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

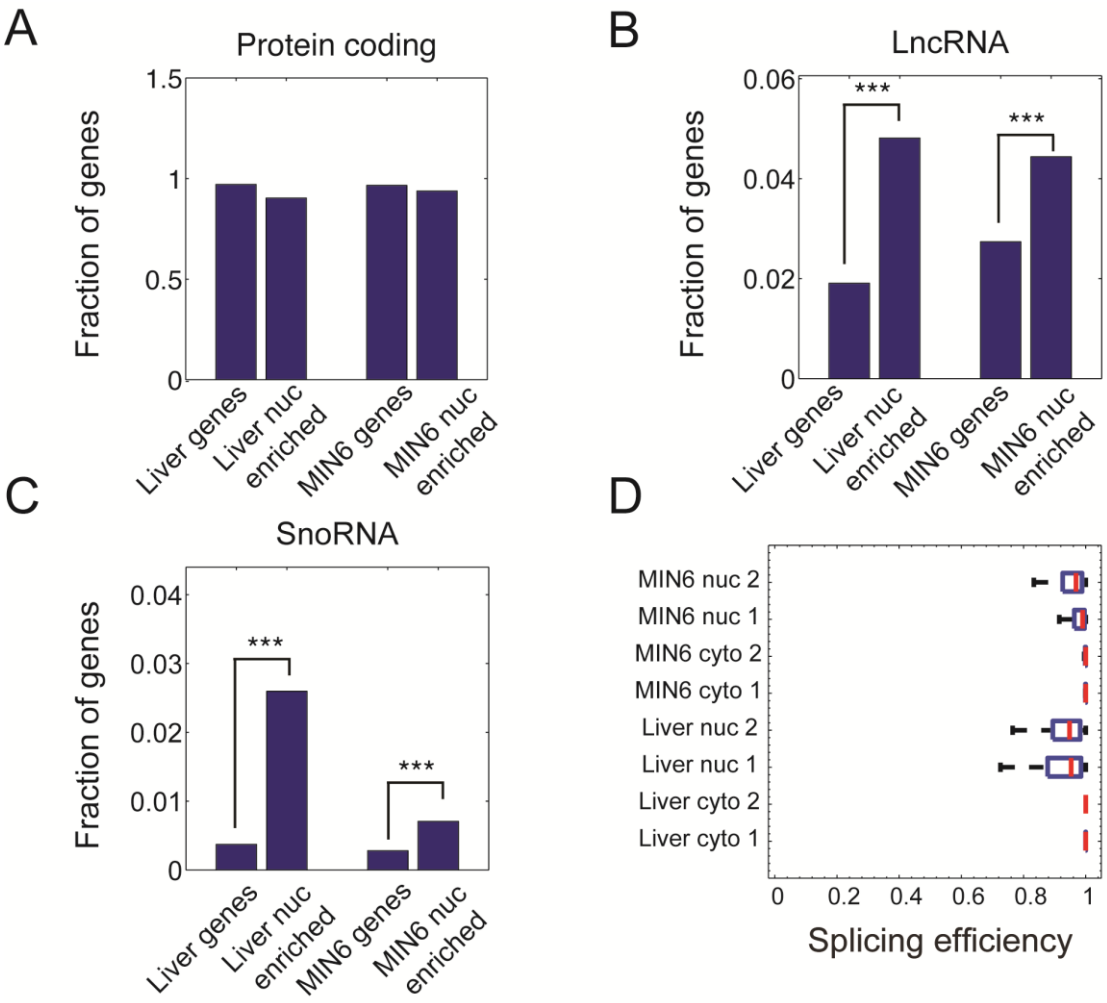

Figure S1 (Related to Figure 1) – RNAseq of nuclear and cytoplasmic fractions. (A) Protein-coding genes have a comparable representation among the genes with higher numbers of mRNA in the nucleus compared to the cytoplasm (protein-coding genes are 97% of the entire transcriptome sequenced and 91% of 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. *** pval<0.001.

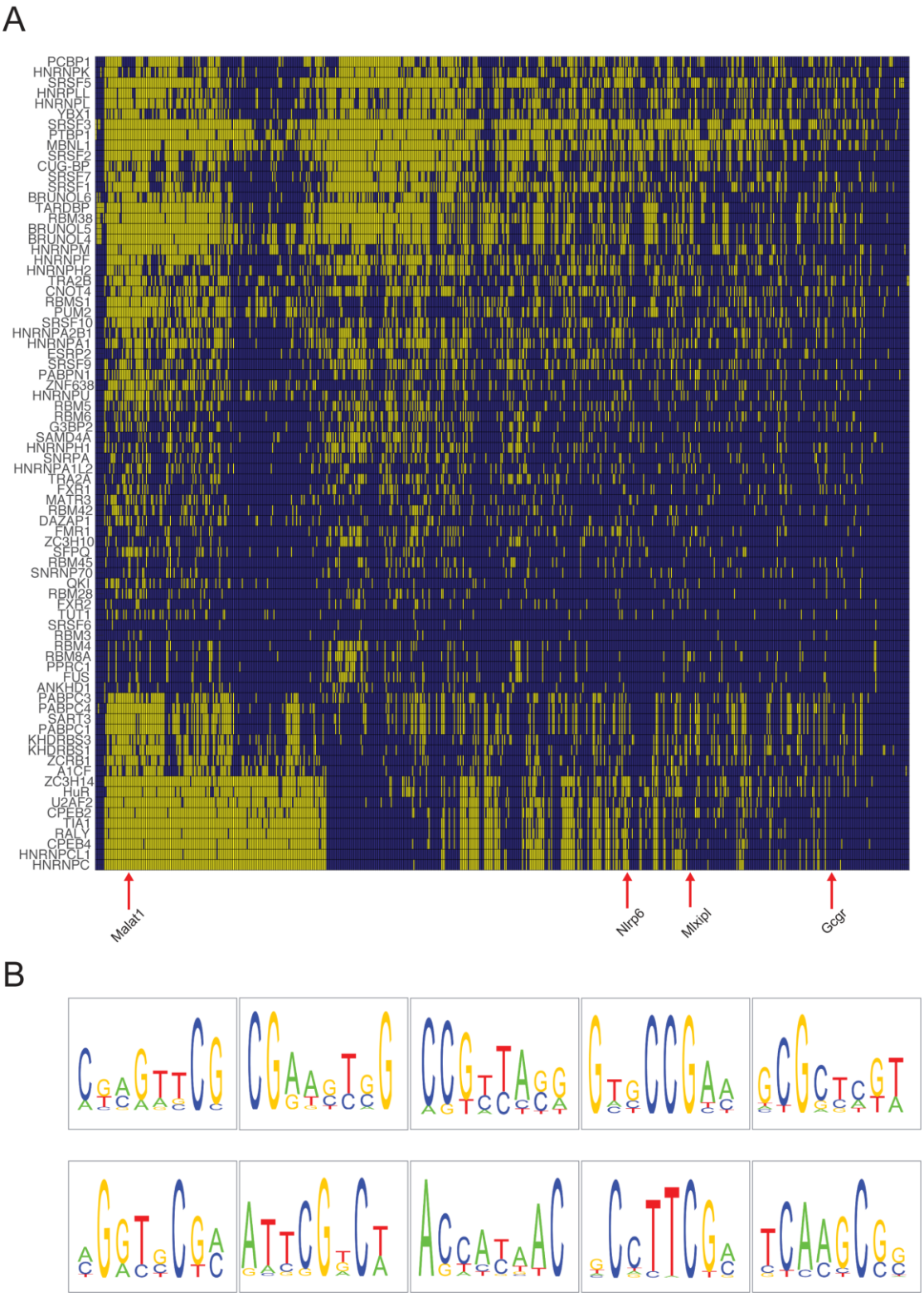

Figure S2 (related to Figure 1) – Putative RNA binding proteins and enriched motifs for the nuclearly enriched genes. (A) Putative binding interactions between known RNA binding proteins and the 3'UTR of the most nuclearly retained genes. Columns are the 654 most nuclearly retained genes, rows are RNA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

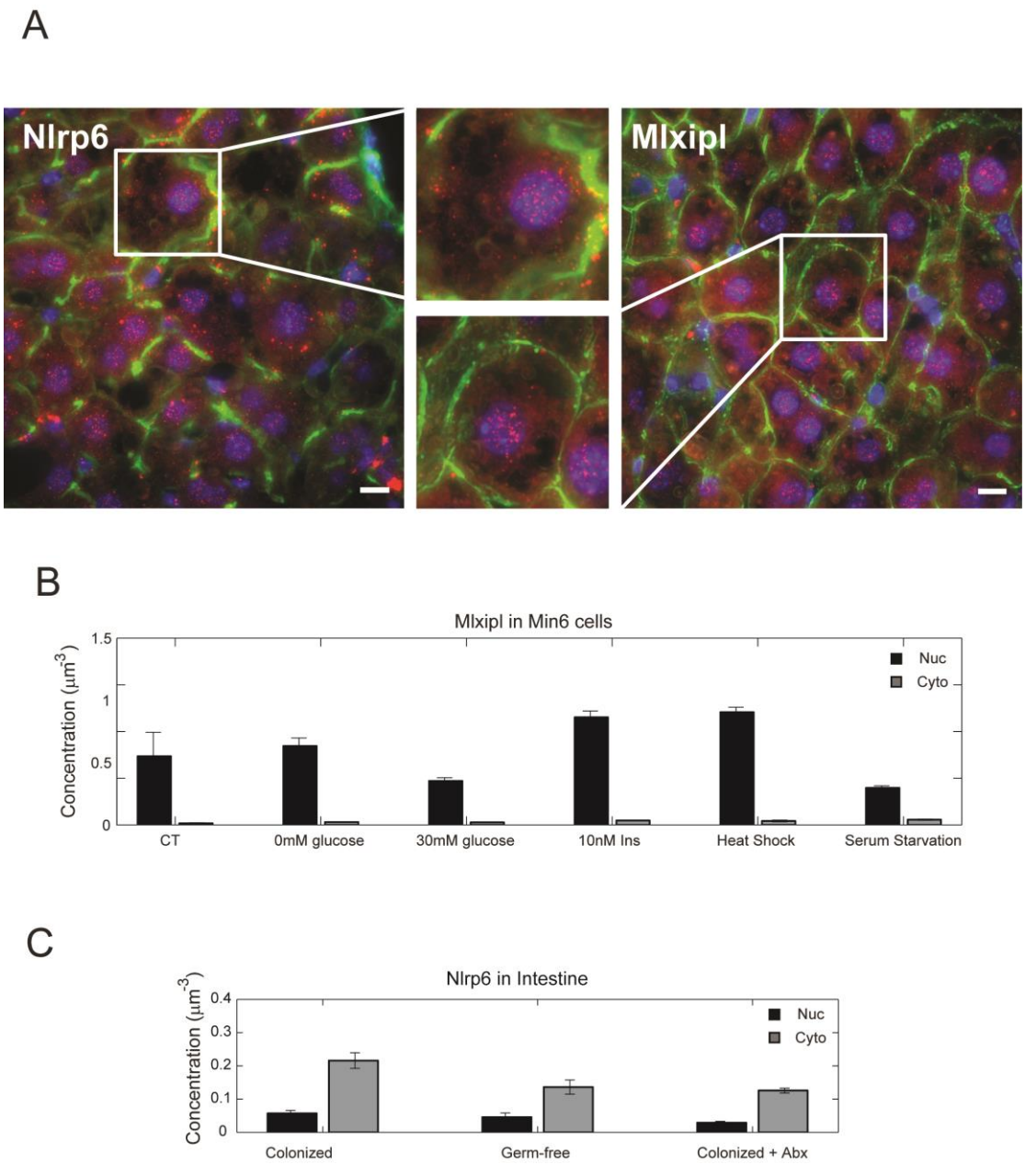

Figure S3 (related to Figure 3) – (A) Transcripts of Nlrp6 and Mlxipl remain nuclear in livers of mice after 8 weeks of high-fat diet. Red dots are single mRNA of Nlrp6 (left) or Mlxipl (right). Green - phalloidin-stained membranes, blue - DAPI-stained nuclei. (B) Mlxipl mRNA remains

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.

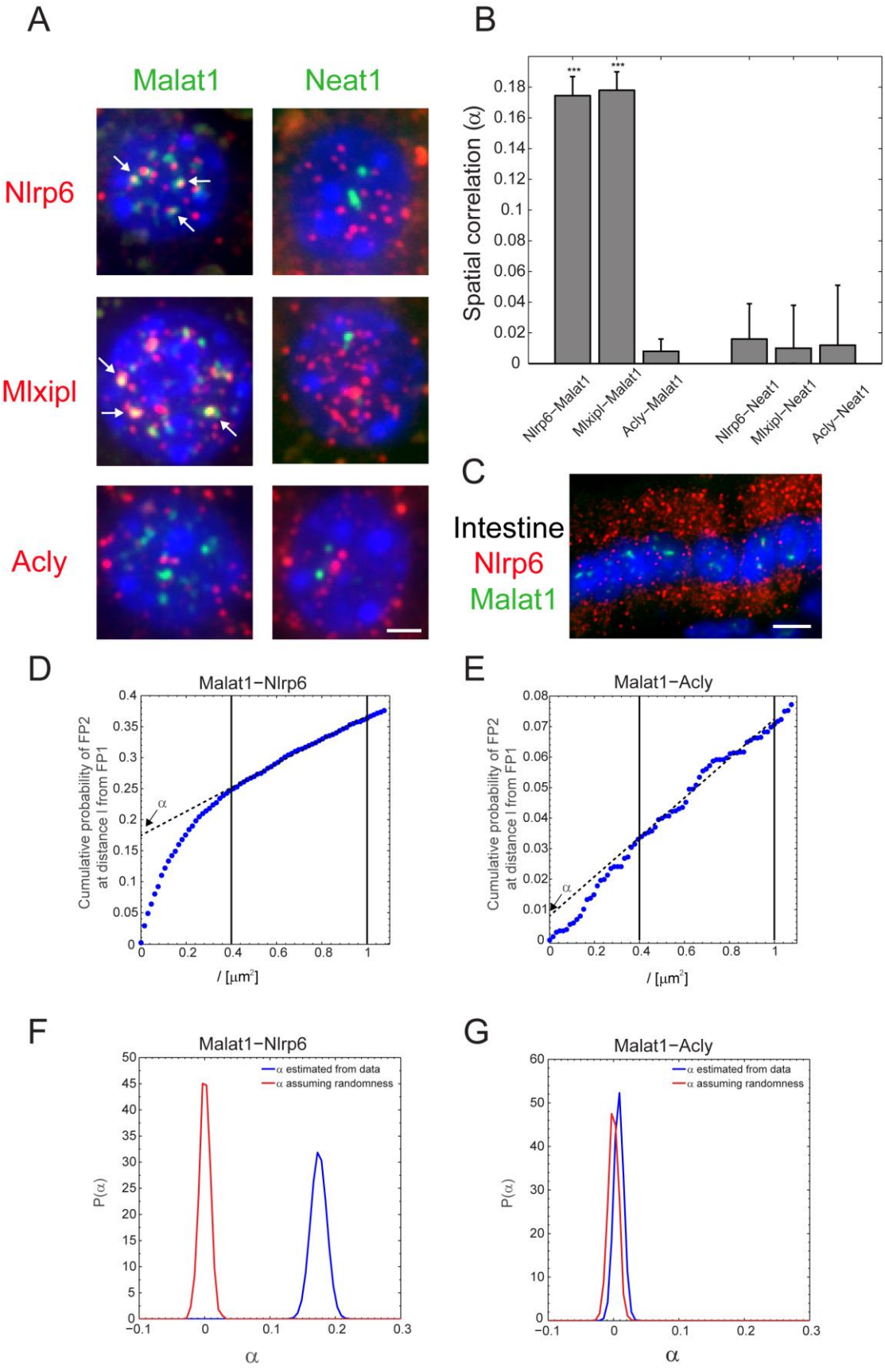

Figure S4 (related to Figure 3) – Nuclear transcripts of Mlxipl and Nlrp6 co-localize with nuclear speckles. (A) Dual color smFISH for Malat1 (green dots left column) or Neat1 (green

dots right column) with Mlxipl, Nlrp6 and Acly (red dots). Images are maximal projections of 5 optical sections space 0.3µm apart. Scale bar is 2µm (B) Mlxipl and Nlrp6 co-localized with nuclear speckles. Shown are the spatial correlations (α) between the relevant genes and either Malat1 or Neat1. *** pval<1e-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 5µm (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 0.4-1µm 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 ($\alpha = 0.1745 \pm 0.0124$, pval<1e-15) but not for Acly ($\alpha = 0.008 \pm 0.0008$, pval=0.23).

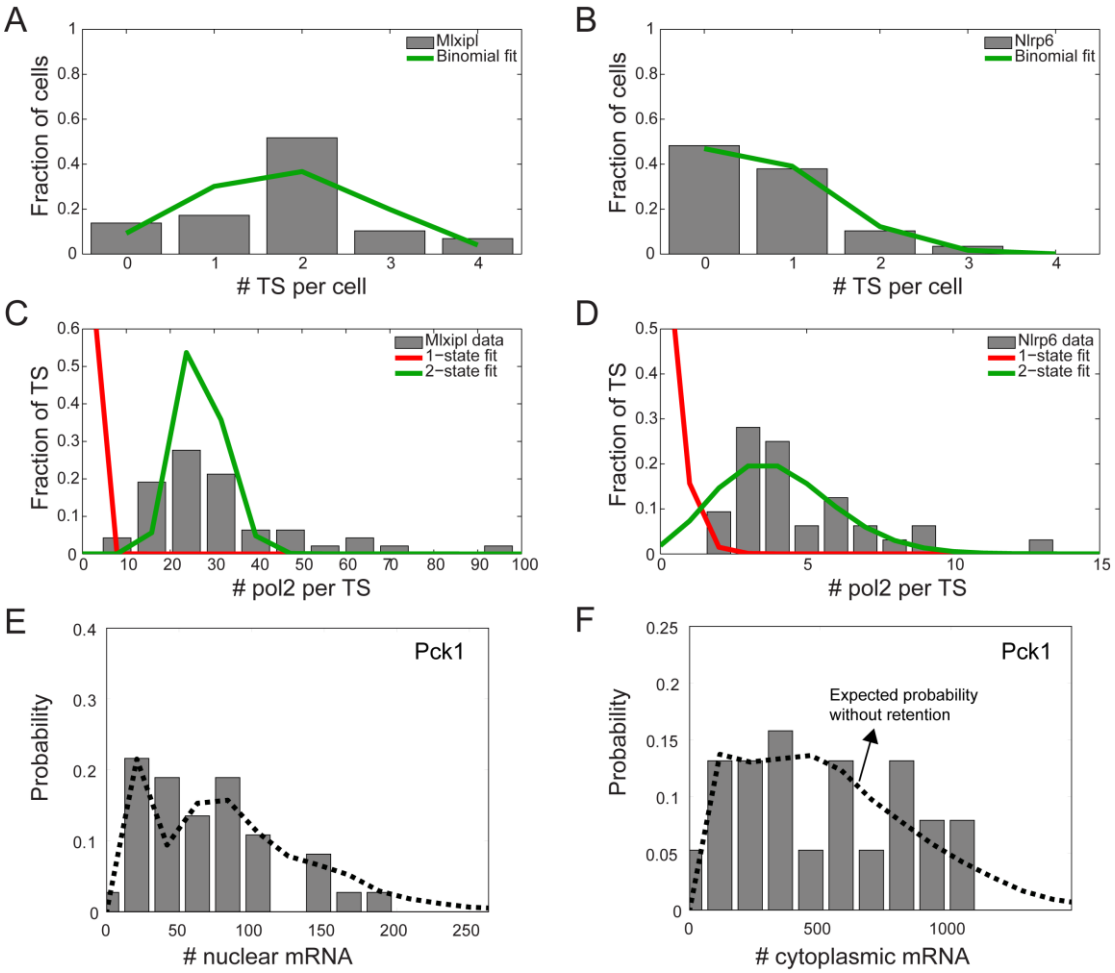

Figure S5 (related to Figure 5) – Mlxipl and Nlrp6 are expressed in a bursty manner. (A) Distribution of the number of active TS per nucleus for Mlxipl. (B) Distribution of the number of active TS per nucleus for Nlrp6. Green lines in A-B are binomial fits, demonstrating that promoters burst independently. (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: $k_{ON} = 0.23\ \mathrm{hr}^{-1}$, $k_{OFF} = 0.78\ \mathrm{hr}^{-1}$. 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, d12492I). 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 µm or 6 µm 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℃ 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 -20^{0}C 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 (~$2x10^6$) were first trypsinized and washed with cold PBS. Cell pellet was then treated with 175 µl 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*=34bp/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 Mlxipl, 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:

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

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 Acly 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 1µm. We considered only FP1 transcripts that were distant from the nucleus edge by at least 1µm. 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] $C_{norm}(l) = \mathrm{C}(l)\frac{N_{FP1}}{N_{FP2}}$,

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, $C_{norm}(\mathrm{l})$ 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] $\Delta\alpha = \sqrt{\frac{N_{sam-1}}{N_{sam}}\sum_{i=1}^{N_{sam}}(\alpha_i - \bar{\alpha}_i)}$,

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 $P_{sim}(\propto)$ 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 $0.4 \leq l \leq 1\ \mu m\ l$ 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 IFNgamma. 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 Lundeberg, J. (2014). Sequencing Degraded RNA Addressed by 3' Tag Counting. PLoS ONE *9*.

Tilgner, H., Knowles, D.G., Johnson, R., Davis, C.A., Chakrabortty, 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.