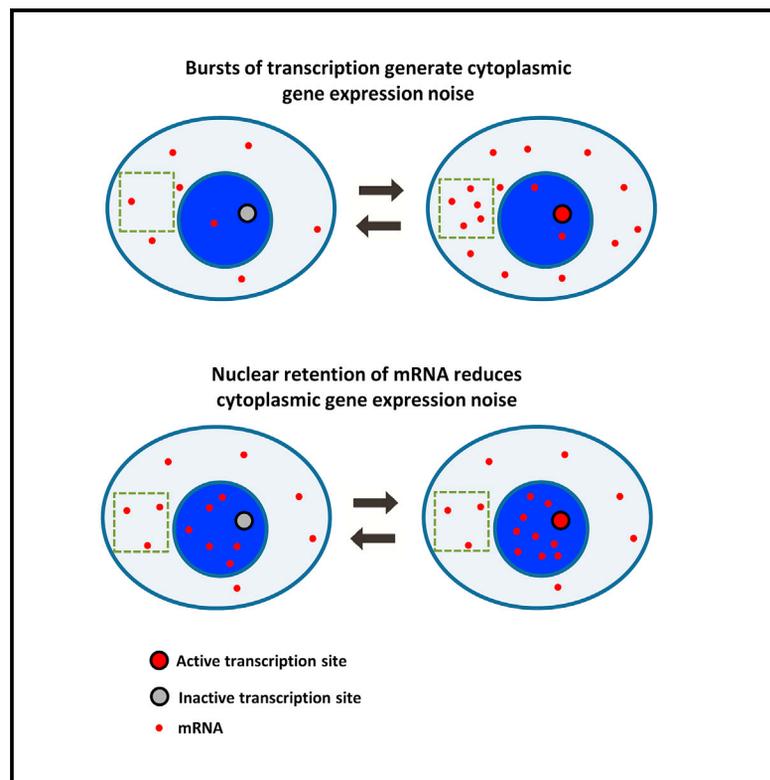


Nuclear Retention of mRNA in Mammalian Tissues

Graphical Abstract



Highlights

- Genome-wide catalog of nuclear and cytoplasmic mRNA in mouse tissues
- Spliced, polyadenylated mRNA is retained in the nucleus for many protein-coding genes
- Retained genes include ChREBP and liver Nlrp6, co-localized with nuclear speckles
- Nuclear retention of mRNA reduces cytoplasmic gene expression noise

Authors

Keren Bahar Halpern, Inbal Caspi, Doron Lemze, ..., Eran Elinav, Igor Ulitsky, Shalev Itzkovitz

Correspondence

shalev.itzkovitz@weizmann.ac.il

In Brief

Bahar Halpern et al. combine whole-transcriptome and single-molecule approaches to demonstrate that a substantial fraction of genes have higher levels of spliced, polyadenylated mRNA in the nucleus compared to the cytoplasm in mammalian tissues. This nuclear retention reduces cytoplasmic gene expression noise created by transcriptional bursts.

Accession Numbers

GSE73977



Nuclear Retention of mRNA in Mammalian Tissues

Keren Bahar Halpern,^{1,4} Inbal Caspi,^{1,4} Doron Lemze,¹ Maayan Levy,² Shanie Landen,¹ Eran Elinav,² Igor Ulitsky,³ and Shalev Itzkovitz^{1,*}

¹Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot 76100, Israel

²Department of Immunology, Weizmann Institute of Science, Rehovot 76100, Israel

³Department of Biological Regulation, Weizmann Institute of Science, Rehovot 76100, Israel

⁴Co-first author

*Correspondence: shalev.itzkovitz@weizmann.ac.il

<http://dx.doi.org/10.1016/j.celrep.2015.11.036>

This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

SUMMARY

mRNA is thought to predominantly reside in the cytoplasm, where it is translated and eventually degraded. Although nuclear retention of mRNA has a regulatory potential, it is considered extremely rare in mammals. Here, to explore the extent of mRNA retention in metabolic tissues, we combine deep sequencing of nuclear and cytoplasmic RNA fractions with single-molecule transcript imaging in mouse beta cells, liver, and gut. We identify a wide range of protein-coding genes for which the levels of spliced polyadenylated mRNA are higher in the nucleus than in the cytoplasm. These include genes such as the transcription factor ChREBP, Nlrp6, Glucokinase, and Glucagon receptor. We demonstrate that nuclear retention of mRNA can efficiently buffer cytoplasmic transcript levels from noise that emanates from transcriptional bursts. Our study challenges the view that transcripts predominantly reside in the cytoplasm and reveals a role of the nucleus in dampening gene expression noise.

INTRODUCTION

The life course of mRNA begins with transcription, splicing, and processing, which generally occur at the nuclear sites of transcription, and ends with cytoplasmic translation and degradation. Nuclear export of mRNA is considered a transient phase, lasting only a few minutes in mammalian cells (Oeffinger and Zenklusen, 2012; Shav-Tal et al., 2004; Vargas et al., 2005), a negligible time compared to the other phases. Recent studies applied deep sequencing of RNA from cellular fractions to identify RNA molecules that are retained in the nucleus (Bhatt et al., 2012; Djebali et al., 2012; Pandya-Jones et al., 2013; Tilgner et al., 2012). These, however, predominantly included long non-coding RNA (lncRNA), such as Xist, Malat1, and Neat1; hyper-edited mRNA (Chen and Carmichael, 2009); or incompletely spliced mRNA (Boutz et al., 2015; Shalgi et al., 2014) rather than mature protein-coding mRNA. Though rare examples exist for nuclearly retained transcripts (Prasanth et al., 2005), a global picture of mRNA nuclear retention in mammalian tissues is lacking.

Here, to explore the extent and possible roles of nuclear mRNA retention, we combined deep sequencing of RNA from nuclear and cytoplasmic fractions with single-molecule transcript imaging in intact mouse tissues. Surprisingly, we found a wide range of spliced polyadenylated protein-coding mRNA, which are nuclearly retained for the majority of their lifetime. These include Glucokinase and Glucagon receptor in beta cells; Nlrp6 in the liver; and, most strikingly, the transcription factor Mlxipl, also known as ChREBP, the transcripts of which are highly retained in nuclear speckles in liver, beta cells, and intestinal tissue. We developed a single-molecule in situ technique to quantify nuclear mRNA lifetimes and found that the transcripts of these genes can spend hours in the nucleus before being exported to the cytoplasm.

To study the potential role of nuclear retention, we analyzed its impact on fluctuations in cytoplasmic mRNA levels. Mammalian genes are transcribed in bursts (Larson et al., 2011; Darzacq et al., 2007; Suter et al., 2011; Bahar Halpern et al., 2015; Dar et al., 2012; Senecal et al., 2014), leading to temporal fluctuations in cellular mRNA levels and variability among identical cells (Blake et al., 2003; Eldar and Elowitz, 2010; Golding et al., 2005; Kaern et al., 2005; Maheshri and O'Shea, 2007; Paulsson, 2004; Raj and van Oudenaarden, 2008). We demonstrate theoretically and experimentally that nuclear retention can effectively buffer these fluctuations, facilitating lower variability in cytoplasmic mRNA.

RESULTS

RNA Sequencing of Cell Fractions Reveals Broad Nuclear Localization of mRNA in Beta Cells and Liver

To obtain a genome-wide catalog of genes in mammalian metabolic tissues that are potentially nuclearly retained, we extracted nuclear and cytoplasmic fractions from MIN6 pancreatic beta cell line (Miyazaki et al., 1990) and mouse liver and performed whole-transcriptome RNA sequencing (RNA-seq). We used single-molecule fluorescence in situ hybridization (smFISH) (Bahar Halpern et al., 2015; Itzkovitz et al., 2011; Lyubimova et al., 2013) of representative genes to convert the number of reads to estimates of cytoplasmic and nuclear mRNA numbers per cell (Tables S1 and S2). Our analysis revealed that most genes had more transcripts in the cytoplasm compared to the nucleus in MIN6 cells (mean

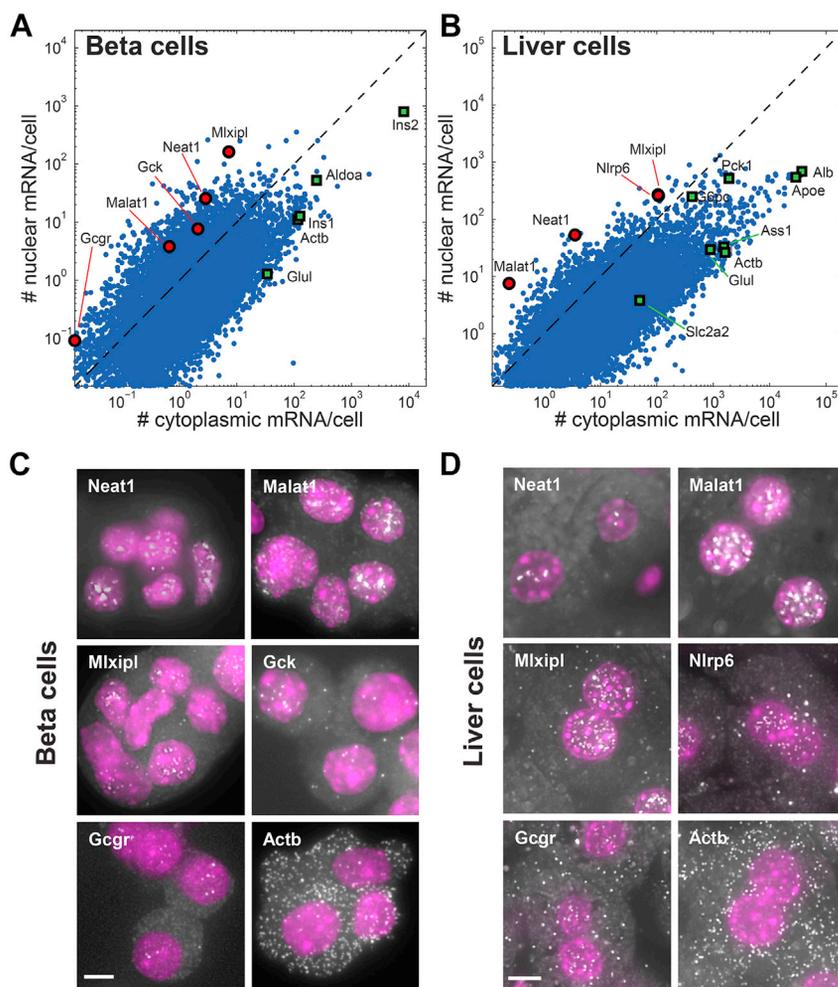


Figure 1. Deep Sequencing of Cellular Fractions Reveals Broad Nuclear Retention of mRNA

(A and B) RNA-seq of nuclear and cytoplasmic fractions of MIN6 cells (A) and mouse liver cells (B). Each dot represents a gene, x axis is the number of cytoplasmic mRNA molecules per cell, and y axis is the number of nuclear mRNA molecules. Dashed line represents the locus of genes that have equal numbers of nuclear and cytoplasmic mRNA copies. Green squares mark representative genes with higher cytoplasmic mRNA numbers, and red circles mark representative genes with higher nuclear mRNA numbers.

(C) Validation in primary pancreatic islet cells of some of the nuclear genes identified in (A) using smFISH. Images are maximum projections of 20 optical sections spaced 0.3 μm apart.

(D) Validation in intact liver frozen sections of some of the nuclear genes identified in (B) using smFISH. Images are maximum projections of eight optical sections spaced 0.3 μm apart. Scale bars, 5 μm (C and D). See also Figures S1 and S2.

cytoplasm/nucleus = 3.8 ± 0.05 , Figure 1A). Examples include the insulin genes *Ins1* (cytoplasm/nucleus = 13.2 ± 4.6) and *Ins2* (cytoplasm/nucleus = 10.2 ± 0.45), as well as house-keeping genes such as beta-actin (*Actb*, cytoplasm/nucleus = 10.6 ± 1.1).

A substantial fraction (30%) of the genes in MIN6 cells, however, had equal or higher levels of mRNA in the nucleus. These genes included the lncRNAs *Malat1* (cytoplasm/nucleus = 0.33 ± 0.27) and *Neat1* (cytoplasm/nucleus = 0.11 ± 0.02), as well as small nucleolar RNA (snoRNA; Weinstein and Steitz, 1999; Figure S1C). Interestingly, they also included key protein-coding genes such as Glucokinase (*Gck*, cytoplasm/nucleus = 0.29 ± 0.12), Glucagon receptor (*Gcgr*, cytoplasm/nucleus = 0.53 ± 0.46), and the transcription factor *Mlxipl*, also known as ChREBP (cytoplasm/nucleus = 0.05 ± 0.004 ; Herman et al., 2012; Postic et al., 2007; Uyeda and Repa, 2006; Figure 1A).

We next performed RNA-seq of nuclear and cytoplasmic fractions of liver cells isolated from mice (Figure 1B). Here, as well, we found that the majority of genes had predominantly cytoplasmic mRNA (mean cytoplasm/nucleus = 6.5 ± 1.3). As in MIN6 cells, however, a non-negligible 13.1% of protein-coding genes had more mRNA in the nucleus compared to the cyto-

plasm. Notably, *Mlxipl* was nucleary retained in this tissue as well (cytoplasm/nucleus = 0.38 ± 0.01 ; Figure 1B). Another notable nuclear gene was the inflammasome component nucleotide-binding oligomerization domain protein-like receptor 6 (*Nlrp6*; Anand et al., 2012; Elinav et al., 2011; Strowig et al., 2012; cytoplasm/nucleus = 0.41 ± 0.03 ; Figure 1B).

To validate the nuclear enrichment, we imaged individual mRNA molecules of representative genes in primary pancreatic

islets and in liver frozen sections using smFISH (Figures 1C and 1D). This revealed the absolute numbers and intra-cellular localizations of the transcripts of interest, clearly demonstrating that transcripts of *Gck*, *Gcgr*, *Nlrp6*, and *Mlxipl* were indeed substantially more numerous in the nucleus compared to the cytoplasm.

We used the RBPmap tool (Paz et al., 2014) to identify several putative target sites for known RNA-binding proteins in the 3' UTR of the most nucleary retained genes in both liver and MIN6 (Figure S2A; Table S4). Moreover, the 3' UTR sequences of the nucleary retained genes exhibited common sequence motifs that were not identified in a size-controlled group of the most cytoplasmic genes (Figure S2B). While genes with higher nuclear mRNA were enriched in lncRNA and snoRNA (Figures S1B and S1C; $p < 0.001$), the vast majority of genes with nuclear mRNA were protein-coding genes (Figure S1A). The median intron splicing efficiency for the nuclear mRNAs was $>95\%$ (Figure S1D; Table S3), and only 20% of the nuclear liver genes we identified have been shown to have intron detention (Boutz et al., 2015). Thus, our analysis revealed that a substantial fraction of genes in liver and MIN6 cells have higher levels of spliced, polyadenylated mRNA in the nucleus than in the cytoplasm.

Single-Molecule Transcript Imaging Reveals Increased Nuclear Retention for Key Protein-Coding Transcripts

Nuclear localization of mRNA does not necessarily imply increased nuclear lifetime, namely, low export rate of mRNA. High transcription rates combined with high cytoplasmic mRNA degradation rates can give rise to large numbers of nuclear mRNA and low numbers of cytoplasmic mRNA, even when nuclear export rate is high. To understand this effect, we considered a simple mathematical model describing the dynamics of nuclear (X) and cytoplasmic (Y) mRNAs. In this model, nuclear mRNA is produced at rate β , exported from the nucleus at rate λ , and degraded in the cytoplasm at rate δ (we considered only properly spliced mRNA for which nuclear degradation rate is negligible; [Garneau et al., 2007](#)).

$$dX/dt = \beta - \lambda \cdot X \quad (\text{Equation 1})$$

$$dY/dt = \lambda \cdot X - \delta \cdot Y \quad (\text{Equation 2})$$

[Equations 1](#) and [2](#) yield the following results for the levels of nuclear and cytoplasmic mRNAs at steady state (X_{st} , Y_{st}):

$$X_{st} = \beta / \lambda \quad (\text{Equation 3})$$

$$Y_{st} = \beta / \delta \quad (\text{Equation 4})$$

[Equations 3](#) and [4](#) indicate that the ratio between the amount of mRNA in the nucleus and that in the cytoplasm equals the ratio of the rates of cytoplasmic degradation and nuclear export ($X_{st}/Y_{st} = \delta/\lambda$). The amount of nuclear mRNA at steady state (X_{st}) increases with increasing transcription rates (β) and decreases with increasing export rate (λ). The ratio of transcription rate and total nuclear mRNA levels can thus be used to estimate nuclear export rates ($\lambda = \beta/X_{st}$).

To quantify nuclear export rates in situ, we developed a method to jointly quantify transcription rates (β) and total nuclear mRNA (X_{st}) ([Figure 2A](#)). We designed pairs of smFISH probe libraries targeting the introns and exons of the genes of interest and coupled them to two spectrally resolvable fluorophores, enabling quantification of the transcription rates, β ([Figure 2A](#); [Supplemental Experimental Procedures](#)). We also counted the total number of mRNA molecules in the nucleus (X_{st}) and used [Equation 3](#) to extract the nuclear export rate ($\lambda = \beta/X_{st}$). Similarly, we counted the number of cytoplasmic mRNA and used [Equation 4](#) to extract the cytoplasmic degradation rate ($\delta = \beta/Y_{st}$).

To validate our estimates of nuclear export rates, we sought to measure the temporal decline in nuclear mRNA following cessation of transcription. In such cases, nuclear mRNA should exponentially decline at rate λ . Since actinomycin D treatment on primary hepatocytes caused extensive perturbation to cellular physiology, we reverted to measure *G6pc*, a gene that is highly expressed in fasting mice but rapidly shuts down upon refeeding ([Bahar Halpern et al., 2015](#)). Indeed, we observed complete shut-down of transcription upon refeeding, as evident by the lack of double-labeled intronic-exonic nuclear dots at 15 and 30 min ([Figures 2B](#) and [2D](#)). Nuclear mRNA diluted at a rate of $5.3 \pm 1.24 \text{ hr}^{-1}$, consistent with our estimates of $\lambda = 4.99 \pm 0.99 \text{ hr}^{-1}$ obtained from in situ measurements of mice at the fasting state ([Fig-](#)

[ure 2E](#)). Similarly, our estimates of degradation rates were within 15% error of the validated values ([Bahar Halpern et al., 2015](#)).

We next applied this methodology to representative liver genes ([Figure 2F](#); [Table S5](#)). Most export rates were higher than the cytoplasmic degradation rates, and they conformed to previous estimates of nuclear lifetimes of a few minutes. Notably, however, *Nlrp6* and *Mlxipl* had substantially longer nuclear lifetimes of $1.98 \pm 0.96 \text{ hr}$ for *Nlrp6* and $0.75 \pm 0.37 \text{ hr}$ for *Mlxipl*. The nuclear export rates of *Mlxipl* and *Nlrp6* were also substantially lower than their cytoplasmic degradation rates (cytoplasmic lifetimes were $0.85 \pm 0.4 \text{ hr}$ for *Nlrp6* and $0.18 \pm 0.09 \text{ hr}$ for *Mlxipl*). For these genes, mRNA spends more time in the nucleus than in the cytoplasm.

Nuclear Localization of *Mlxipl* and *Nlrp6* mRNA in Diverse Tissues and Metabolic Conditions

Next, we turned to characterize the patterns of nuclear mRNA localization for *Mlxipl* and *Nlrp6*, two of the most prominent nuclear genes we uncovered. *Mlxipl* encodes the ChREBP transcription factor, a key regulator of lipogenic and glycolytic genes in metabolic tissues ([Herman et al., 2012](#); [Postic et al., 2007](#); [Uyeda and Repa, 2006](#)). We found that *Mlxipl* mRNA was predominantly nuclear in liver, intestine, and beta cells ([Figure 3](#)), as well as in different metabolic conditions, such as after intraperitoneal (i.p.) injection of glucose or insulin and following a high-fat diet (HFD) ([Figure S3A](#)). As controls, we measured the genes *Pck1* and *Actb* in the liver and the gene *Slc2a2* (also known as *Glut2*) in the intestinal epithelium. Unlike *Mlxipl*, these genes had substantially higher mRNA concentrations in the cytoplasm compared to the nucleus ([Figure 3](#)).

To examine whether nuclear retention of *Mlxipl* could be regulated by external conditions, we applied diverse stimuli on MIN6 cells and used smFISH to examine the patterns of *Mlxipl* mRNA nuclear localization. We found that *Mlxipl* remains nuclearly enriched following glucose challenges, insulin stimulation, heat shock, and serum starvation ([Figure S3B](#)).

Nlrp6, encoding a component of the inflammasome, which orchestrates diverse functions during homeostasis and inflammation including steady-state regulation of the composition and function of the intestinal microbiome ([Anand et al., 2012](#); [Elinav et al., 2011](#); [Strowig et al., 2012](#)), is expressed in both the liver and the intestinal epithelium. We found that *Nlrp6* transcripts were predominantly nuclear in the livers of mice fed a normal diet ([Figure 3](#)) or an HFD ([Figure S3A](#)). Unlike *Mlxipl*, which was nuclear in all tissues we examined, *Nlrp6* was cytoplasmic in the intestinal epithelium ([Figure 3](#)). To assess whether the cytoplasmic localization of intestinal *Nlrp6* mRNA is regulated by the intestinal microbiota, we examined germ-free (GF) mice, as well as colonized mice treated with wide-spectrum antibiotics for 4 weeks. Intestinal *Nlrp6* mRNA remained cytoplasmic in these conditions ([Figure S3C](#)). Thus, exposure to bacteria does not seem to be a cue for regulating nuclear export of intestinal *Nlrp6* mRNA.

Nuclear mRNA Co-localizes with Nuclear Speckles

[Spector \(2001\)](#) have shown that CTN-RNA is retained in the nucleus through sequestration to nuclear paraspeckles ([Prasanth et al., 2005](#)), sites of active RNA editing ([Chen and Carmichael,](#)

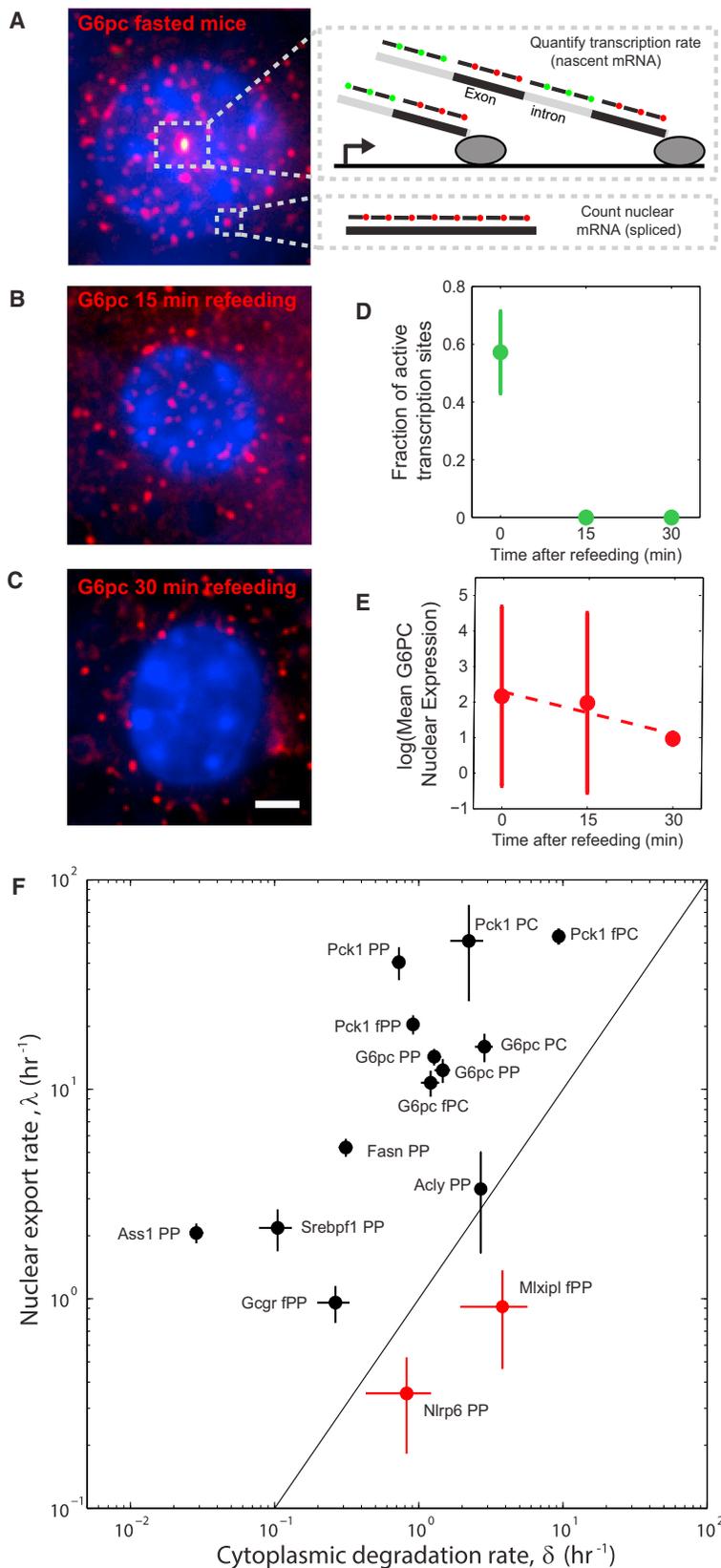


Figure 2. Single-Molecule Approach for Measuring Nuclear Export Rate and Cytoplasmic Lifetime

(A) Example shows identification of active transcription site of G6pc in liver cryosection from a fasting mouse using dual-color labeling of introns (green) and exons (red).

(B and C) G6pc nuclear levels rapidly decline 15 (B) and 30 min (C) after refeeding of fasting mice.

Images in (A)–(C) are single optical sections. Scale bar, 2 μm .

(D) Active transcription sites disappear 15 and 30 min after refeeding, indicating a complete shutdown of transcription.

(E) Quantification of the number of nuclear transcripts of G6pc at 5 hr fasting (time 0), as well as 15 and 30 min after refeeding. Data were divided by the expression at time 0 (145 mRNA per nucleus). Nuclear mRNA declined at a rate of $5.3 \pm 1.24 \text{ hr}^{-1}$, compatible with the in situ estimation based on the fasting state of $4.99 \pm 0.99 \text{ hr}^{-1}$ ($n = 2$ mice per time point).

(F) Degradation and nuclear export rates of liver genes estimated from in situ measurements in intact liver tissue. Solid line represents the locus of genes with equal rates of nuclear export and cytoplasmic degradation. Nlrp6 and Mxipl (marked in red) have significantly lower nuclear export rates. PC, pericentral; PP, periportal; f, fast. Error bars represent SEM.

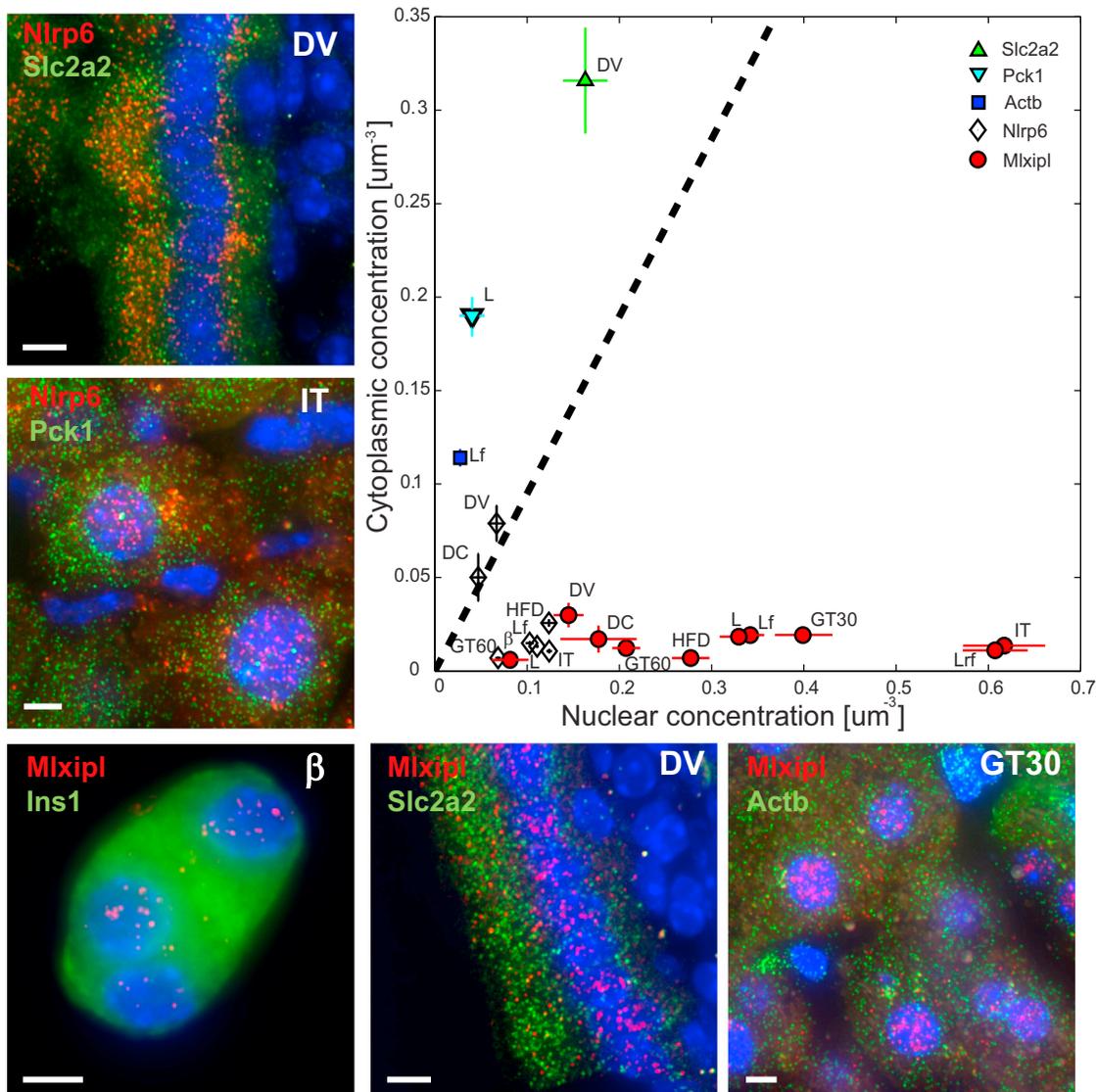


Figure 3. Mlxipl and Nlrp6 Are Highly Nuclearily Retained in Diverse Metabolic Tissues and Conditions

In contrast, Pck1, Actb, and Slc2a2 are highly enriched in the cytoplasm. Shown are the nuclear and cytoplasmic concentrations, as well as example images. DV, Duodenum, Villus; DC, Duodenum, Crypt; L, liver, ad libitum; Lf, liver, fasting; Lrf, liver, re-fed; IT, liver, insulin tolerance; GT30 (GT60), liver from mice sacrificed 30 min (60 min) after glucose injection; HFD, liver, high-fat diet; β , pancreatic beta cell. Images are maximum projections of 15–20 optical sections spaced 0.3 μm apart, respectively. Scale bar, 5 μm . Error bars represent SEM. See also Figures S3 and S4.

2009). To explore whether the retained genes found in our study are spatially correlated with nuclear domains, we performed dual-color smFISH of our nuclear genes and lncRNA markers of speckles (Malat1) and paraspeckles (Neat1). We used particle image cross-correlation spectroscopy (PICCS) (Semrau et al., 2011) to assess the spatial correlation, α , between the nuclear transcripts and either Malat1 or Neat1 foci (Figure S4). We found a highly significant spatial correlation between Malat1 foci and both Mlxipl ($\alpha = 0.178$, $p < 1e-15$) and Nlrp6 ($\alpha = 0.175 \pm 0.012$, $p < 1e-15$; Figure S4). Interestingly, Malat1 and Nlrp6 were not significantly correlated in the intestine, tissue in which Nlrp6 mRNA exhibited cytoplasmic localization ($\alpha = -0.026 \pm 0.037$, $p = 0.71$; Figure S4C). Mlxipl was also signifi-

cantly co-localized with Malat1 in beta cells ($\alpha = 0.12$, $p = 0.003$) and in the intestine ($\alpha = 0.121$, $p = 0.002$). In contrast, mRNA of ATP citrate lyase (Acl), which was not nuclearily retained (Figure 2F), was not co-localized with speckles and none of the genes tested were spatially correlated with paraspeckles in liver tissue (Figure S4). These results indicate that preferential binding or detention of mRNA of Mlxipl and Nlrp6 in nuclear speckles could facilitate their nuclear retention.

Nuclear mRNA Retention Can Reduce Cytoplasmic Gene Expression Noise

What could be the role of nuclear retention of mature mRNA? At first glance nuclear retention seems like an inefficient strategy for

regulating gene expression, as most of the RNA molecules do not reside in the cytoplasmic compartment where they should function. A possible advantage of nuclear retention could involve robustness to noise generated by stochastic mRNA production (Battich et al., 2015; Singh and Bokes, 2012; Xiong et al., 2010). Transcription in a wide range of organisms, including mammals, has been shown to be a pulsatile process, consisting of stochastic bursts of production followed by periods of promoter quiescence (Figure 4A; Larson et al., 2011; Darzacq et al., 2007; Suter et al., 2011; Bahar Halpern et al., 2015; Dar et al., 2012; Senecal et al., 2014). Bursty transcription can lead to profound variations in cellular mRNA content, a phenomenon termed gene expression noise. When promoters are in a transcriptionally active state, the cell accumulates mRNA, whereas when the promoters switch to an off state, mRNA levels decline (Figure 4B). Compartmentalization of mRNA could potentially reduce these burst-associated fluctuations in cytoplasmic mRNA concentrations, the fluctuations that eventually propagate to protein levels.

To assess the potential noise-reduction feature of low nuclear export rate on cytoplasmic variability, we performed Gillespie simulations (Gillespie, 1977) of a bursty promoter that stochastically transitions between on and off states at rates k_{OFF} and k_{ON} , producing transcripts at rate μ only when the promoter is on (Raj et al., 2006). While nuclear mRNA levels fluctuated in line with the promoter dynamics, cytoplasmic levels exhibited damped fluctuations compared to those expected when nuclear export was immediate (Figures 4A and 4B). The coefficient of variation (CV) of cytoplasmic transcripts reduced substantially when nuclear export rates were lower than the cytoplasmic mRNA degradation rates (Figures 4B–4D). Thus, reduced nuclear export rate can decrease cytoplasmic variability without changing the average cytoplasmic mRNA level (as evident by Equation 4), at the expense of accumulating more nuclear transcripts (Figure 4D).

Nuclear Retention of Mlxipl and Nlrp6 mRNAs Reduces Their Cytoplasmic Gene Expression Noise

Assessing whether nuclear retention buffers cytoplasmic gene expression noise requires comparing the observed single-cell distribution of cytoplasmic mRNA for a nuclearly retained gene with the distribution that would be expected if nuclear export were immediate. To this end, we used our previously reported method (Bahar Halpern et al., 2015; Bahar Halpern and Itzkovitz, 2015) to identify transcription sites and quantify their bursting dynamics in the intact liver lobule (Figure 2) for the nuclearly retained genes Mlxipl and Nlrp6. We found that both genes were expressed in a bursty manner; 43% of Mlxipl sites were actively transcribing at any given moment and had on average $M = 38 \pm 6$ polymerase molecules, a number that was too high to be compatible with a non-bursty transcription model (Figures 5 and S5). Similarly, Nlrp6 exhibited rare transcription sites with only 17% of sites transcriptionally active at any given moment and an average occupancy of $M = 5 \pm 2$ polymerase molecules.

Next, we fitted the model of Raj et al. (2006; Bahar Halpern et al., 2015) to the nuclear mRNA distributions to extract the rates of promoter opening and closing, k_{ON} and k_{OFF} . The distributions of nuclear mRNA for both genes were well fitted by a two-state bursty model (Figures 5A and 5C). In contrast, cytoplasmic

mRNA levels for both Mlxipl and Nlrp6 were significantly narrower, compared with the distribution expected based on the same burst parameters but immediate export (Figures 5B and 5D; CV = 0.46 versus CV = 0.56 for Mlxipl, $p < 0.002$; CV = 0.42 versus CV = 0.55 for Nlrp6, $p < 0.0001$; Supplemental Experimental Procedures). Cytoplasmic mRNA noise level of Pck1, for which export rate was substantially higher than the cytoplasmic degradation rate (Figure 2F), was identical to the noise predicted based on the fitted two-state bursty model (CV = 0.62 versus 0.56, $p = 0.91$; Figure S5F). Thus, nuclear retention of mRNA decreases cytoplasmic gene expression noise emanating from promoter bursts, when the mRNA is retained in the nucleus for time periods that exceed its cytoplasmic lifetime.

DISCUSSION

Our experiments revealed a surprisingly wide range of genes in metabolic tissues for which fully spliced, polyadenylated mRNA molecules are retained in the nucleus for time periods that exceed their cytoplasmic lifetimes. Since mRNAs are transcribed and processed at the sites of transcription and translated in the cytoplasm, this lengthy retention period raises the intriguing possibility that the nucleus may have previously overlooked roles.

The nuclear retention of the genes we followed up on in our study (Mlxipl and Nlrp6) appeared to be constitutive, rather than regulated, at least for the stimuli we applied. These included acute exposure to glucose, a condition that has been shown to elicit a potent response from the ChREBP protein (Herman et al., 2012; Postic et al., 2007; Uyeda and Repa, 2006), but which did not yield higher cytoplasmic mRNA levels. In addition, exposure to the intestinal microbiota, a potential regulator of Nlrp6, does not seem to be the stimulus responsible for the tissue-specific cytoplasmic localization of Nlrp6 mRNA in the intestine, but not the liver. It would be important to test additional stimuli that might give rise to differential nuclear retention for other genes we identified in our study.

The ubiquitous nuclear enrichment of transcripts of Mlxipl and Nlrp6 under diverse conditions prompted us to consider additional roles for lengthy mRNA nuclear retention periods. Gene expression in unicellular organisms, as well as in mammalian tissues, consists of transcriptional bursts that can generate profound variability in mRNA levels among identical cells and in a given cell over time. While several papers demonstrated the advantage of this variability as a bet-hedging strategy in unicellular organisms (Chalancon et al., 2012; Eldar and Elowitz, 2010), it is yet unclear if variability is advantageous in mammalian tissues or simply a by-product of the promoter bursting dynamics. Fundamental processes in gene expression can either reduce or amplify burst-associated noise. Lifetimes of mRNA and proteins are key in modulating this variability. Long-lived transcripts render the cell insensitive to the fluctuations in mRNA production by temporally averaging several burst events. Extended protein lifetimes also can achieve a similar effect of time-averaging of fluctuations in mRNA levels, even when cytoplasmic mRNA lifetimes are short (Raj et al., 2006). Nuclear retention has a similar effect, since the nucleus averages the stochastic promoter

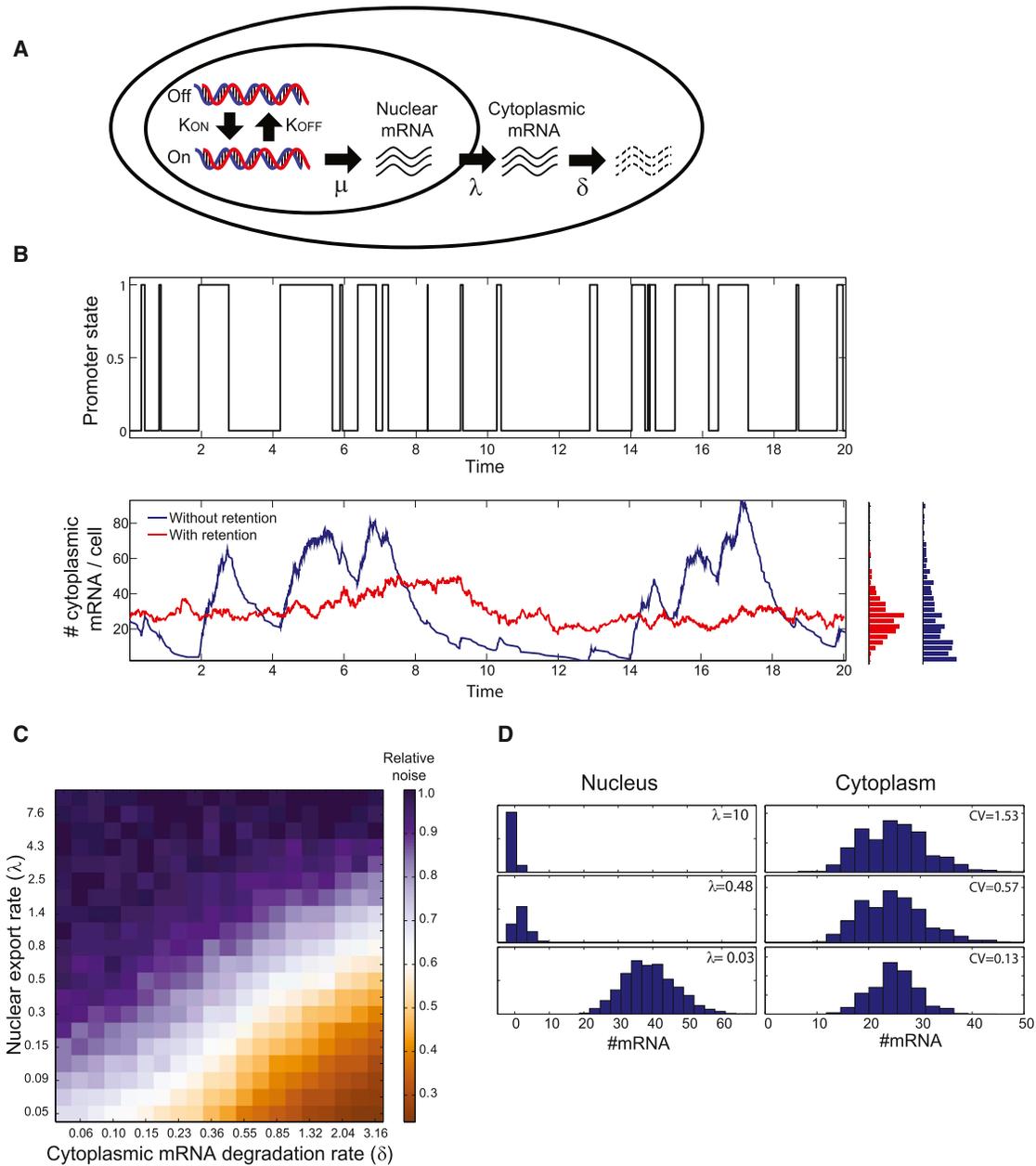


Figure 4. Nuclear Retention Can Buffer Gene Expression Noise

(A) Schematic diagram of a two-state burst model that includes a nuclear retention phase. Promoters switch between off and on states at rates k_{ON} and k_{OFF} , generating mRNA from the on state at rate μ .

(B) Nuclear retention significantly reduces cytoplasmic variability. Stochastic simulation of a bursty promoter with $k_{ON} = 1 \text{ hr}^{-1}$, $k_{OFF} = 3 \text{ hr}^{-1}$, $\mu = 100 \text{ mRNA/hr}$, and $\delta = 1 \text{ hr}^{-1}$ is shown. (Top) Plot of promoter state versus time is shown. (Bottom) Time course of mRNA levels in the cytoplasm without (blue) and with (red) nuclear retention ($\lambda = 0.2 \text{ hr}^{-1}$) is shown. Side histograms demonstrate a substantially lower cytoplasmic variability of $CV = 0.33$ with retention compared to $CV = 0.91$ without retention (CV, coefficient of variation).

(C) Cytoplasmic noise decreases with increased nuclear retention (decreased export rate λ). Heatmap of CVs for different combinations of degradation rates and nuclear export rates is shown. Values are normalized to the maximal CV for each degradation rate. Values are averages of 2,000 stochastic simulations of a bursty promoter with $k_{ON} = 1 \text{ hr}^{-1}$ and $k_{OFF} = 3 \text{ hr}^{-1}$. Transcription rate was set to $\mu = 100 \cdot \delta$ so that average cytoplasmic levels were 25 mRNA for all combinations.

(D) Histograms of nuclear (left) and cytoplasmic (right) mRNA levels. As nuclear retention increases, average cytoplasmic levels remain identical but noise is decreased. Cytoplasmic degradation was $\delta = 3.16$ for all simulations.

bursts. An attractive feature of nuclear retention is that it can decrease cytoplasmic noise without affecting the average steady-state levels (Equation 4). In contrast, noise reduction

through lengthened mRNA or protein lifetimes requires fine-tuning of the rates of transcription or translation, respectively, to maintain the same steady-state levels.

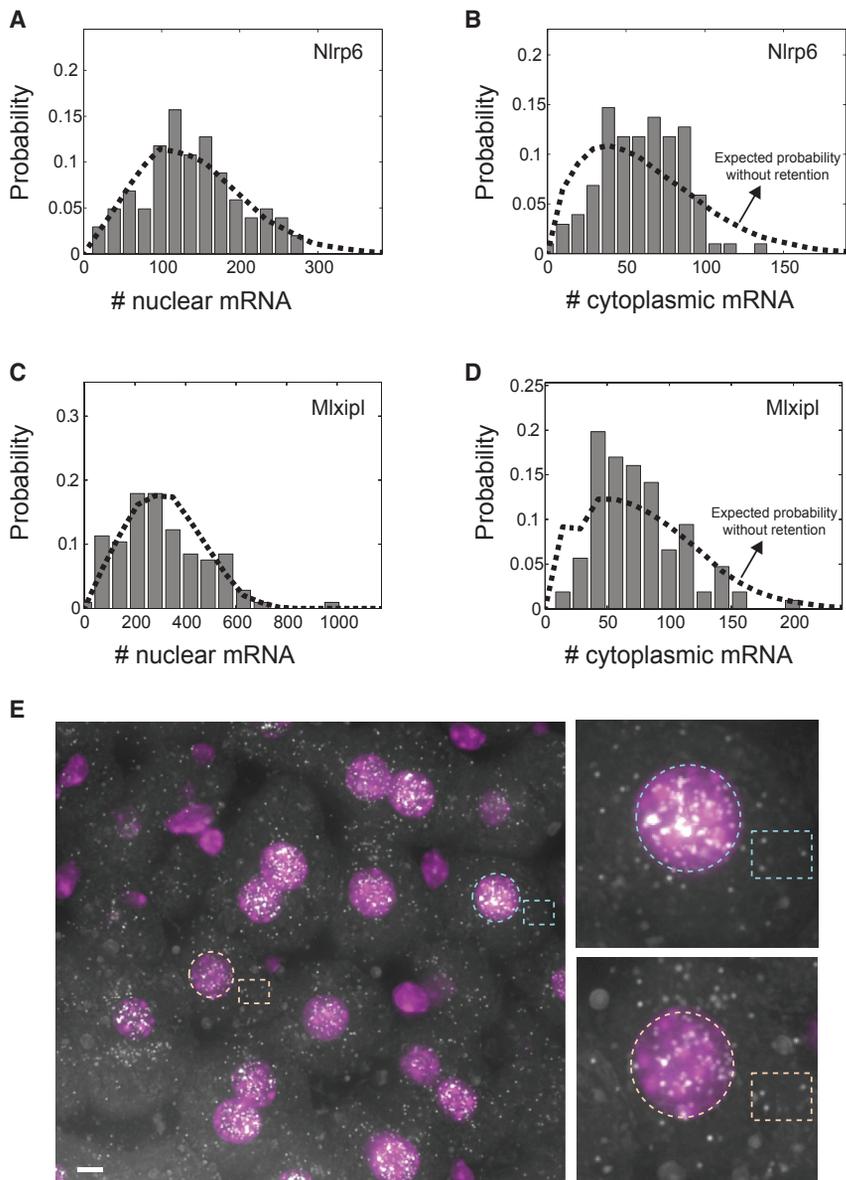


Figure 5. Nuclear Retention of Nlrp6 and Mlxipl Reduces Cytoplasmic Gene Expression Noise

(A–D) Probability distributions of mRNA levels in the nucleus (A and C) and cytoplasm (B and D) of hepatocytes measured in the intact mouse liver. Dashed lines are the fits of a two-state bursty model to the nucleus (A and C) and indicate the expected probability distributions of cytoplasmic mRNA if export was immediate (B and D). Fitted burst parameters were $k_{ON} = 0.48 \text{ hr}^{-1}$ and $k_{OFF} = 2.34 \text{ hr}^{-1}$ for Nlrp6 and $k_{ON} = 0.17 \text{ hr}^{-1}$ and $k_{OFF} = 0.23 \text{ hr}^{-1}$ for Mlxipl. While the two-state model fits the nuclear mRNA distributions (A and C), the measured cytoplasmic distributions are significantly narrower compared to the distributions expected based on the promoter bursting dynamics and no nuclear retention (B, D, and E).

(E) Example of Mlxipl expression in liver section from fasting mouse. Dashed yellow and blue circles label nuclei of two tetraploid hepatocytes with variable mRNA content, and dashed yellow and cyan boxes label their cytoplasmic areas, demonstrating the low variability in cytoplasmic concentration. Scale bar, 5 μm . Image is maximum projection of 15 optical sections spaced 0.3 μm apart. See also Figure S5.

tissues were harvested and fixed as described previously (Bahar Halpern et al., 2015) and in the Supplemental Experimental Procedures. Primary pancreatic islets were isolated from 6- to 8-week-old mice, cultured up to 1 day, and fixed in 4% paraformaldehyde for 15 min (Supplemental Experimental Procedures). At least two mice were analyzed for each time point and condition.

Cell Fractionation and RNA-Seq

Fractionation of nuclear and cytoplasmic liver RNAs was performed according to Menet et al. (2012), except for minor modifications (Supplemental Experimental Procedures). Fractionation of nuclear and cytoplasmic RNAs from MIN6 cell line (passage 30) is described in the Supplemental Experimental Procedures. RNA-seq was performed using Illumina HiSeq 2500. Read analysis is described in detail in the Supplemental Experimental Procedures. MIN6

RNA-seq results were based on RNA extractions in two independent experiments. Liver RNA-seq was performed on two fasting mice independently processed and analyzed.

Hybridization and Imaging

Probe library constructions, hybridization procedures, and imaging conditions were described previously (Itzkovitz et al., 2011; Lyubimova et al., 2013).

Computational Procedures

To assess the nuclear export rates, cytoplasmic degradation rates, and burst parameters, we used our previously reported method (Bahar Halpern et al., 2015; Bahar Halpern and Itzkovitz, 2015). We detected active transcription sites of the genes of interest based on dots that appeared in both the intronic and exonic channels. The burst fraction f , transcription rate from active transcription sites μ , and overall transcription rate per cell β were calculated as described in the Supplemental Experimental Procedures. The bursting rates k_{OFF} and k_{ON} were computed by fitting the model of Raj et al. (2006; Supplemental Experimental Procedures). To assess the noise that would be observed

Given the wide range of nuclearly retained mRNAs described here (13% and 30% in liver and beta cells, respectively), it seems that nuclear retention of mRNA is a meaningful, previously underappreciated step in the mRNA life cycle. Nuclear retention likely has diverse mechanisms and roles. Our study opens the way to exploring this unique mode of gene regulation in diverse physiological and pathological states.

EXPERIMENTAL PROCEDURES

Mice and Tissues

C57bl6 male mice (5 months old) were fed normal chow ad libitum, fasted, or re-fed for the indicated times. HFD was applied to 2-month-old mice for a duration of 8 weeks. Mice were stimulated with insulin and glucose by i.p. injection 30 or 60 min prior to sacrifice. GF C57bl6 mice were housed in sterile isolators. For the antibiotic treatment, mice were given a combination of antibiotics in their drinking water (Supplemental Experimental Procedures). Tis-

without nuclear retention, we used Equation S1 (Supplemental Experimental Procedures) with δ and our inferred k_{ON} and k_{OFF} .

ACCESSION NUMBERS

The accession number for all sequencing data reported in this paper is GEO: GSE73977.

SUPPLEMENTAL INFORMATION

Supplemental Information includes Supplemental Experimental Procedures, five figures, and five tables and can be found with this article online at <http://dx.doi.org/10.1016/j.celrep.2015.11.036>.

AUTHOR CONTRIBUTIONS

Conceptualization, K.B.H., I.C., D.L., and S.I.; Methodology, K.B.H., D.L., I.C., I.U., M.L., S.L., E.E., and S.I.; Software, D.L., I.U., and S.I.; Validation, K.B.H., I.C., and M.L.; Formal Analysis, D.L., I.U., and S.I.; Investigation, K.B.H. and I.C.; Data Curation, D.L., I.U., and S.I.; Writing – Original Draft, K.B.H., D.L., and S.I.; Writing – Review & Editing, K.B.H., D.L., and S.I.; Funding Acquisition, S.I.; Supervision, S.I.

ACKNOWLEDGMENTS

We thank Yuval Dor, Agnes Klochendler, and Jerome Menet for help with reagents and experimental protocols and Phillip Sharp and Paul Boutz for the liver intron retention data. We thank Shlomit Gilad and Inbal Paz for technical help and Nir Drayman and all members of our lab for valuable comments on the manuscript. E.E. is supported by Yael and Rami Ungar, Leona M. and Harry B. Helmsley Charitable Trust, the Gurwin Family Fund for Scientific Research, Crown Endowment Fund for Immunological Research, estate of Jack Gitlitz, estate of Lydia Hershkovich, the Benozio Endowment Fund for the Advancement of Science, Adelis Foundation, John L. and Vera Schwartz, Alan Markovitz, Cynthia Adelson, Centre National de la Recherche Scientifique (CNRS), estate of Samuel and Alwyn J. Weber, and Mr. and Mrs. Donald L. Schwarz. E.E. is the incumbent of the Rina Gudinski Career Development Chair. I.U. is supported by the Sygnet Career Development Chair for Bioinformatics by grants from the Israeli Science Foundation (1242/14 and 1984/14), the I-CORE Program of the Planning and Budgeting Committee and the Israel Science Foundation (grant 1796/12), the Minerva Foundation, the Fritz-Thyssen Foundation, the Rising Tide Foundation, and by a research grant from The Abramson Family Center for Young Scientists. S.I. is supported by the Henry Chanoch Kreuter Institute for Biomedical Imaging and Genomics, The Leir Charitable Foundations, Richard Jakubskind Laboratory of Systems Biology, Cymerman-Jakubskind Prize, The Lord Sieff of Brimpton Memorial Fund, The Human Frontiers Science Program, the I-CORE program of the Planning and Budgeting Committee and the Israel Science Foundation (grants 1902/12 and 1796/12), and the European Research Council (ERC) under the European Union's Seventh Framework Programme (FP7/2007-2013)/ERC grant agreement 335122. S.I. is the incumbent of the Philip Harris and Gerald Ronson Career Development Chair.

Received: June 18, 2015

Revised: September 15, 2015

Accepted: November 10, 2015

Published: December 17, 2015

REFERENCES

Anand, P.K., Malireddi, R.K.S., Lukens, J.R., Vogel, P., Bertin, J., Lamkanfi, M., and Kanneganti, T.-D. (2012). NLRP6 negatively regulates innate immunity and host defence against bacterial pathogens. *Nature* **488**, 389–393.

Bahar Halpern, K., Tanami, S., Landen, S., Chapal, M., Szlak, L., Hutzler, A., Nizhberg, A., and Itzkovitz, S. (2015). Bursty gene expression in the intact mammalian liver. *Mol. Cell* **58**, 147–156.

Bahar Halpern, K., and Itzkovitz, S. (2015). Single molecule approaches for quantifying transcription and degradation rates in intact mammalian tissues. *Methods* **15**, 30155–30159.

Battich, N., Stoeger, T., and Pelkmans, L. (2015). Control of transcript variability in single mammalian cells. *Cell* **163**, 1596–1610.

Bhatt, D.M., Pandya-Jones, A., Tong, A.-J., Barozzi, I., Lissner, M.M., Natoli, G., Black, D.L., and Smale, S.T. (2012). Transcript dynamics of proinflammatory genes revealed by sequence analysis of subcellular RNA fractions. *Cell* **150**, 279–290.

Blake, W.J., Kaern, M., Cantor, C.R., and Collins, J.J. (2003). Noise in eukaryotic gene expression. *Nature* **422**, 633–637.

Boutz, P.L., Bhutkar, A., and Sharp, P.A. (2015). Detained introns are a novel, widespread class of post-transcriptionally spliced introns. *Genes Dev.* **29**, 63–80.

Chalancón, G., Ravarani, C.N.J., Balaji, S., Martínez-Arias, A., Aravind, L., Jothi, R., and Babu, M.M. (2012). Interplay between gene expression noise and regulatory network architecture. *Trends Genet.* **28**, 221–232.

Chen, L.-L., and Carmichael, G.G. (2009). Altered nuclear retention of mRNAs containing inverted repeats in human embryonic stem cells: functional role of a nuclear noncoding RNA. *Mol. Cell* **35**, 467–478.

Dar, R.D., Razoooky, B.S., Singh, A., Trimeloni, T.V., McCollum, J.M., Cox, C.D., Simpson, M.L., and Weinberger, L.S. (2012). Transcriptional burst frequency and burst size are equally modulated across the human genome. *Proc. Natl. Acad. Sci. USA* **109**, 17454–17459.

Darzacq, X., Shav-Tal, Y., de Turrís, V., Brody, Y., Shenoy, S.M., Phair, R.D., and Singer, R.H. (2007). In vivo dynamics of RNA polymerase II transcription. *Nat. Struct. Mol. Biol.* **14**, 796–806.

Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., et al. (2012). Landscape of transcription in human cells. *Nature* **489**, 101–108.

Eldar, A., and Elowitz, M.B. (2010). Functional roles for noise in genetic circuits. *Nature* **467**, 167–173.

Elinav, E., Strowig, T., Kau, A.L., Henao-Mejia, J., Thaiss, C.A., Booth, C.J., Peaper, D.R., Bertin, J., Eisenbarth, S.C., Gordon, J.I., and Flavell, R.A. (2011). NLRP6 inflammasome regulates colonic microbial ecology and risk for colitis. *Cell* **145**, 745–757.

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

Gillespie, D.T. (1977). Exact stochastic simulation of coupled chemical reactions. *J. Phys. Chem.* **81**, 2340–2361.

Golding, I., Paulsson, J., Zawilski, S.M., and Cox, E.C. (2005). Real-time kinetics of gene activity in individual bacteria. *Cell* **123**, 1025–1036.

Herman, M.A., Peroni, O.D., Villoria, J., Schön, M.R., Abumrad, N.A., Blüher, M., Klein, S., and Kahn, B.B. (2012). A novel ChREBP isoform in adipose tissue regulates systemic glucose metabolism. *Nature* **484**, 333–338.

Itzkovitz, S., Lyubimova, A., Blat, I.C., Maynard, M., van Es, J., Lees, J., Jacks, T., Clevers, H., and van Oudenaarden, A. (2011). Single-molecule transcript counting of stem-cell markers in the mouse intestine. *Nat. Cell Biol.* **14**, 106–114.

Kaern, M., Elston, T.C., Blake, W.J., and Collins, J.J. (2005). Stochasticity in gene expression: from theories to phenotypes. *Nat. Rev. Genet.* **6**, 451–464.

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

Lyubimova, A., Itzkovitz, S., Junker, J.P., Fan, Z.P., Wu, X., and van Oudenaarden, A. (2013). Single-molecule mRNA detection and counting in mammalian tissue. *Nat. Protoc.* **8**, 1743–1758.

Maheshri, N., and O'Shea, E.K. (2007). Living with noisy genes: how cells function reliably with inherent variability in gene expression. *Annu. Rev. Biophys. Biomol. Struct.* **36**, 413–434.

- Menet, J.S., Rodriguez, J., Abruzzi, K.C., and Rosbash, M. (2012). Nascent-Seq reveals novel features of mouse circadian transcriptional regulation. *eLife* 1, e00011.
- Miyazaki, J., Araki, K., Yamato, E., Ikegami, H., Asano, T., Shibasaki, Y., Oka, Y., and Yamamura, K. (1990). Establishment of a pancreatic β cell line that retains glucose-inducible insulin secretion: special reference to expression of glucose transporter isoforms. *Endocrinology* 127, 126–132.
- Oeffinger, M., and Zenklusen, D. (2012). To the pore and through the pore: a story of mRNA export kinetics. *Biochim. Biophys. Acta* 1819, 494–506.
- Pandya-Jones, A., Bhatt, D.M., Lin, C.-H., Tong, A.-J., Smale, S.T., and Black, D.L. (2013). Splicing kinetics and transcript release from the chromatin compartment limit the rate of Lipid A-induced gene expression. *RNA* 19, 811–827.
- Paulsson, J. (2004). Summing up the noise in gene networks. *Nature* 427, 415–418.
- Paz, I., Kosti, I., Ares, M., Jr., Cline, M., and Mandel-Gutfreund, Y. (2014). RBPmap: a web server for mapping binding sites of RNA-binding proteins. *Nucleic Acids Res.* 42, W361–W367.
- Postic, C., Dentin, R., Denechaud, P.-D., and Girard, J. (2007). ChREBP, a transcriptional regulator of glucose and lipid metabolism. *Annu. Rev. Nutr.* 27, 179–192.
- Prasanth, K.V., Prasanth, S.G., Xuan, Z., Hearn, S., Freier, S.M., Bennett, C.F., Zhang, M.Q., and Spector, D.L. (2005). Regulating gene expression through RNA nuclear retention. *Cell* 123, 249–263.
- Raj, A., and van Oudenaarden, A. (2008). Nature, nurture, or chance: stochastic gene expression and its consequences. *Cell* 135, 216–226.
- Raj, A., Peskin, C.S., Tranchina, D., Vargas, D.Y., and Tyagi, S. (2006). Stochastic mRNA synthesis in mammalian cells. *PLoS Biol.* 4, e309.
- Semrau, S., Holtzer, L., González-Gaitán, M., and Schmidt, T. (2011). Quantification of biological interactions with particle image cross-correlation spectroscopy (PICCS). *Biophys. J.* 100, 1810–1818.
- Senecal, A., Munsky, B., Proux, F., Ly, N., Braye, F.E., Zimmer, C., Mueller, F., and Darzacq, X. (2014). Transcription factors modulate c-Fos transcriptional bursts. *Cell Rep.* 8, 75–83.
- Shalgi, R., Hurt, J.A., Lindquist, S., and Burge, C.B. (2014). Widespread inhibition of posttranscriptional splicing shapes the cellular transcriptome following heat shock. *Cell Rep.* 7, 1362–1370.
- Shav-Tal, Y., Darzacq, X., Shenoy, S.M., Fusco, D., Janicki, S.M., Spector, D.L., and Singer, R.H. (2004). Dynamics of single mRNPs in nuclei of living cells. *Science* 304, 1797–1800.
- Singh, A., and Bokes, P. (2012). Consequences of mRNA transport on stochastic variability in protein levels. *Biophys. J.* 103, 1087–1096.
- Spector, D.L. (2001). Nuclear domains. *J. Cell Sci.* 114, 2891–2893.
- Strowig, T., Henao-Mejia, J., Elinav, E., and Flavell, R. (2012). Inflammasomes in health and disease. *Nature* 481, 278–286.
- Suter, D.M., Molina, N., Gatfield, D., Schneider, K., Schibler, U., and Naef, F. (2011). Mammalian genes are transcribed with widely different bursting kinetics. *Science* 332, 472–474.
- Tilgner, H., Knowles, D.G., Johnson, R., Davis, C.A., Chakraborty, S., Djebali, S., Curado, J., Snyder, M., Gingeras, T.R., and Guigó, R. (2012). Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. *Genome Res.* 22, 1616–1625.
- Uyeda, K., and Repa, J.J. (2006). Carbohydrate response element binding protein, ChREBP, a transcription factor coupling hepatic glucose utilization and lipid synthesis. *Cell Metab.* 4, 107–110.
- Vargas, D.Y., Raj, A., Marras, S.A.E., Kramer, F.R., and Tyagi, S. (2005). Mechanism of mRNA transport in the nucleus. *Proc. Natl. Acad. Sci. USA* 102, 17008–17013.
- Weinstein, L.B., and Steitz, J.A. (1999). Guided tours: from precursor snoRNA to functional snoRNP. *Curr. Opin. Cell Biol.* 11, 378–384.
- Xiong, L.P., Ma, Y.Q., and Tang, L.H. (2010). Attenuation of transcriptional bursting in mRNA transport. *Phys. Biol.* 7, 016005.

Cell Reports

Supplemental Information

Nuclear Retention of mRNA in Mammalian Tissues

Keren Bahar Halpern, Inbal Caspi, Doron Lemze, Maayan Levy, Shanie Landen, Eran Elinav, Igor Ulitsky, and Shalev Itzkovitz

SUPPLEMENTAL TABLES AND FIGURES

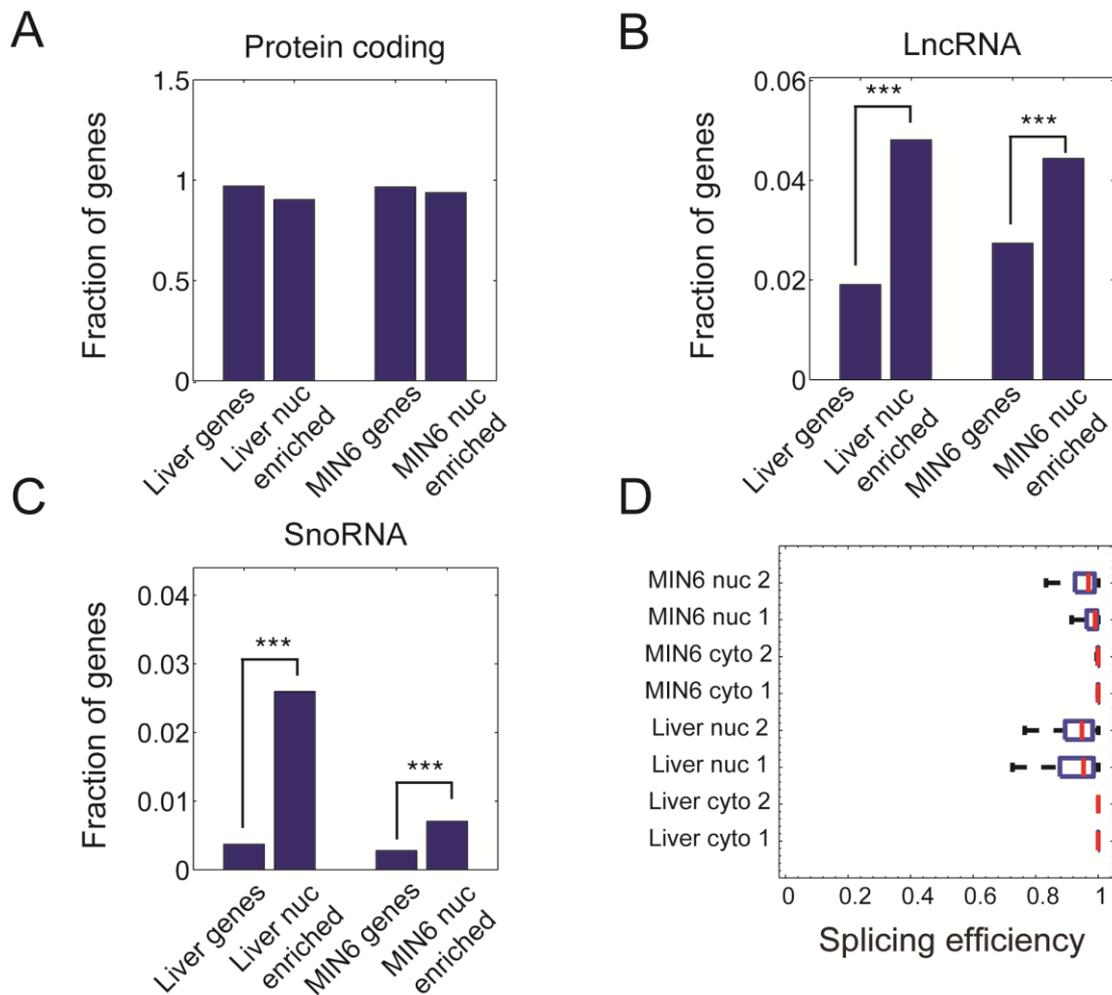
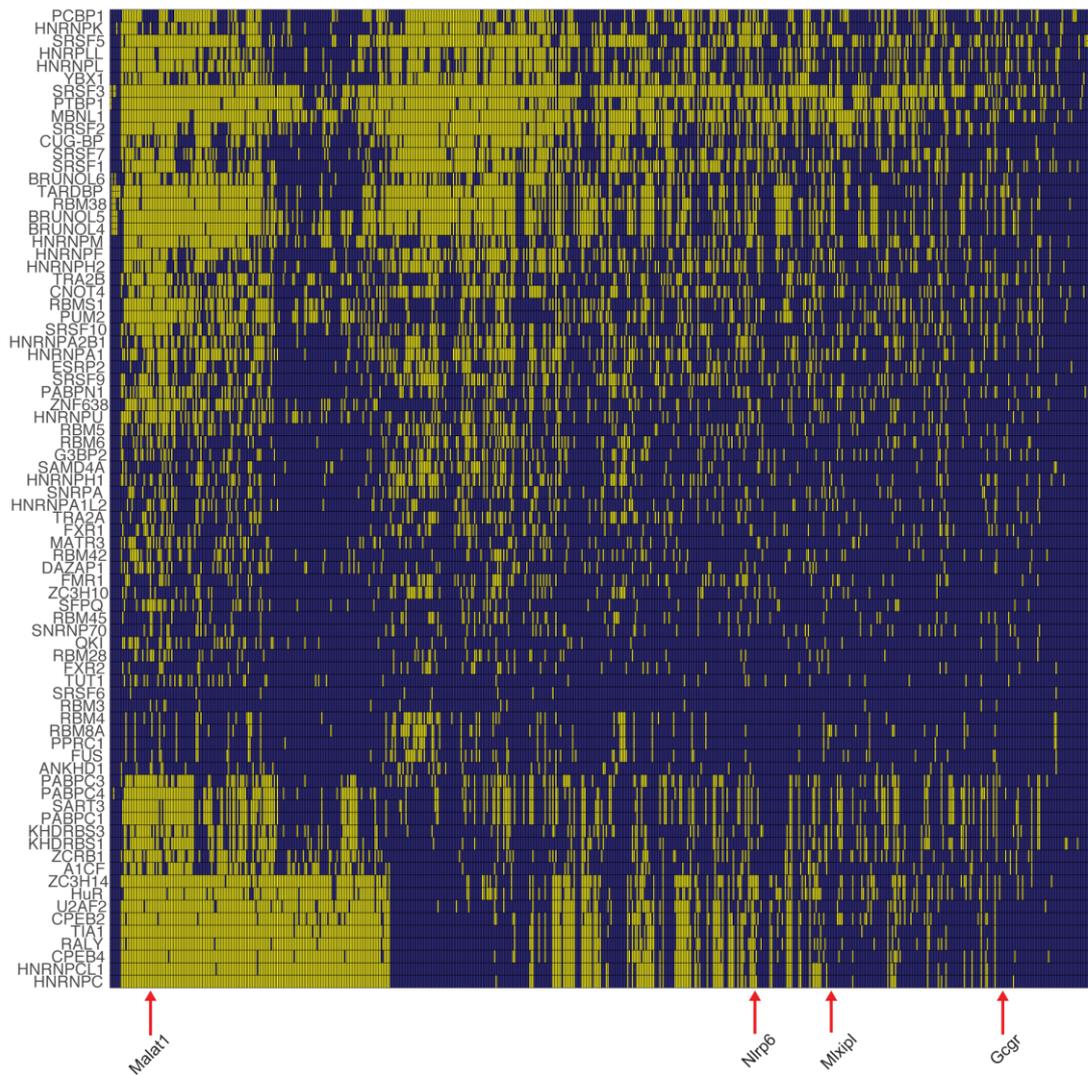


Figure S1 (Related to Figure 1) – RNAseq of nuclear and cytoplasmic fractions. (A) Protein-coding genes have a comparable representation among the genes with higher numbers of mRNA in the nucleus compared to the cytoplasm (protein-coding genes are 97% of the entire transcriptome sequenced and 91% of the nuclear genes). (B) lncRNA are enriched among the nuclear genes (lncRNA are 2% of the transcriptome sequenced but 4.8% of the nuclear genes). (C) snoRNA are enriched among the nuclear genes (snoRNA are 0.4% of the transcriptome sequenced but 2.6% of the nuclear genes). Results for (A-C) include only genes with more than 1 copy per cell for the liver or 0.1 copy per cell for MIN6 cells. D – Nuclear poly-adenylated mRNA are predominantly spliced. *** $p < 0.001$.

A



B

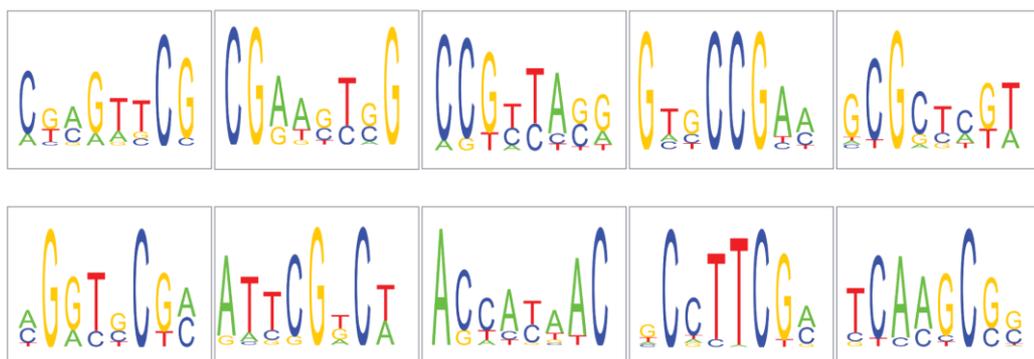
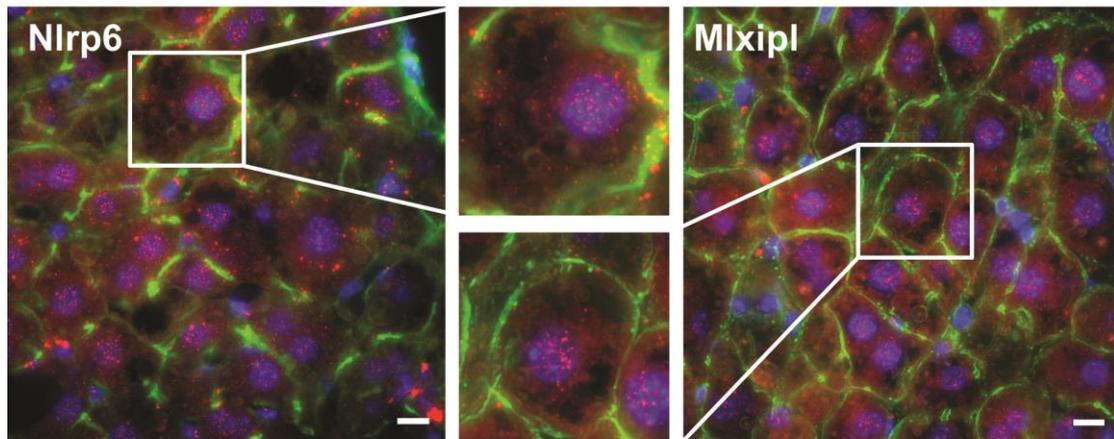


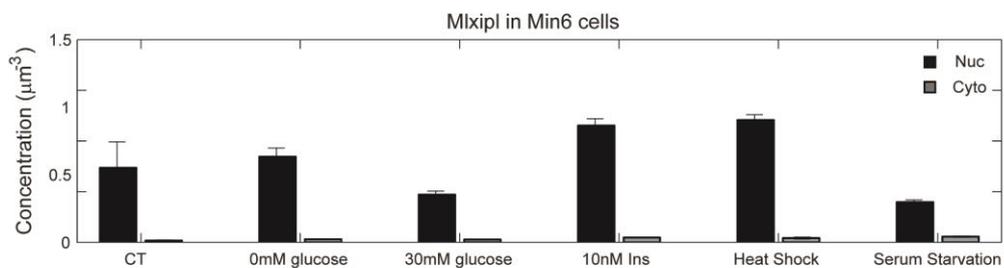
Figure S2 (related to Figure 1) – Putative RNA binding proteins and enriched motifs for the nuclearly enriched genes. (A) Putative binding interactions between known RNA binding proteins and the 3'UTR of the most nuclearly retained genes. Columns are the 654 most nuclearly retained genes, rows are RNA binding proteins from the RBPmap database (Paz et

al., 2014). For every RNA-binding protein and gene combination, yellow marks interactions for which the 3'UTR of the gene has at least one binding motif with $pval < 0.001$. (B) Sequence motifs found in the 3'UTR of the most nuclearly retained genes with the Amadeus software (Linhart et al., 2008).

A



B



C

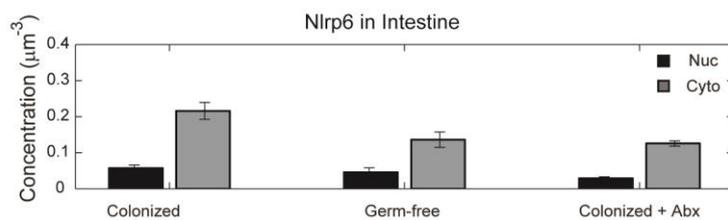


Figure S3 (related to Figure 3) – (A) Transcripts of Nlrp6 and Mlxipl remain nuclear in livers of mice after 8 weeks of high-fat diet. Red dots are single mRNA of Nlrp6 (left) or Mlxipl (right). Green - phalloidin-stained membranes, blue - DAPI-stained nuclei. (B) Mlxipl mRNA remains

nuclearly retained in diverse conditions in MIN6 treated cells, including exposure to high concentrations of glucose and insulin, heat shock and serum starvation. All analyses are for at least 30 cells per condition. (C) Intestinal Nlrp6 is cytoplasmic regardless of microbiota composition, Shown are the nuclear and cytoplasmic mRNA concentrations in control colonized mice, germ-free mice as well as colonized mice after four weeks of antibiotics treatment.

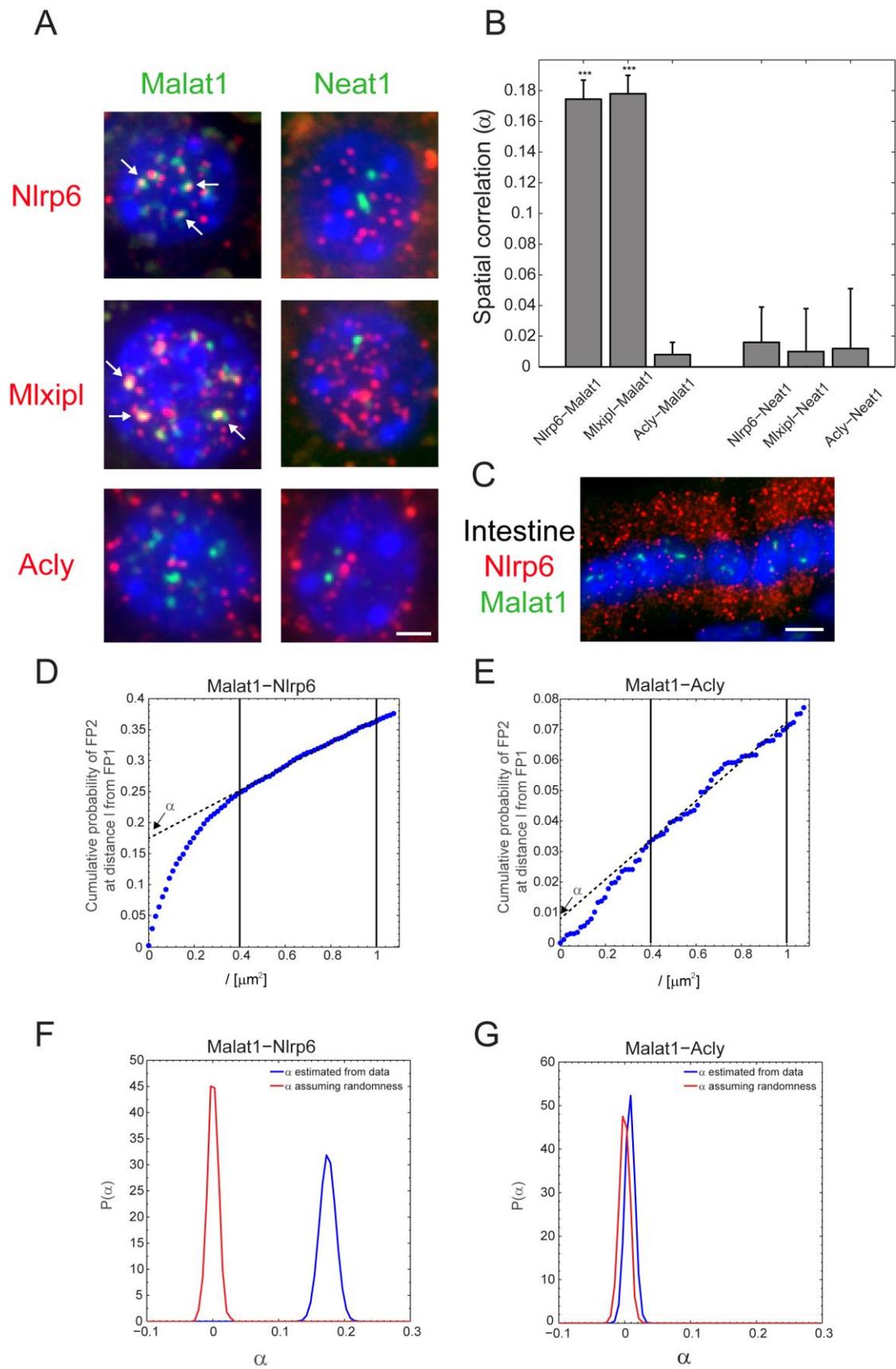


Figure S4 (related to Figure 3) – Nuclear transcripts of Mxipl and Nlrp6 co-localize with nuclear speckles. (A) Dual color smFISH for Malat1 (green dots left column) or Neat1 (green

dots right column) with Mlxipl, Nlrp6 and Acly (red dots). Images are maximal projections of 5 optical sections space $0.3\mu\text{m}$ apart. Scale bar is $2\mu\text{m}$ (B) Mlxipl and Nlrp6 co-localized with nuclear speckles. Shown are the spatial correlations (α) between the relevant genes and either Malat1 or Neat1. *** $p\text{val}<1\text{e-}15$. (C) Nlrp6 nuclear transcripts do not co-localize with nuclear speckles in the intestine. Dual color smFISH for Malat1 (green dots) Nlrp6 (red dots) in the intestinal epithelium. Scale bar is $5\mu\text{m}$ (D-G) PICCS method for estimating spatial co-localization of mRNA with nuclear domains. (D, E) Cumulative probability functions of observing an FP2 particle (Nlrp6 in (D), Acly in (E)) at distance l from an FP1 particle (Malat1). A linear fit at $0.4\text{-}1\mu\text{m}$ yields the spatial correlation α - the probability that a FP2 is co-localized with an FP1 particle, as the y-axis intercept. (F, G) Distributions of measured spatial correlations for the data (blue) and randomized FP2 dots (red) for Nlrp6 (F) and Acly (G). The correlation with Malat1 was significant for Nlrp6 ($\alpha = 0.1745 \pm 0.0124$, $p\text{val}<1\text{e-}15$) but not for Acly ($\alpha = 0.008 \pm 0.0008$, $p\text{val}=0.23$).

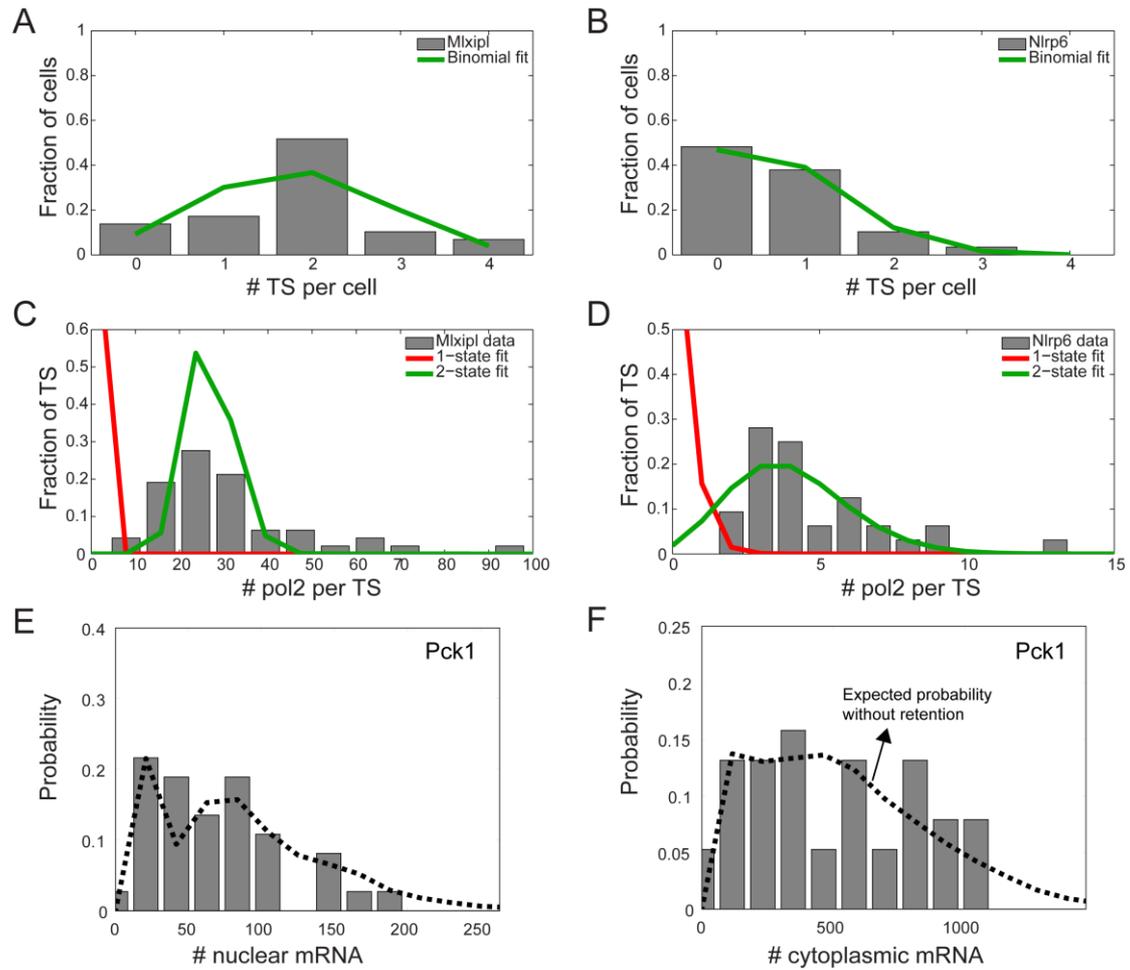


Figure S5 (related to Figure 5) – Mlxipl and Nlrp6 are expressed in a bursty manner. (A) Distribution of the number of active TS per nucleus for Mlxipl. (B) Distribution of the number of active TS per nucleus for Nlrp6. Green lines in A-B are binomial fits, demonstrating that promoters burst independently. (C,D) Distribution of Pol2 occupancy among TS of Mlxipl (C) and Nlrp6 (D). Green lines are binomial fits, red lines are the expected Pol2 occupancy distribution in a 1-state non-bursty model (Bahar Halpern et al., 2015). (E,F) Cytoplasmic variability for Pck1 is not smaller than that expected based on its bursting properties. Shown are the probability distributions of mRNA levels in the nucleus (E) and cytoplasm (F) of hepatocytes residing in the central vein in an ad-libitum fed mouse. Best-fit burst parameters are: $k_{ON} = 0.23 \text{ hr}^{-1}$, $k_{OFF} = 0.78 \text{ hr}^{-1}$. Cytoplasmic coefficient of variance was not significantly different than that expected from a 2-state bursty model with immediate export ($CV=0.62$ vs. 0.56 , $pval=0.91$).

Table S1 (related to Figure 1) – Calibration factor for obtaining the numbers of nuclear and cytoplasmic mRNA per cell from the RNAseq experiments. The factors were computed by dividing the nuclear or cytoplasmic sequencing read counts of selected genes by the number of nuclear or cytoplasmic mRNA counted using smFISH.

Sample	factor	Genes used for calibration
MIN6 nucleus	55±13	Actb, Acly, Fasn
MIN6 cytoplasm	24±6.5	Actb, Acly, Fasn
Liver nucleus	9.3±1.4	Ass1, Actb
Liver cytoplasm	1.88±0.36	Ass1, Actb

Table S2 (related to Figure 1) – Numbers of nuclear and cytoplasmic mRNAs per cell in MIN6 cells and liver cells. Reads were normalized to estimated numbers per cell based on the calibration factors of Table S1.

Table S3 (related to Figure 1) – Splicing efficiency of the introns in the nuclear and cytoplasmic fractions.

Table S4 (related to Figure 1) – Minimal P-values for the interaction between 83 RNA binding proteins and 448 nuclearly retained genes, obtained using the RBPmap software (Paz et al., 2014).

Table S5 (related to Figure 2) – Sequences and probe weight factors of the probe libraries used in this study. Additional probe libraries are described in (Bahar Halpern et al., 2015). W is the probe library weight factor and L is the gene length. Probe weight factors were computed as described in (Bahar Halpern et al., 2015). These depend on the physical location of the smFISH probes along the genes of interest, and are used to convert the intensities of the exon channel to number of Pol2 molecules per TS. The factors for the additional genes studied appear in (Bahar Halpern et al., 2015).

SUPPLEMENTAL EXPERIMENTAL PROCEDURES

Mice and tissues

All animal studies were approved by the Institutional Animal Care and Use Committee of WIS. C57bl6 male mice age 5 month were fed normal chow ad libitum, fasted or re-fed for the indicated times. Mice were sacrificed at 9AM (fed state) and 12 PM (fast state, for these mice food was removed at 8AM). In the RNAseq liver experiment (Figure 1B) and the re-feeding experiment (Figure 2) mice were housed under reverse phase cycle, and fasted for 5 hours starting at 7AM. RNA was extracted from the fasted mice and processed for RNAseq. Mice were then re-fed ad libitum for the indicated times and sacrificed immediately after the end of the feeding time. For the insulin tolerance test, (IT Figure 3) mice were fasted for 5 hours, injected with 0.75 U/Kg Insulin (SIGMA, I1882) and sacrificed 30 minute after injection. For the glucose tolerance test, (GT Figure 3) mice were fasted for 5 hours, injected with 2 gr/Kg glucose (D-Glucose SIGMA, G-6152) and sacrificed 30 minutes (GT30) or 1 hour (GT60) after injection. High fat diet (HFD) was applied to 2 months old mice for 8 weeks (Research Diets, d124921). Germ-free (GF) C57bl6 mice were housed in sterile isolators. For the antibiotic treatment mice were given a combination of the following antibiotics for 4 weeks, vancomycin (1 g/l), ampicillin (1 g/l), kanamycin (1 g/l), and metronidazole (1 g/l) in their drinking water (Fagarasan et al., 2002; Ichinohe et al., 2011; Rakoff-Nahoum et al., 2004). All antibiotics were obtained from Sigma Aldrich. All mice were sacrificed by cervical dislocation. Liver and duodenum tissues were harvested and fixed in 4% paraformaldehyde for 3 hours; incubated overnight with 30% sucrose in 4% paraformaldehyde and then embedded in OCT. 25 μ m or 6 μ m cryosections were used for hybridization for liver or duodenum respectively.

Hybridization and imaging

Probe library constructions, hybridization procedures and imaging conditions were previously described (Itzkovitz et al., 2011; Lyubimova et al., 2013). To detect cell borders alexa fluor 488 conjugated phalloidin (Rhenium A12379) was added to the GLOX buffer wash (Lyubimova et al., 2013). Portal node was identified morphologically on DAPI images based on bile ductile, central vein was identified using smFISH for Glutamine Synthetase performed on serial sections. Only tetraploid hepatocytes within the first three layers of the portal node (up to ~50 um distance) were used for noise analysis, to ensure analysis of a homogenous cell population, since the liver is a polyploid tissue in which gene expression is spatially zonated (Bahar Halpern et al., 2015). All quantifications of smFISH data are based on 30-100 cells.

Cell culture

Pancreatic islets were maintained and expanded up to one day in RPMI 1640 media (Biological Industries) supplemented with 10% Fetal bovine serum (Biological Industries), 1% of Penicillin-Streptomycin (Biological Industries) and 1% L-Glutamine (Biological Industries). MIN6 cells were maintained and expanded in DMEM media (Biological Industries) supplemented with 15% fetal bovine serum (Biological Industries), 1% of Penicillin-Streptomycin (Biological Industries), 1% L-Glutamine (Biological Industries) and 0.1 mM β -mercaptoethanol (Sigma). All treatment on MIN6 cells were performed on passages 20-30. Cells seeded on cover-slips in 6 well plates. For different glucose concentrations cells were starved in glucose free DMEM (Sigma D5030) supplemented with MIN6 medium components for 16 hr then were treated for 1 hr with no addition of glucose (0mM glucose) or addition of 30 mM glucose to the medium. For serum starvation, cells were maintained in serum free medium for 16 hr. For insulin treated cells, cells were serum starved for 16 hr and then were treated with 10nM insulin (Biological Industries 01-818-1H) for 1 hr. For heat shock, the 6 well plate was floated in 45°C bath for 1 hr.

Isolation of primary pancreatic islet cells

Pancreatic islets from C57bl6 mice between the ages of 6-8 weeks, were prepared with a solution of collagenase P (Roche, 11-213-865) diluted in RPMI 1640 (Biological Industries) at a concentration of 1 mg/ml. The solution was first injected into the bile duct before removal of the pancreas, followed by digestion for 6-7 min at 37 °C. The isolated pancreas was washed twice with fresh RPMI and centrifuged in cold centrifuge for 1 minute at 200g. Pellet was resuspended with 4 ml Histopaque 1119 (Sigma), 4 ml of Histopaque 1117 (Sigma) and then 3 ml of RPMI 1640 were layered on top of the resuspended pellet. Tubes were then centrifuged in cold centrifuge with no break or acceleration for 20 minute at 1000g. Individual islets were separated and selected by hand using a microscope, and were then trypsinized into single cells, cultured up to one day and fixed in 4% paraformaldehyde for 15 minutes.

Cell fractionation

Isolation of nuclear and cytoplasmic liver mRNA was performed according to the Nascent-SEQ protocol (Menet et al., 2012) except for minor modifications. In order to isolate cytoplasmic mRNA the supernatant was collected following nuclei isolation by sucrose gradient. For isolation of nuclear mRNA the supernatant was collected following chromatin isolation. For RNA extraction, 1/50 volumes of 5M NaCl and 2.5 volumes of 100% EtOH were added to the supernatants collected, and the mixture was incubated at -20⁰C for 1 hour and then centrifuged for 20 minutes at full speed. The pellet was resuspended in 0.5 ml 0.5% SDS buffer (0.5% SDS, 0.1M NaCl, 1mM EDTA, 10mM Tris-HCl). Similar volume of acid phenol:chloroform (Ambion AM9722) mixture was added. The mixture was then vortexed and centrifuged at full speed for 5 minutes at RT. The aqueous phase was transferred to a new tube and 1 ml of 0.5% SDS buffer was added to the phenol phase for re-extraction. The

two aqueous phases were combined and re-extracted with acid phenol:chloroform. The RNA from the aqueous phase was then isolated using standard EtOH precipitation.

For isolation of nuclear and cytoplasmic RNA from MIN6, the cells ($\sim 2 \times 10^6$) were first trypsinized and washed with cold PBS. Cell pellet was then treated with 175 μ l RLN buffer (Tris pH8.0 50mM, NaCl 140mM, MgCl₂ 1.5mM, NP-40 0.15mM, EDTA 10mM, DTT 1mM, RNase inhibitor 10U/ml) and incubated for 5 minutes on ice. Lysate was centrifuged at 300g for 5 minutes in a cold centrifuge. The supernatant (cytoplasmic fraction) was separated and the pellet was resuspended with same volume of RLN buffer and immediately centrifuged at 500g for 1 minute. Pellet was resuspended in 1 ml S1 buffer (sucrose 250mM, MgCl₂ 10mM, RNase inhibitor 10U/ml), layered over 3 ml of S3 buffer (sucrose 880mM, MgCl₂ 10mM, RNase inhibitor 10U/ml) and centrifuged for 10 minutes in cold centrifuge at 2800g. RNA from the pellet (nuclear fraction) and the cytoplasmic fraction was isolated using RNeasy Mini Kit (QIAGEN) according to the manufacturer's instructions.

RNA sequencing

Libraries were prepared with Illumina TrueSeq kits and sequenced on Illumina HiSeq. Reads were aligned to the mouse mm10 reference genome using TopHat2 (Trapnell et al., 2010) by using default parameters. Read counts for individual mouse genes annotated in Ensembl 80 were computed using HTseq (Anders et al., 2014). Reads for nucleoplasmic and cytoplasmic transcripts and mRNAs were calculated by counting exonic reads in the last 500bp from the 3' end of the gene. We only included the last 500bp of the spliced gene since RNA degradation in the nuclear fraction was significantly elevated further upstream (Sigurgeirsson et al., 2014). We converted the number of reads per gene to the number of nuclear or cytoplasmic mRNA copies per single cell using smFISH measurements. These measurements were performed on 30-100 cells for each calibration gene in MIN6 cells or liver tissue sections from mice that were sacrificed at the same hour and were fasted as the

ones used in the RNA sequencing. For MIN6 calibration we used the genes *Acly*, *Fasn* and *Actb*. For liver calibration we used the genes *Ass1* and *Actb*, genes which we have shown to be relatively stable in their expression levels in diverse metabolic conditions (Bahar Halpern et al., 2015) (Table S1). When analyzing the statistics of ratios of cytoplasmic and nuclear mRNA the minimal number of copies per cell was set as 0.01 in MIN6 and 0.1 in liver for both the nuclear and cytoplasmic fractions.

Splicing efficiencies were analyzed similar to the approach described in (Tilgner et al., 2012). For each intron annotated in Ensembl 80 we counted the number of reads mapping across the exon boundaries into the adjacent intron sequence (originating from primary unspliced mRNA molecules), and compared them to the number of reads split-mapping across the exon–exon junctions (originating from a successfully spliced transcript). When estimating the fraction of nuclear genes and the enrichment of different classes (Figure S1A-C) we only considered genes with more than 0.1 copies per cell in MIN6 or 1 copy per cell in liver.

Sequence Motif analysis

To identify sequence motifs that are over-represented in the 3' UTR of retained genes, we analyzed 717 genes that had more than 0.1 copies per cell in MIN6 cells or 1 copy per cell in liver cells, and that had nuc/cyto ratios above 1.4 in MIN6 cell and above 1 in liver cells. We used the AmadeusPBM_v1.0 software (Linhart et al., 2008) to identify common motifs in the 3' UTR of this gene set. In AmadeusPBM_v1.0, data type was set to “Target set”, sequence type to “3' UTR”, and the variant in scores for ranking motifs to “Binned” to control for length and GC biases of the analyzed sequences. The motif length was kept to the default value, 8. The ten significant motifs found are presented in Figure S2B.

To identify putative RNA binding protein motifs at the 3' UTR of our retained gene selection we used the RBPmap software (Paz et al., 2014), which includes a comprehensive

database of 94 RNA binding proteins, the recognition sequences of which have been defined. We removed 16 RBP that were expressed at less than 1 mRNA copy per liver cell. Of the 717 retained genes, RBPmap found 654 valid sequences. For this set, we computed the binding probabilities (minimal pvalue) of each of the 78 RNA binding proteins (Figure S2A, Table S4).

In order to estimate the significance of the similarities between the 10 common motifs found in the 3'UTR of the 717 genes with the most retained mRNAs and the RBP motifs, we followed the procedure presented in Itzkovitz et al. (Itzkovitz et al., 2006). Shortly, for each 3'UTR motif – RBP motif combination, we performed all pairwise comparisons of the shifted versions of their PWMs, with the condition of a 5-positions overlap minimum. For each relative shift, we summed over all the overlapping motif elements the similarity of the two elements in the two motifs found at the same position, weighted by the product of the element information content of both motifs. The similarity was taken to be one minus the Shannon-Jensen distance. Finally, the combination similarity was taken to be the maximum value out of all the shifts. For each pairwise comparison, we estimated the P-value by generating 1000 randomized realizations of the two motifs. In each realization, we randomly exchanged the A-T and C-G positions in each column of the motif's PWM, thus preserving the GC content. In addition, we randomly permuted the different positions within the motif. The P-value was taken to be the per cent of realizations with similarity larger than the estimated one (for further details, see the section “Measurement of sequence similarity” in Itzkovitz et al., 2006). We did not find combinations that were significant with an $FDR < 0.2$.

Measurements of nuclear export rates

To assess the nuclear export rates, cytoplasmic degradation rates and burst parameters we used the method of Bahar Halepern et al. We first identified mono-nucleated tetraploid hepatocytes by nuclear size and transcription sites (TS) of Pck1, a ubiquitously expressed

gene that exhibited close to 4 active TS in each tetraploid nucleus (Bahar Halpern et al., 2015). Active TS of the genes of interest were then identified in these nuclei based on dots that appeared in both the intronic and exonic channels. The burst fraction f was computed as the average number of active TS per cell divided by 4. Only cells for which the entire nucleus appeared in the stacks were considered.

We estimated the transcription rate from active TS, μ , by inferring the number of Pol2 molecules per active TS (M) (Bahar Halpern et al., 2015). This was inferred from the intensity of the exonic dots, using correction factors for the spread of the smFISH probes along the genes of interest (Table S5). We used the Pol2 occupancy, M , the length of the gene, L , and the speed of RNA polymerase ($v=34\text{bp/s}$, Bahar Halpern et al., 2015) to obtain the average transcription rate from an active TS: $\mu = M \cdot v/L$. Overall transcription rate per cell was calculated as $\beta = 4 \cdot f \cdot \mu$. We next used equations [3-4] to obtain the nuclear export rate and cytoplasmic degradation rates by dividing the cellular transcription rate by the average numbers of nuclear and cytoplasmic mRNA respectively.

To quantify the number of nuclear mRNA molecules we counted the number of nuclear exonic dots in 5 consecutive optical sections around the stack in which the nucleus had the largest area, divided by the quantified nuclear volume to obtain concentrations and multiplied it by the total nuclear volume, obtained from Martin et al. (Martin et al., 2002). Cytoplasmic mRNA was quantified similarly using cytoplasmic counts and volumes. For Mxipl, where nuclear mRNAs were often clustered, we used the summed nuclear dot intensity divided by the average intensity of a single cytoplasmic dot, instead of the number of nuclear dots.

Fitting a 2-state bursty transcription model

The bursting rates k_{ON} and k_{OFF} were computed by fitting the model of Raj et al. (Raj et al., 2006) to the distribution of nuclear mRNA. According to this model the distribution of mRNA per cell, Y , generated by a single bursty promoter is:

$$[S1] P(Y) = \frac{\Gamma(\frac{k_{ON}+Y}{\lambda})}{\Gamma(Y+1)\Gamma(\frac{k_{ON}+k_{OFF}+Y}{\lambda})} \frac{\Gamma(\frac{k_{ON}+k_{OFF}}{\lambda})}{\Gamma(\frac{k_{ON}}{\lambda})} (\frac{\mu}{\lambda})^Y {}_1F_1(\frac{k_{ON}}{\lambda} + Y, \frac{k_{ON}}{\lambda} + \frac{k_{OFF}}{\lambda} + Y, -\frac{\mu}{\lambda})$$

Where ${}_1F_1$ is a confluent hypergeometric function of the first kind. Since our cells are tetraploid we convolved the distribution with itself 4 times. This was justified since the promoter state of each chromosomal locus was independent of the states of the other promoters in that cell (Figure S5). Importantly, in Equation [S1] nuclear export rate λ was used instead of the degradation rate δ , since it plays a similar role in generating the nuclear variability (Equation [3]). Since we measured $f=k_{ON}/(k_{ON}+k_{OFF})$ as well as μ and λ our fit had only a single free parameter. To assess the noise that would be observed without nuclear retention we used equation [S1] with δ and our inferred k_{ON} , k_{OFF} . To assess the differences in noise we performed 10,000 sampling events of N cells from this analytical distribution, where N is the number of cells quantified for the gene of interest. We measured the coefficient of variance of each random sample and computed a p-value as the fraction of sampled sets that had a CV that was lower than the experimental one.

When fitting the distributions of nuclear and cytoplasmic mRNA we corrected for the effect of subsampling a partial volume of the nucleus and cytoplasm (Bahar Halpern et al., 2015). To minimize the broadening of the mRNA distributions due to a small subsample effect we quantified the mRNA concentration in 15 consecutive optical sections around the stack with maximal DAPI area, rather than 5 optical sections, as was done when computing nuclear export rates (Figure 2F). While quantifying large number of optical sections could potentially result in inclusion of mRNA molecules that are either below or above the nucleus this phenomenon was negligible for the genes in which we analyzed noise distributions, for which cytoplasmic mRNA concentrations were small.

Computing the spatial correlations of nuclear transcripts with nuclear domains

We estimated 2D spatial correlation α (co-localization) between fluorescently labeled transcripts of two different genes, using the particle image cross-correlation spectroscopy (PICCS) method (Semrau et al., 2011). The first sets of particles were the foci of either Malat1 or Neat1, lncRNA markers for speckles or paraspeckles respectively. The second set of particles included the transcripts of the gene with nuclearly retained mRNA (Mlxipl, Nlrp6, or Acl1 as a control). For simplicity, we denote the sets of dots from the two fluorescence channels by FP1 and FP2. We used 2D image slices rather than 3D, as in (Semrau et al., 2011).

We corrected shifts between the fluorescent channels using normalized image cross-correlation. We used DAPI staining to identify the nuclei and the dense chromatin regions within them. For each image, to reduce axial dependent sensitivity, we normalized all axial layers to have the same DAPI median intensity as the first axial layer. Next, we pooled each Z-stack in each nucleus that had at least 1 FP1 transcript within it, to obtain N samples, to obtain N samples, N_{sam} . In each sample, we counted the number of FP2 around each FP1 within an increasing distance l until a limited length of $1\mu\text{m}$. We considered only FP1 transcripts that were distant from the nucleus edge by at least $1\mu\text{m}$. We averaged the profiles from all the N samples and obtained the average normalized cumulative distribution of FP2 transcript around an FP1 transcript, i.e.

$$[S2] C_{norm}(l) = C(l) \frac{N_{FP1}}{N_{FP2}},$$

Where $C(l)$ is the number of FP2 particles within a circle of radius l around an FP1 particle. N_{FP1} and N_{FP2} are the numbers of FP1 and FP2 particles in that sample. The normalization is important to control for multiple FP1 particles within the same sample. The cumulative distribution of FP2, $C_{norm}(l)$ has the following form

$$[S3] C_{norm}(l) = \alpha P(l) + (1 - \alpha) \Sigma_{norm} \pi l^2 ,$$

Where α is the fraction of the FP2 transcripts which are correlated with the FP1 transcripts, $P(l)$ is the cumulative probability to find a distance smaller than l between FP1 and FP2 transcripts, and Σ_{norm} is the 2D spatial density of FP2 particles. At l large enough distances, $P(l)=1$ and the added FP2 transcripts are completely uncorrelated with the FP1 transcript, so the $C_{norm}(l)$ vs. l^2 form is linear. We estimated α by fitting a line to $C_{norm}(l)$ at large l . We found $C_{norm}(l)$ linear at $0.4 \leq l \leq 1 \mu m$ for all data sets, so we used this l range for the fit (Figure S4). For estimating the uncertainty of α , we used the jackknife resampling technique:

$$[S4] \Delta\alpha = \sqrt{\frac{N_{sam}-1}{N_{sam}} \sum_{i=1}^{N_{sam}} (\alpha_i - \bar{\alpha}_i)^2},$$

Where α_i is α estimated from all the samples except for sample i .

When assessing whether the spatial correlation measured by α is significant it is critical to take into account the fact that mRNA are not randomly distributed in the nucleus. Vargas et al. have shown that regions of dense chromatin are largely depleted of mRNA (Vargas et al., 2005), a phenomenon that we also observed using our smFISH approach. To account for this non-random exclusion of mRNA we generated randomized datasets in which the FP2 dots were randomly distributed within the allowed nuclear region and recomputed the spatial correlations between FP1 and the randomized FP2. This calculation yielded the probability to have any α value when there is no correlation, $P_{sim}(\alpha)$, and was compared to the α probability distribution from the data, $P_{data}(\alpha)$. For obtaining $P_{sim}(\alpha)$ we ran 1000 simulations. For each simulation and each sample, we kept the positions of the FP1 transcripts and randomly placed the number of FP2 transcripts in that sample within all the allowed pixels (excluding the dense chromatin regions). Then for each simulation we estimated α , as mentioned above, by counting $C_{norm}(l)$ and estimating α by fitting a line at

the $0.4 \leq l \leq 1 \mu m$ range. For estimating $P_{data}(\alpha)$, we assumed a Gaussian distribution where the mean and standard deviation were taken to be the α and $\Delta\alpha$, respectively, which were estimated from the data.

The P-value is evaluated in the following way:

$$[S5] pval(\alpha') = \int_{\alpha'}^1 P_{sim}(\alpha) d\alpha$$

$$[S6] pval = \int_0^1 pval(\alpha') P_{data}(\alpha') d\alpha' .$$

SUPPLEMENTAL REFERENCES

Bahar Halpern, K., Tanami, S., Landen, S., Chapal, M., Szlak, L., Hutzler, A., Nizhberg, A., and Itzkovitz, S. (2015). Bursty gene expression in the intact Mammalian liver. *Mol. Cell* 58, 147–156.

Fagarasan, S., Muramatsu, M., Suzuki, K., Nagaoka, H., Hiai, H., and Honjo, T. (2002). Critical roles of activation-induced cytidine deaminase in the homeostasis of gut flora. *Science* 298, 1424–1427.

Ichinohe, T., Pang, I.K., Kumamoto, Y., Peaper, D.R., Ho, J.H., Murray, T.S., and Iwasaki, A. (2011). Microbiota regulates immune defense against respiratory tract influenza A virus infection. *Proc. Natl. Acad. Sci. U. S. A.* 108, 5354–5359.

Itzkovitz, S., Tlusty, T., and Alon, U. (2006). Coding limits on the number of transcription factors. *BMC Genomics* 7, 239.

Itzkovitz, S., Lyubimova, A., Blat, I., Maynard, M., van Es, J., Lees, J., Jacks, T., Clevers, H., and van Oudenaarden, A. (2011). Single molecule transcript counting of stem cell markers in the mouse intestine. *Nat. Cell Biol.* 14, 106–114.

Linhart, C., Halperin, Y., and Shamir, R. (2008). Transcription factor and microRNA motif discovery: The Amadeus platform and a compendium of metazoan target sets. *Genome Res.* 18, 1180–1189.

Lyubimova, A., Itzkovitz, S., Junker, J.P., Fan, Z.P., Wu, X., and van Oudenaarden, A. (2013). Single-molecule mRNA detection and counting in mammalian tissue. *Nat. Protoc.* 8, 1743–1758.

Martin, N.C., McCullough, C.T., Bush, P.G., Sharp, L., Hall, A.C., and Harrison, D.J. (2002). Functional analysis of mouse hepatocytes differing in DNA content: volume, receptor expression, and effect of IFN γ . *J. Cell. Physiol.* 191, 138–144.

- Menet, J.S., Rodriguez, J., Abruzzi, K.C., and Rosbash, M. (2012). Nascent-Seq reveals novel features of mouse circadian transcriptional regulation. *eLife* 1.
- Paz, I., Kosti, I., Ares, M., Cline, M., and Mandel-Gutfreund, Y. (2014). RBPmap: a web server for mapping binding sites of RNA-binding proteins. *Nucleic Acids Res.* gku406.
- Raj, A., Peskin, C.S., Tranchina, D., Vargas, D.Y., and Tyagi, S. (2006). Stochastic mRNA synthesis in mammalian cells. *PLoS Biol.* 4, e309.
- Rakoff-Nahoum, S., Paglino, J., Eslami-Varzaneh, F., Edberg, S., and Medzhitov, R. (2004). Recognition of Commensal Microflora by Toll-Like Receptors Is Required for Intestinal Homeostasis. *Cell* 118, 229–241.
- Semrau, S., Holtzer, L., González-Gaitán, M., and Schmidt, T. (2011). Quantification of Biological Interactions with Particle Image Cross-Correlation Spectroscopy (PICCS). *Biophys. J.* 100, 1810–1818.
- Sigurgeirsson, B., Emanuelsson, O., and Lundberg, J. (2014). Sequencing Degraded RNA Addressed by 3' Tag Counting. *PLoS ONE* 9.
- Tilgner, H., Knowles, D.G., Johnson, R., Davis, C.A., Chakraborty, S., Djebali, S., Curado, J., Snyder, M., Gingeras, T.R., and Guigó, R. (2012). Deep sequencing of subcellular RNA fractions shows splicing to be predominantly co-transcriptional in the human genome but inefficient for lncRNAs. *Genome Res.* 22, 1616–1625.
- Vargas, D.Y., Raj, A., Marras, S.A.E., Kramer, F.R., and Tyagi, S. (2005). Mechanism of mRNA transport in the nucleus. *Proc. Natl. Acad. Sci. U. S. A.* 102, 17008–17013.