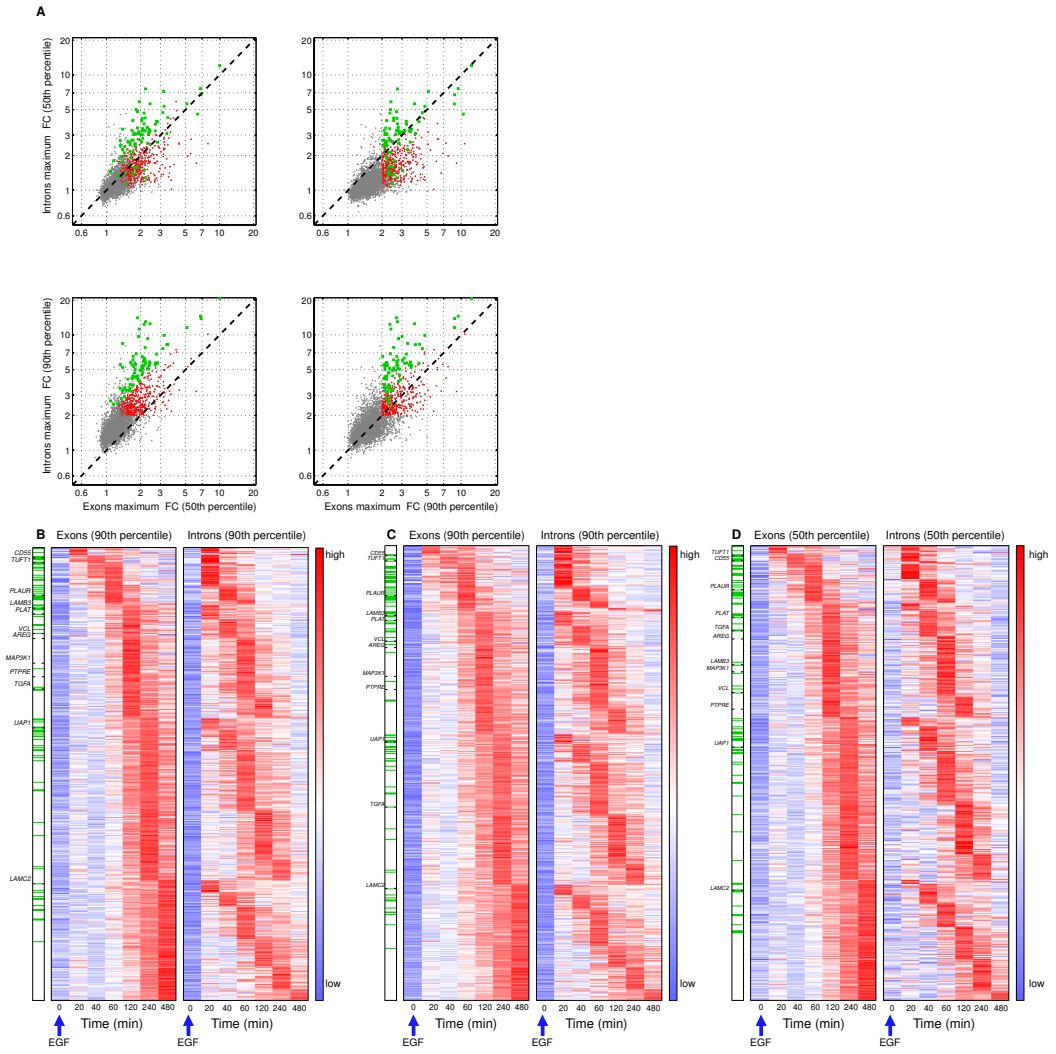
Supplementary Figure S7: Pre-mRNA production overshoot accelerates the response time of mRNAs Same analysis as shown in Figure 6, but with mRNA degradation times determined by metabolic labelling with 4sU for 60 minutes followed by microarray analyses of labelled and unlabelled RNA in (Friedel et al, 2009). Genes that were transcriptionally induced by EGF stimulation of MCF10A cells (as described in Legend to Figure 2) were grouped according to their mRNA half-lives Results are shown for three representative groups with half-lives between (A) 30 to 90 minutes (B) 360 to 480 minutes, and (C) 480 to 720 minutes, respectively (due to the labelling period no information was available for mRNAs with half-lives $<$ 30 minutes). The two columns on the left depict the percentage of genes, whose exons and introns peaked at the indicated time points, green for overshooting and brown for non-overshooting genes. The two columns on the right present the mean temporal expression FC profiles for exons and introns, for the overshooting (green) and non-overshooting (brown) genes. Note that within each group of genes with similar half-lives we observe earlier mRNA peak times for the genes with pre-mRNA production overshoot. Comparison with Figure 6 indicates that the main observation (faster response of overshooting genes) is unaltered, even though the degradation times obtained by the two methods did differ.



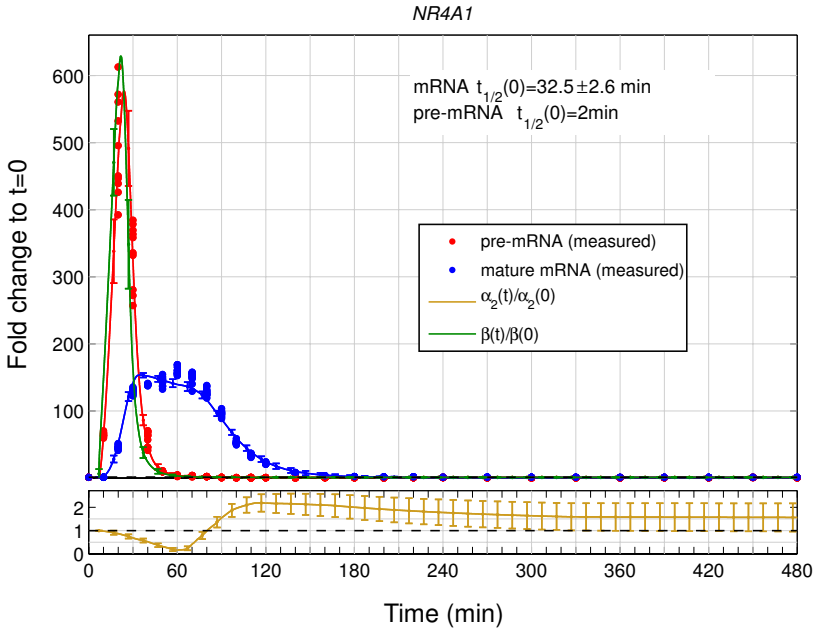
Supplementary Figure S8: LPS-induced activation of dendritic cells (DCs) and retinoic acid (RA)-induced differentiation of human embryonic stem cells (hESCs) (A) Left panel: Bone-marrow derived DCs were prepared as described in *Supplementary Information*, incubated with the indicated antibodies and subjected to FACS analysis. Left panel: Expression of the DC marker CD11c in unstimulated DCs. Right panel: FACS analysis of unstimulated DCs (light grey) and of DCs stimulated for 24 hours with LPS (100ng/ml, dark grey) showing LPS-induced upregulation of the activation marker CD40. (B) hESCs were incubated with RA (1μM). RNA was isolated at the indicated time points and expression of pluripotency markers (*NANOG, OCT4*) and differentiation markers (*RARB, HOXA1*) was measured by qPCR.



Supplementary Figure S9: Real time quantitative PCR (qPCR) measurements of pre-mRNAs and mRNAs of genes induced by LPS stimulation of dendritic cells (DCs) and by retinoic acid stimulation in human embryonic stem cells (hESCs) (A) DCs were harvested, cultured and LPS-stimulated as described under *Supplementary Information*. RNA was isolated at the indicated time points and qPCR was used to measure the expression of pre-mRNAs (red) and mRNAs (blue). Data shown are means and standard deviations from technical triplicates. (B) hESCs were cultured and stimulated with RA as described under *Supplemental Information*. RNA was isolated at the indicated time points and qPCR was used to measure the expression of pre-mRNAs (red) and mRNAs (blue). Data shown are means and standard deviations from technical triplicates.



Supplementary Figure S10: Scatter plots and heatmaps comparing fold changes (FC) of introns and exons according to different cutoffs (A) The peak FC of gene-level introns observed after EGF stimulation is plotted against the peak FC of exons, with each plot using different percentiles of the averaged intron and exon signals to define gene-level FC. Grey dots indicate genes for which, over the whole time course, exon FC (using 50th percentile) was ≤ 1.5. Transcriptionally induced genes (for which gene-level exon FC (using 50th percentile) exceeded 1.5 at least at one time point) are marked by either green or red dots. Green dots denote overshooting genes for which the maximum gene-level intronic FC (using 90th percentile) exceeded the maximum gene-level exonic FC by a factor of 2 or more, and red dots non overshooting genes. The four subfigures correspond to different definitions of the gene-level exonic/intronic FC, as indicated on the axes. Note that the behavior of genes with production overshoot is robust to the choice of percentiles used to define exonic and intronic gene-level FC. (B) to (D): Throughout our analyses presented in the main text, we used gene-level 50th (90th) percentiles for exons (introns) to define 441 genes as transcriptionally induced for which the gene-level exonic FC exceeded 1.5 and the gene-level intronic FC exceeded 2.1 (see *Supplementary Information* for reasons for choice of these percentiles and cutoffs). Heatmaps presented here show, for the same 441 genes, the effect of changing the percentiles used to define the exonic and intronic gene-level expression. The left pair of heatmaps (B) used 50th percentile for exons and 90th for introns, the middle pair (C) used the 90th percentile for both and the right pair (D) used the 50th percentile for both. For each such pair the genes were grouped first according to their exonic peak expression FC times. Next, genes were sorted within each group according to the intronic peak FC times. Note the robustness of our main findings (e.g. wide range of delays between pre-mRNA and mRNA peak times, imperfect correlation between the two profiles) against the particular definition of gene-level expression used.



Supplementary Figure S11: Estimation of the error of fitted functions: Estimation of the error of the fitted and inferred functions, on the basis of the estimated qPCR measurement noise. The error bars were estimated by simulating 10 measurements of 10 pre-mRNA and mRNA profiles using the noise distribution of the three technical replicates of qPCR measurements. After generating 10 randomly drawn pre-mRNA and mRNA profiles, (see *Supplementary Information*) we performed the fitting process for each of the 10 cases, and calculated the means and standard deviations of the fitted curves $\hat{P}^{fit}(t)$ and $\hat{M}^{fit}(t)$, and of the inferred functions $\hat{\beta}(t)$ and $\hat{\alpha}_2(t)$.