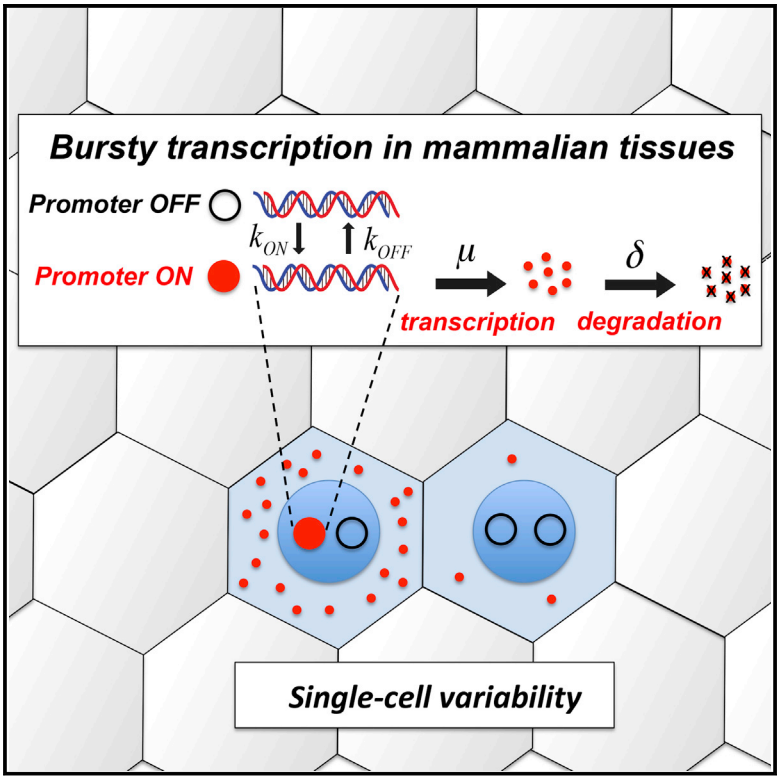
Article

Molecular Cell

Bursty Gene Expression in the Intact Mammalian Liver

Graphical Abstract



Authors

Keren Bahar Halpern, Sivan Tanami, ..., Anna Nizhberg, Shalev Itzkovitz

Correspondence

shalev.itzkovitz@weizmann.ac.il

In Brief

Bahar Halpern et al. use single-molecule approaches to demonstrate that gene expression in the intact mammalian liver consists of transcriptional bursts. Tight coordination of burst parameters and mRNA lifetimes, as well as liver polyploidy, reduce the gene expression noise created by these bursts.

Highlights

- Single-molecule approach to quantify all gene expression stages in intact tissues
- Gene expression in the mammalian liver consists of transcriptional bursts
- These bursts generate variability (noise) in the mRNA content of individual cells
- Liver polyploidy and temporal averaging reduce the resulting noise

Bahar Halpern et al., 2015, Molecular Cell *58*, 147–156
April 2, 2015
http://dx.doi.org/10.1016/j.molcel.2015.01.027

CellPress

Molecular Cell
Article

Bursty Gene Expression in the Intact Mammalian Liver

Keren Bahar Halpern,[1] Sivan Tanami,[1] Shanie Landen,[1] Michal Chapal,[1] Liran Szlak,[1] Anat Hutzler,[1] Anna Nizhberg,[1] and Shalev Itzkovitz[1,*]
[1]Department of Molecular Cell Biology, Weizmann Institute of Science, Rehovot, 76100, Israel
*Correspondence: shalev.itzkovitz@weizmann.ac.il
http://dx.doi.org/10.1016/j.molcel.2015.01.027

SUMMARY

Bursts of nascent mRNA have been shown to lead to substantial cell-cell variation in unicellular organisms, facilitating diverse responses to environmental challenges. It is unknown whether similar bursts and gene-expression noise occur in mammalian tissues. To address this, we combine single molecule transcript counting with dual-color labeling and quantification of nascent mRNA to characterize promoter states, transcription rates, and transcript lifetimes in the intact mouse liver. We find that liver gene expression is highly bursty, with promoters stochastically switching between transcriptionally active and inactive states. Promoters of genes with short mRNA lifetimes are active longer, facilitating rapid response while reducing burst-associated noise. Moreover, polyploid hepatocytes exhibit less noise than diploid hepatocytes, suggesting a possible benefit to liver polyploidy. Thus, temporal averaging and liver polyploidy dampen the intrinsic variability associated with transcriptional bursts. Our approach can be used to study transcriptional bursting in diverse mammalian tissues.

INTRODUCTION

Gene expression in unicellular organisms and in mammalian cell lines has been shown to be highly bursty (Blake et al., 2003, 2006; Chong et al., 2014; Dar et al., 2012; Bar-Even et al., 2006; Friedman et al., 2006; Golding et al., 2005; Kaern et al., 2005; Newman et al., 2006; Pedraza and Paulsson, 2008; Raj and van Oudenaarden, 2008; Suter et al., 2011). Promoters tend to stochastically transition between a closed, transcription-prohibitive state and an open permissive state, generating bursts of nascent transcripts. This bursting phenomenon can generate intrinsic variability, or "noise," in the mRNA content of isogenic cells. In unicellular organisms such variability may constitute a "bet-hedging" strategy, improving the chances that a clonal population adapts to variable conditions (Chalancon et al., 2012; Eldar and Elowitz, 2010).

An open question is whether single-cell variability is an advantage or a disadvantage in tissues that maintain organismal homeostasis. Unlike unicellular organisms, cells residing in homeostatic tissues coordinately function toward a common physiological goal. One may think that gene expression in such systems would be tuned to an optimal set point with minimal variability among cells. Conversely, expression variability could give rise to sub-populations of cells that can rapidly respond to changing environmental stimuli. Such division of labor may be advantageous for metabolic tissues that modulate their function on a rapid timescale. The extent of bursty transcription and expression variability in mammalian tissues has so far not been explored.

Quantifying intrinsic variability in a tissue is challenging, because tissues are highly heterogeneous and spatially structured. Tissues are often polarized by directional blood flow or morphogenes and thus the physical location of a cell within the tissue is a major extrinsic determinant of gene expression that must be controlled for. The mammalian liver is a prime example of these features. The liver is composed of repeating anatomical units termed lobules, which are polarized by blood flowing from an upstream "periportal zone" to a downstream "pericentral zone" (Hoehme et al., 2010). These zones differ in the levels of oxygen, nutrients, and hormones, as well as the expression levels of genes (Jungermann and Kietzmann, 1996; Gebhardt, 1992; Braeuning et al., 2006). An additional source of heterogeneity in the liver is its polyploidy (Celton-Morizur and Desdouets, 2010; Duncan et al., 2010; Pandit et al., 2013). The liver consists of a mixture of hepatocytes with either one or two nuclei, where each nucleus has 2, 4, 8, or 16 copies of each chromosome. Thus, the liver is composed of multiple sub-populations distinguished by ploidy and tissue location. Fewer than 0.1% of hepatocytes are cycling at any given time (Duncan 2013), and so the contribution of cell-cycle stage to gene expression variability in the liver can be neglected. Assessing intrinsic variability among hepatocytes must take into account the key sources of heterogeneity in this tissue and focus on expression differences between cells of identical ploidy residing at the same spatial zone.

Several techniques enabled inference of promoter bursting kinetics from fluorescent reporters (Elowitz et al., 2002; Raser and O'Shea, 2004; Suter et al., 2011) or single-molecule time-lapse studies of promoter dynamics (Darzacq et al., 2007; Larson et al., 2011); however, these are not suitable for intact tissues. Single molecule fluorescence in situ hybridization (smFISH) can identify mature as well as nascent transcripts of endogenous genes (Raj et al., 2008; Zenklusen et al., 2008; So et al., 2011; Boettiger and Levine, 2013; Itzkovitz et al., 2012; Little et al., 2013; Senecal et al., 2014) and has the

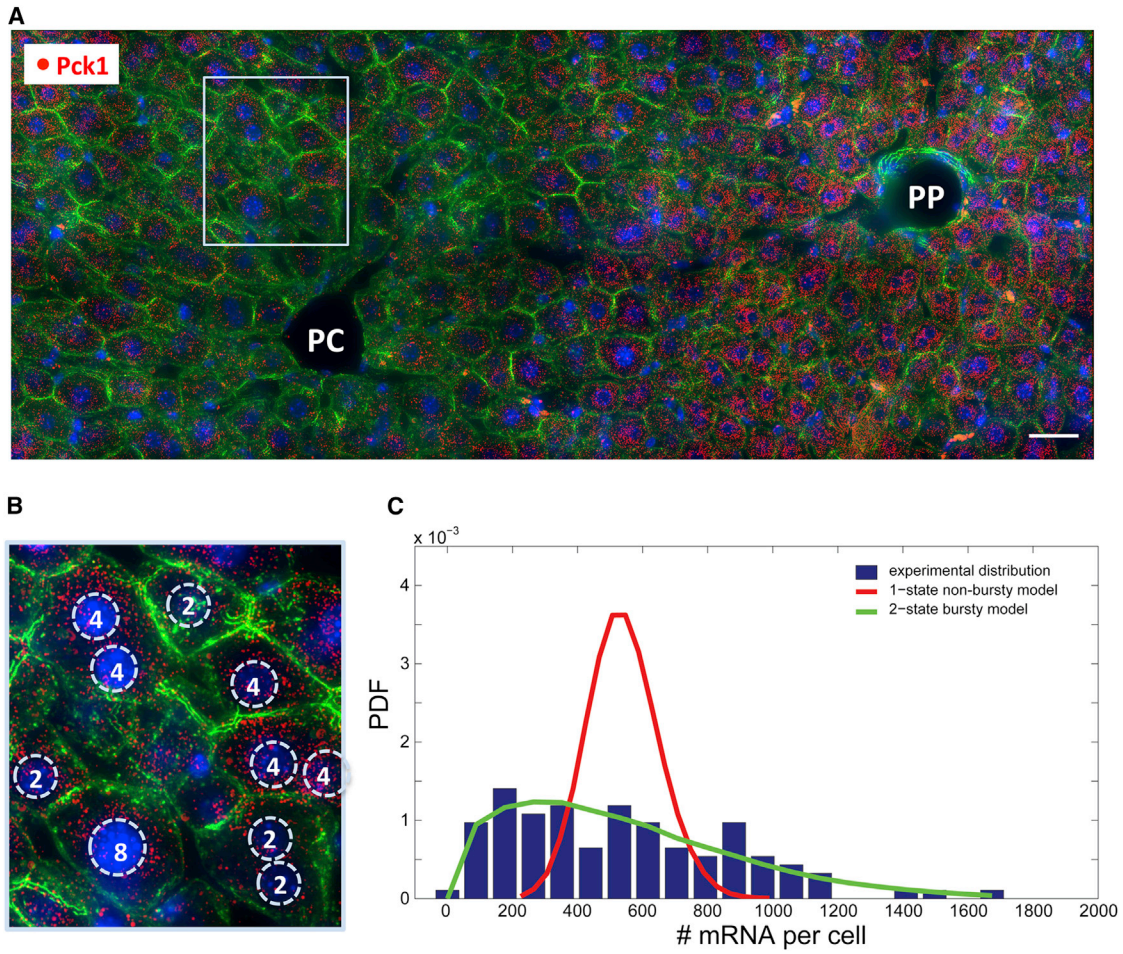


Figure 1. Single Molecule Measurements of Intrinsic Variability in the Intact Mouse Liver

(A) Single molecule transcript counting enables controlling for ploidy and spatial location of cells. Red dots are single mRNA molecules of Pck1, blue are DAPI-stained nuclei, and green is phalloidin membrane staining. PP, periportal zone. PC, pericentral zone. Scale bar is 30 μm. Image is a maximal projection of 12 optical sections spaced 0.3 μm apart.

(B) Magnified view of the boxed region in (A) showing polyploid hepatocytes with one or two nuclei, each with either two, four, or eight copies of each chromosome.

(C) Hepatocytes of the same ploidy and tissue zone exhibit substantial intrinsic variability in gene expression. Blue bars are the distributions of the numbers of Pck1 mRNA per cell in tetraploid hepatocytes in the pericentral zone. Red, theoretical probability distribution function (PDF) expected from a one-state non-bursty model; green, distribution of a bursty two-state model. See also Figure S1.

potential of controlling for both cell type and physical locations of cells within a tissue. Here we develop a methodology based on smFISH to quantitatively characterize promoter states and bursting properties of cells in intact mammalian tissues. We find that bursty transcription is the common mode of gene expression in the liver; however, tight coordination of transcription and degradation as well as liver polyploidy reduce the burst-associated noise.

RESULTS

Hepatocytes Exhibit Extensive Intrinsic Gene Expression Variability

To assess the intrinsic variability in the expression of liver genes, we imaged individual mRNA molecules in mouse liver frozen sections using smFISH (Itzkovitz et al., 2012; Lyubimova et al., 2013) (Figures 1 and S1). We used simultaneous DAPI nuclear staining and phalloidin membrane staining to assign mRNA dots to individual cells. We developed an in situ ploidy classification algorithm (Supplemental Experimental Procedures) that enabled stratifying our single-cell mRNA counts by both tissue zone and ploidy class (Figures S1D–S1F). Strikingly, hepatocytes of the same ploidy level and at the same lobule zone exhibited highly variable mRNA levels, spanning up to two orders of magnitude (Figure 1C). Cellular mRNA distributions within these apparently uniform populations were much broader than expected based on a non-bursty one-state promoter model. In contrast, they were well fitted by a two-state bursty model (Supplemental Experimental Procedures; Figures 1C and 2A), indicating that promoter bursting could be at play.

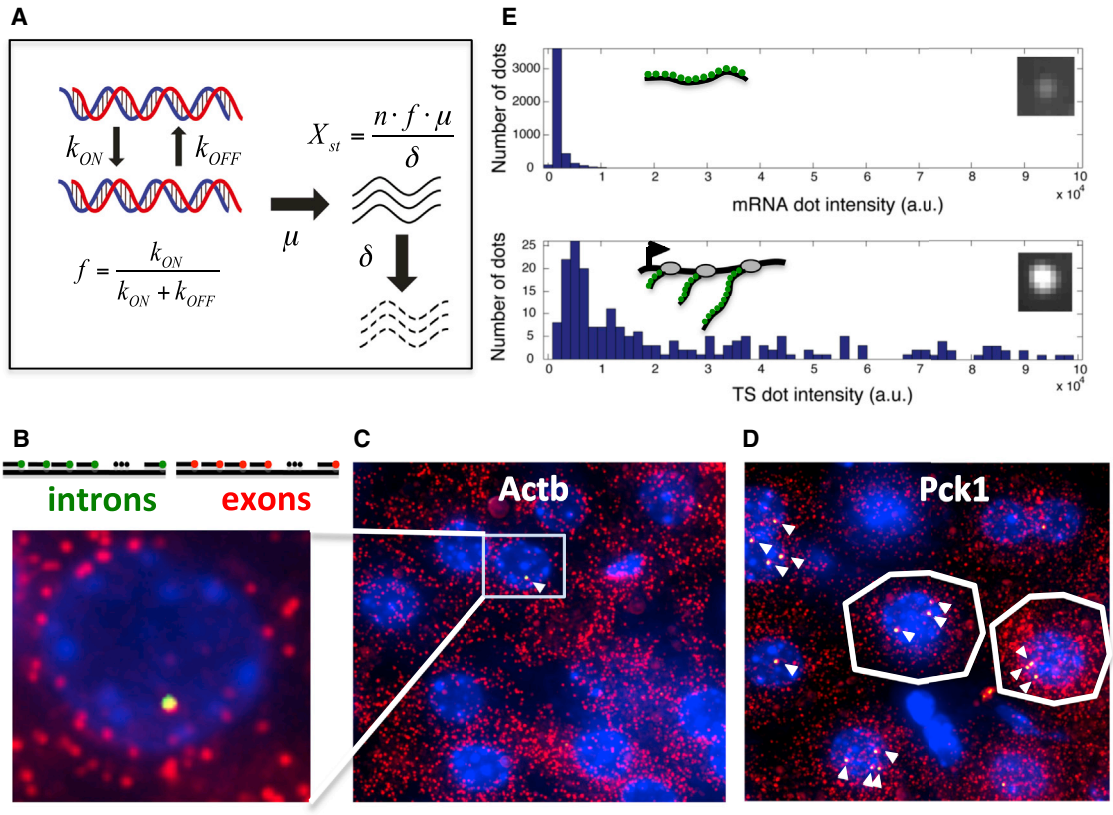


Figure 2. Single Molecule Detection and Quantification of Bursting Promoters in the Intact Mouse Liver

(A) Two-state bursting model of gene expression. *f* is the fraction of promoters that are actively transcribing, μ is the transcription rate from an active promoter, δ is the mRNA degradation rate, and *n* is the number of gene copies (ploidy). k_{ON} and k_{OFF} are the rates of promoter opening and closing, respectively.

(B) Dual color labeling of introns and exons reveals active TSs. Red dots are Actb mRNA detected using a probe library targeting the exons. Green dots are pre-mRNA detected using a probe library targeting the introns.

(C) Actb exhibits rare TSs with low transcription rate and stable mRNA (low *f*, μ, and δ).

(D) Pck1 exhibits abundant intense TSs and high degradation rates (high *f*, μ, and δ). Outlined are two adjacent hepatocytes with substantial difference in transcript counts. Arrowheads mark TSs.

(E) Ratio of intensities of exon dots at TSs to those of mature mRNA facilitates extracting polymerase occupancy and transcription rate (μ). Shown are the distributions of the exonic channel intensities of Pck1 for non-TS dots (top) and TS dots (bottom). Inset shows dot examples. Gray ovals represent Pol2 molecules; green dots represent smFISH probes. See also Figures S2 and S3.

Dual-color Labeling of Introns and Exons Facilitates Quantification of Promoter States, Transcription Rates, and Transcript Lifetimes In Situ

The single-cell variability we observed in hepatocytes may not necessarily be a result of promoter bursting. There could be additional extrinsic sources of variability other than physical location and ploidy among the studied cell population. To directly link variability to promoter bursting, we therefore sought to develop a method that would enable unambiguous identification and quantification of promoter states and transcription rates in situ (Figure 2A). To this end, we designed a second smFISH probe library coupled to a different fluorophore, targeting the intron gene segments. Introns are spliced and degraded co-transcriptionally (Levesque and Raj, 2013; Vargas et al., 2011) and indeed intronic probe libraries yielded dots, which resided only in the nucleus (Figures 2B–2D). The number of double-labeled dots was smaller or equal to the expected number of loci, estimated from our ploidy classification. These dots disappeared following actinomycin D treatment (Figure S2), demonstrating they are indeed active transcription sites (TSs). Thus, the fraction of actively transcribing promoters ("burst fraction," denoted by *f*; Figure 2A) could be readily computed from the ratio of the double-labeled nuclear dots and the expected number of gene loci.

We next calculated the number of nascent mRNA residing at each TS, as the ratio of intensities of the exonic TS dots to the intensity of cytoplasmic dots, representing single mRNA (Figure 2E) (we included a correction factor for the physical spread of the library along the gene; Supplemental Experimental Procedures; Figure S3). Using intensity ratios of probe libraries targeting both ends of the gene, we demonstrated that >85% of nascent mRNAs at TSs are attached to actively transcribing polymerase molecules (Figure S3). Thus, the number of nascent mRNA can be used as a proxy for polymerase occupancy (*M*). We next used polymerase occupancy to infer the transcription rate μ, the average rate of mRNA production from an active TS, using the equation $\mu = M \cdot v / L$, where *L* is the length of the

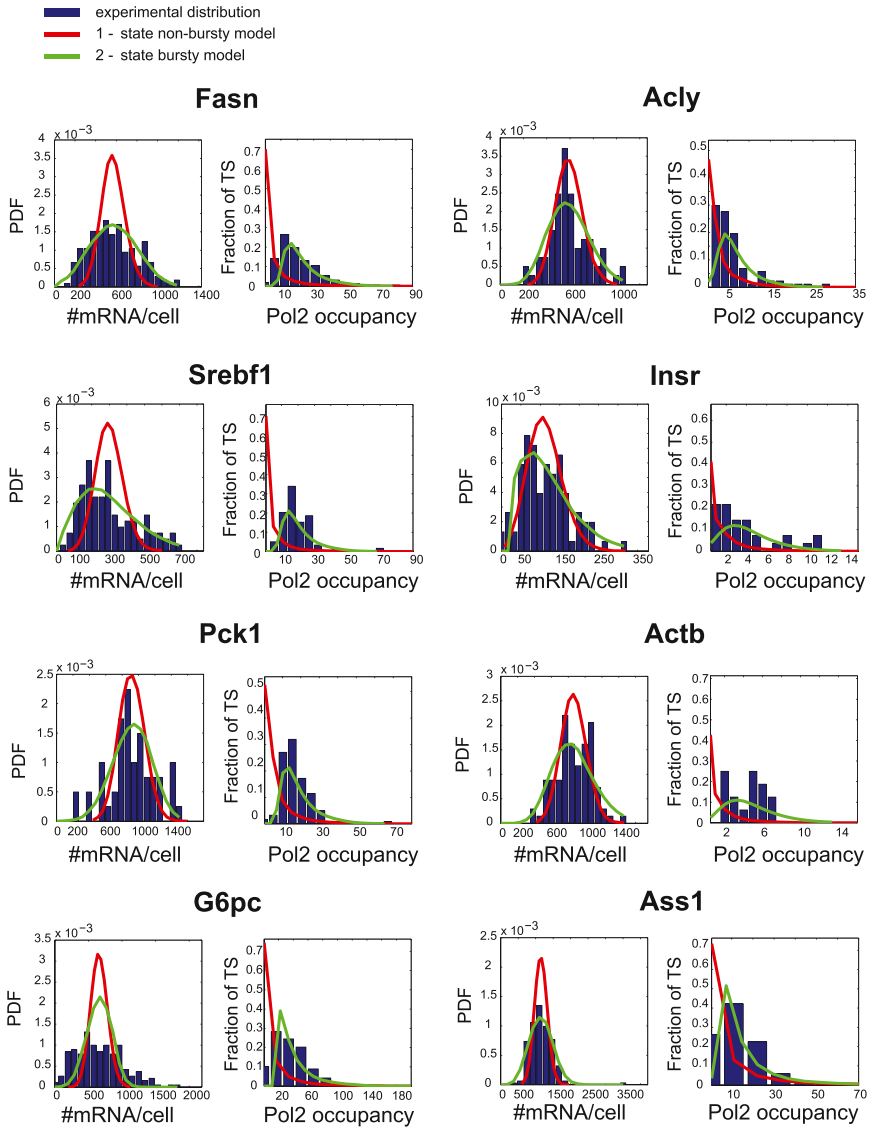


Figure 3. Distributions of Cellular mRNA Content and of Polymerase Occupancies Fit a Two-State Bursty Model

Left plots for each gene are the probability distribution functions (PDFs) of the number of cytoplasmic mRNA per cell; right plots show the PDF of Pol2 occupancies (*M*). Red is a one-state non-bursty model fit; green is a two-state bursty model fit. Both distributions show a better fit to a two-state bursty model for all genes (Table S1 provides mean square errors of the model fits). All data is for tetraploid hepatocytes in the periportal zone. See also Figures S4 and S5.

gene and v = 34 ± 11 bp/s is the polymerase speed, which we calibrated using actinomycin D treatment (Experimental Procedures). Finally, we computed the transcript degradation rate δ as the ratio of summed cellular TS transcription rates and total number of cytoplasmic mRNA. We verified our estimates by tracking mRNA decline following cessation of transcription via actinomycin D treatment (Supplemental Experimental Procedures), yielding estimate errors of 15%. Thus, our approach enabled measurement of all gene expression parameters (Figure 2A)—the fraction of time a promoter is actively transcribing (burst fraction, f), the transcription rate from an active promoter state (μ), and the mRNA degradation rates (δ), in addition to the total cellular mRNA levels.

Liver Genes Are Expressed in Transcriptional Bursts

We applied our measurements to nine liver genes, the gluconeogenic genes Phosphoenolpyruvate carboxykinase 1 (Pck1) and Glucose-6-phosphatase (G6pc), the lipogenic genes ATP citrate lyase (Acly) and Fatty acid synthase (Fasn), the ammonia detoxification genes Argininosuccinate synthase 1 (Ass1) and Glutamine synthetase (Glul), the transcription factor Sterol regulatory element-binding factor 1 (Srebf1), and the Insulin receptor (Insr) as well as a housekeeping gene, Beta-actin (Actb). We measured burst parameters in the pericentral and periportal zones in three metabolic states (fasting, fed, and highly fed; Experimental Procedures). We found that promoters of most genes studied were bursting rather than constantly active (Figures 2B–2D and 3). Only a subset of the expected gene loci in each nucleus was transcriptionally active (exhibiting nascent mRNA); however, those that were active had polymerase occupancies that were too high to be accounted for by a one-state non-bursty model (Figure 3). A notable exception that exhibited non-bursty transcription was Glul, for which all gene loci were transcriptionally active (Figure S4A).

The numbers of active TSs per cell were binomially distributed among cells for all genes studied (Figure S4B), indicating independent bursting. Moreover, combinatorial labeling (Levesque and Raj, 2013) of TSs for Pck1 and G6pc, two functionally related gluconeogenic genes, revealed non-correlated bursting, further supporting intrinsic variability as the source of the differences in gene expression among hepatocytes that reside in the same tissue zone (Figures S4B and S4C). We used the independent bursting feature to extend a previous analytical two-state promoter bursting model (Raj et al., 2006) to multiple alleles (Supplemental Experimental Procedures). This enabled inference of the absolute rates of promoter transitions between an ON and OFF state (Figures 2A and 3; Table S1). We find that burst parameters differ widely between genes, ranging from intense frequent bursts for Pck1 ($f \approx 0.73$, $M \approx 17$ Pol2/TS, $k_{ON} \approx 0.65 hr^{-1}$, $k_{OFF} \approx 0.24 hr^{-1}$) to rare and weaker bursts for Actb ($f \approx 0.03$, $M \approx 5$ Pol2/TS, $k_{ON} \approx 0.08 hr^{-1}$, $k_{OFF} \approx 2.2 hr^{-1}$; Table S1).

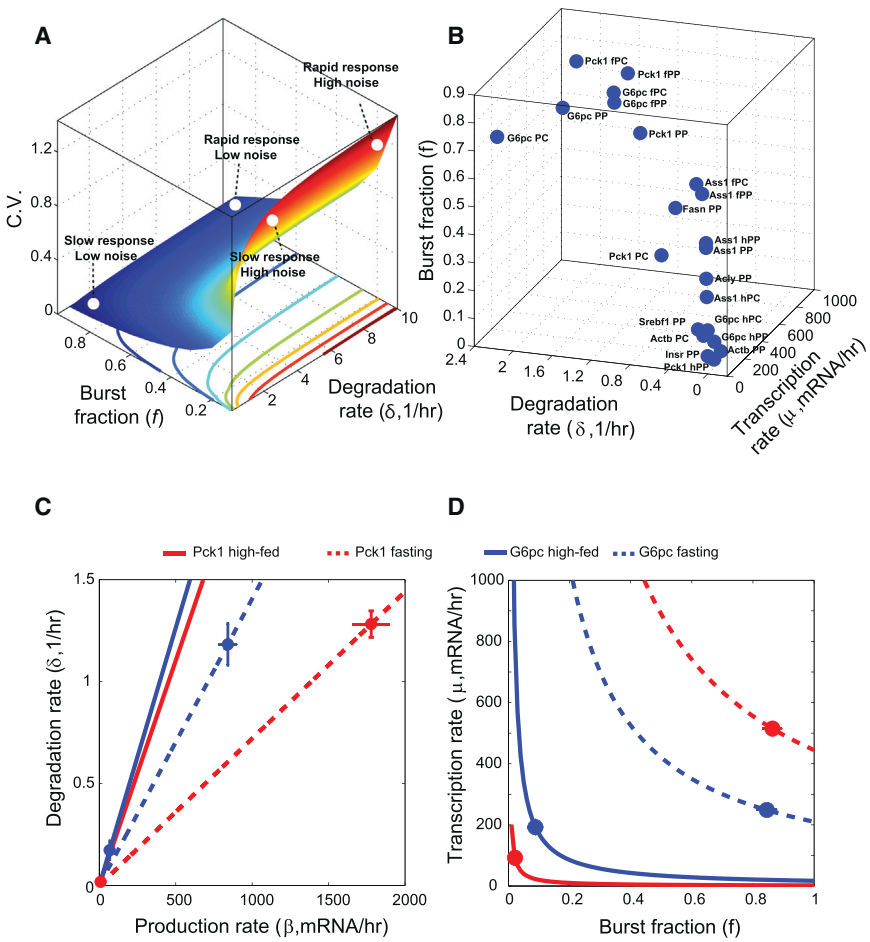


Figure 4. Burst Parameters Facilitate Rapid Response while Minimizing Noise

(A) Coefficient of variation (C.V.) of cellular mRNA among cells with identical mean expression (set to 100 mRNA/cell) increases with degradation rate and decreases with burst fraction. Supplemental Equation [18] was used with $k_{ON} = 0.5$ to generate the distributions used for calculating C.V. μ was tuned for every combination of f and δ to maintain the same steady state.

(B) 3D expression parameter space of liver genes. PP, periportal zone; PC, pericentral zone. f prefix denotes fasting state, and h prefix denotes highly-fed state.

(C) Projection of gene expression space of G6pc (blue) and Pck1 (red) in the periportal zone in fasting (dashed lines) and high-fed (solid lines) metabolic states. x axis is production rate ($\beta = 4 \cdot f \cdot \mu$); y axis is degradation rate (δ). Data are represented as mean ± SEM.

(D) Projections of the gene expression space on transcription rate (μ) and burst fraction (f) axes. Higher expression in fasting states is attained by increasing burst fraction and degradation rates and to a lesser extent transcription rate. Data are represented as mean ± SEM. Lines in (C) and (D) have identical mean transcript level, obtained by different combinations of the gene expression parameters.

To assess whether the bursty gene expression observed is unique to the liver, we repeated our measurements on an additional metabolic tissue—the mouse small intestinal epithelium. We found that most genes were expressed in transcriptional bursts in this tissue as well (Figure S5). Actb was expressed in a non-bursty manner in the intestine, and in a bursty manner in the liver, whereas Glul, which was non-bursty in the liver, exhibited bursty transcription in the intestine (Figure S5).

Burst Fraction and Transcript Degradation Rates Are Tightly Correlated to Enable Rapid Response while Minimizing Noise

Different combinations of burst fraction, transcription rate, and mRNA stability can achieve a required level of cellular mRNA with differential impact on features such as noise and response time (Rabani et al., 2011; Schwanhäusser et al., 2011). For example, coordinated high transcription and degradation facilitates rapid changes in gene expression but can increase intrinsic noise stemming from promoter bursting (Figure 4A). Increasing burst fraction coordinately with degradation rate can dampen the increase in noise. Indeed, we find a tight positive correlation between burst fraction and degradation rate for all genes and conditions studied ($R = 0.78, p = 2.2 \cdot 10^{-5}$, Figure 4B). Thus, genes that should respond fast tend to be less bursty than genes with long mRNA lifetimes. Conversely, the long mRNA lifetime we observe for the bursty genes could be a mechanism for minimizing burst-associated noise through temporal averaging of the stochastic burst events.

To further explore noise-response tradeoffs in liver gene expression parameter space, we performed our measurements on Pck1 and G6pc, the key genes controlling hepatic glucose output, in fed, fasting, and refed mice. These conditions have been shown to lead to drastic changes in mRNA levels for these genes (Gebhardt, 1992; Jungermann and Kietzmann, 1996). We find that Pck1 and G6pc are upregulated in fasting conditions through a coordinate increase in both transcript production ($\beta = n \cdot f \cdot \mu$) and degradation rates (δ) compared to a high-fed state (Figures 4C and 4D). High degradation rates enable a rapid decline in transcript numbers after a 1 hr period of refeeding (Figure 5). Interestingly, the increased transcript production is mainly a result of an increase in burst fraction (f, 10-fold increase for G6pc and 40-fold increase for Pck1; Table S1) and a more modest increase in transcription rate (μ, 1.4-fold for G6pc and 6-fold for Pck1; Figure 4D; Table S1). The increase in transcript production rate predominantly via increased burst fraction is consistent with a strategy of minimizing burst-associated noise (Figure 4A).

High mRNA Degradation Rates of G6pc Impact Protein Dynamics, mRNA Intra-Cellular Localization, and Correlations of mRNA Content with Transcription Rates

Our in situ measurements indicated that the gluconeogenic gene G6pc has a particularly short transcript lifetime of ~20–30 min

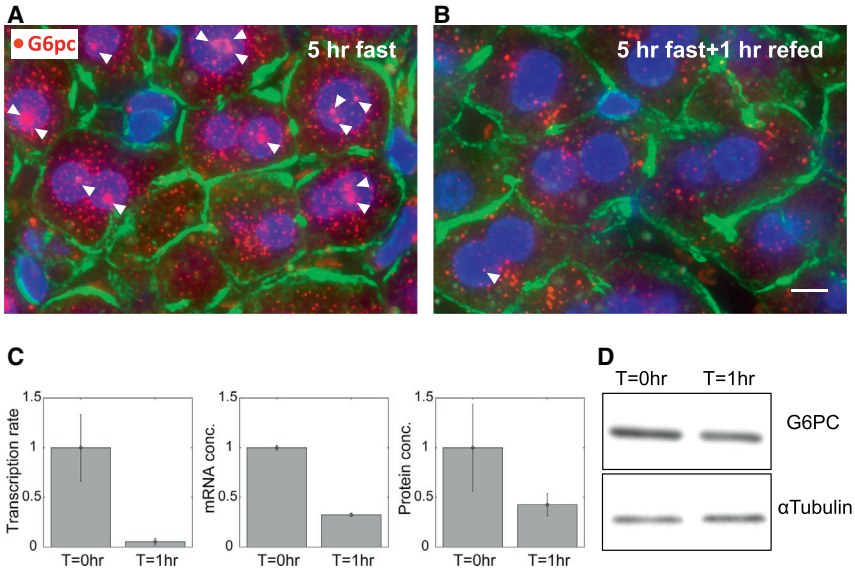


Figure 5. Elevated mRNA and Protein Degradation Rates Facilitate Rapid Decline in mRNA and Protein Levels for G6pc over 1 hr

(A and B) G6pc mRNA (red dots) in mice fasted for 5 hr before (A) and after (B) 1 hr of refeeding. Arrowheads mark TSs. Scale bar is 5 μm.

(C) Decrease in transcription rate (left), cytoplasmic mRNA concentration (middle), and protein concentration (right) over an hour of refeeding.

(D) Representative western blot for G6PC protein used to calculate the decline in G6PC protein levels presented in (C). α-Tubulin (bottom) was used as a loading control. Data are represented as mean ± SEM. See also Figure S6.

(Figures 4B and 5; Table S1). To assess the impact of this feature on protein content, we measured G6PC protein levels before and after 1 hr of refeeding. Strikingly, we observe not only an almost complete shutdown of transcription and a decline of ~70% in mRNA concentrations, but also a decline of ~60% in protein concentrations over this period (Figures 5C and 5D). This decline indicates that G6PC protein is also highly unstable under these conditions (protein half-life of 10–45 min, depending on whether translation rates change; Supplemental Experimental Procedures). This half-life is particularly short, considering that median protein half-lives are on the order of 50 hr in mammalian cells (Schwanhäusser et al., 2011).

Transcripts of G6pc also exhibited non-random localization of mRNA with the majority of mRNAs spatially clustered around the nucleus (Figure S6A). Considering previous estimates of mRNA diffusion rates of ~0.03 $\mu m^2/s$ (Vargas et al., 2005), this clustering could suggest that the mRNA lifetime for this gene may be smaller than the diffusion time within the large hepatocyte volume. At the single-cell level, we observed significant correlation between cellular transcription rates and mRNA levels for unstable transcripts such as G6pc, but not for stable transcripts such as Actb (Figures S6B and S6C). The levels of short-lived mRNA are expected to track the instantaneous promoter activity more tightly compared to long-lived mRNA, supporting our estimates for mRNA lifetimes (Taniguchi et al., 2010). Thus, mRNA degradation rates impact the cellular localization of transcripts in hepatocytes, as well as the correlations between instantaneous transcription rates and cellular mRNA levels.

Hepatocyte Polyploidy Reduces Gene Expression Noise

Unlike most tissues in our body, which consist of mononucleated cells with diploid genomes, the liver is a polyploid tissue, consisting of hepatocytes with either one or two nuclei where each nucleus has either two, four, eight, or 16 copies of each chromosome. This feature is highly ubiquitous with more than ~85% of hepatocytes harboring more than two genomic copies in the tissues studied (15% diploids, 75% tetraploids, and 8% octoploids). The functional advantages of liver polyploidy remain unclear. Our observation of independent promoter bursting (Figure S4) suggested that polyploidy could serve as an additional mechanism to reduce burst-associated noise.

The noise reduction potential of polyploidy could be understood by considering the distinct effects of polyploidy on averages and variability of mRNA concentrations. A tetraploid hepatocyte has twice the number of copies for each gene compared to a diploid hepatocyte, as well as twice the volume (Pandit et al., 2013); hence, the average mRNA concentration should be the same if there is no ploidy-specific regulation. In contrast, variability between the cytoplasmic concentrations among tetraploid cells should be lower than for diploid cells due to the averaging of more stochastic, independent events (Figure 6; Supplemental Experimental Procedures). Indeed, for eight out of ten genes and conditions for which median mRNA concentrations were not significantly different between diploids and tetraploids, the coefficients of variation (C.V.) of the concentrations were significantly lower in tetraploid cells compared to diploid cells (Fisher's combined probability $p < 10^{-16}$ for all genes tested, median reduction of 13% in C.V.), and no gene had a statistically significant higher C.V. in tetraploids (Figure 6C). These results suggest that polyploidy may be an additional mechanism that could serve to reduce gene expression noise.

DISCUSSION

Our work provides direct evidence of bursty gene expression in intact mammalian tissues. We found that liver promoters stochastically switch between transcriptionally active and inactive states, generating intrinsic variability between cells that are considered identical in terms of ploidy and tissue location. Interestingly the liver seems to possess features that can dampen this variability, through temporal averaging and polyploidy. These two mechanisms effectively increase the number of stochastic transcriptional burst events that contribute to cellular mRNA levels.

We found that mRNA lifetimes of the more bursty genes, the ones that are transcriptionally active for shorter periods, tend

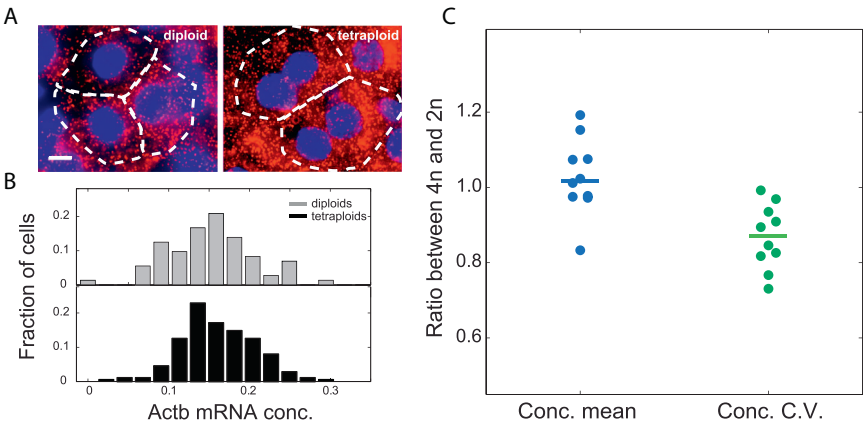


Figure 6. Tetraploid Hepatocytes Have Reduced Gene-Expression Noise Compared to Diploid Hepatocytes

(A) Examples of Actb expression among diploid cells (left) and tetraploid cells (right). Cell outlines (white dashed lines) are based on phalloidin membrane staining. Scale bar is 5 μm.

(B) Probability distribution function of mRNA concentrations in diploid (gray) and tetraploid (black) hepatocytes for Actb in the periportal zone of a 5-month-old mouse in a fed state. Coefficients of variation (C.V.s) are 0.35 for diploids and 0.28 for tetraploids.

(C) Single-cell variability in cytoplasmic mRNA concentrations are smaller in tetraploid (4n) hepatocytes compared to diploid (2n) hepatocytes. Every dot is a gene in one of the conditions studied; shown are the ratios of mean cytoplasmic concentration (blue) and C.V. (green). All genes and conditions analyzed had mean concentrations that were not significantly different, and eight out of these ten genes had significantly lower C.V.s. in tetraploids.

to be longer, reducing the burst-associated variability through temporal averaging. As a result of this tight correlation between mRNA degradation rates and burst fractions, gene expression parameter space (the 3D space spanned by burst fraction, transcription rate, and degradation rate) is relatively sparse (Figure 4B). If the liver evolved strategies to reduce noise, what then could be the functional advantage of bursty transcription in this tissue? Non-bursty transcription, consisting of open chromatin continuously transcribing mRNA, would have been more effective than bursty transcription in reducing variability. One possible advantage of the bursty transcription observed could be protection of "closed" DNA from damage, a feature that has been previously attributed to nucleosome structure (Chen et al., 2012). Such protection could be particularly important given the detoxification roles of the liver. An additional advantage of reducing the amount of accessible DNA may be the minimization of misbinding events of transcription factors (Shinar et al., 2006).

The liver is a polyploid tissue containing either mono-nucleated or bi-nucleated cells where each nucleus contains two, four, eight, or 16 chromosomal copies. The benefits of polyploidy to mammalian cells in general, and particularly for the liver, remain unclear (Duncan 2013; Pandit et al., 2013; Storchova and Pellman, 2004; Tang and Amon, 2013). Existing theories include the ability to harbor backup copies of genes to protect cells against mutations. Other advantages could include the higher biosynthetic capacity of polyploid cells, supporting a larger cell size, the ability to generate functional diversity (Duncan et al., 2010), and the ability to modulate surface to volume ratios (Storchova and Pellman, 2004). Our work uncovered an additional possible benefit of polyploidy-reducing gene expression noise. Tetraploid hepatocytes have twice the number of gene copies as well as twice the volume as diploid hepatocytes, and indeed, we find that the average expression of genes in diploid and tetraploid hepatocytes is mostly unchanged (Figure 6). In contrast to the average expression, variability of mRNA concentrations tends to be lower in tetraploid hepatocytes compared to diploid hepatocytes. This effect can be achieved by spatially averaging more stochastic burst events per cell. Similar spatial averaging effects have been demonstrated in a polyploid mutant of *Bacillus subtilis* (Süel et al., 2007) and in *Drosophila* embryos, which are syncytia containing multiple nuclei (Little et al., 2013). Liver polyploidy increases with age (Celton-Morizur and Desdouets, 2010), as does single-cell variability (Bahar et al., 2006). An intriguing hypothesis is that liver polyploidization may counteract the noise increase caused by genes becoming burstier with age. Examining the relation between polyploidy and gene expression noise in liver aging, in liver pathology, and in other polyploid tissues such as heart and muscles could provide important insight into the role of this enigmatic feature of mammalian tissues. Measuring the impact of hepatocyte noise on the performance functions of different liver metabolic tasks and comparing metabolic performances in mouse models with perturbed polyploidy (Pandit et al., 2012) can reveal the functional significance of the noise reduction feature of liver polyploidization described in this study.

The liver switches between two distinct modes in terms of glucose metabolism—a glucose absorber following a meal and a glucose producer in between meals. Much of the control of this process is achieved by modulating mRNA levels of the gluconeogenic genes Pck1 and G6pc, the expression levels of which are high in a fasting state and low following a meal. Changes in the expression levels of these genes can be achieved by modulating burst fraction, transcription rate, degradation rate, or any combination of these three key features. We found that the higher levels of gluconeogenic genes in a fasting state are achieved via an increase in both transcription and degradation compared to a fed state. The increase in degradation rate comes at a price, as it requires excessive energy expenditure for transcription and accentuates burst-associated noise. A possible explanation for the elevated degradation rates during fasting may lie in the fact that the transition from a fasting to a fed state is rapid, as blood glucose levels rise following a meal within minutes. Upon feeding, it may thus be important to shut

down hepatic glucose output as fast as possible, and indeed, for G6pc, the high degradation rates facilitate rapid reduction in both mRNA and protein levels following refeeding (Figure 5). In contrast, the transition from a fed to a fasting state takes hours, as metabolites are gradually depleted from the circulation, and thus, in a fed state maintaining high mRNA degradation rates (the determinant of response time) may be less important. It will be interesting to explore the mechanisms that facilitate the rapid reduction in mRNA and protein levels during refeeding. Possible regulatory candidates may include microRNA, which have been shown to have important roles in maintaining liver zonation (Sekine et al., 2009).

Our approach for measuring all gene expression parameters, namely burst fractions, transcription rates, and degradation rates in single cells within intact tissues opens new avenues for exploring the design principles that may have shaped gene expression parameter space. This approach can reveal the extent of transcriptional bursts and gene expression noise in diverse mammalian tissues.

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 months were fed normal chow ad libitum, fasting or refed for the indicated times. Mice were sacrificed at 6AM (high-fed state), 9AM (fed state), and 12 PM (fasting state, for these mice food was removed at 8AM). In the refeeding experiment (Figure 5), mice were housed under reverse phase cycle and fasted for 5 hr starting at 7AM. Mice were then refed ad libitum for an hour and sacrificed either before refeeding (two mice) or after refeeding (two mice). Mice were anesthetized and tissues were harvested and fixed in 4% paraformaldehyde for 3 hr, incubated overnight with 30% sucrose in 4% paraformaldehyde, and then embedded in OCT. 25 μm cryosections were used for hybridization of liver tissue, and 12 μm sections were used for intestinal tissue.

Hybridization and Imaging

Probe library constructions, hybridization procedures, and imaging conditions were previously described (Itzkovitz et al., 2012; Lyubimova et al., 2013). To detect cell borders alexa fluor 488 conjugated phalloidin (Rhenium A12379) was added to the GLOX buffer wash. Portal node was identified morphologically on DAPI images based on bile ductile; central vein was identified using smFISH for Glul performed on serial sections. Hepatocytes within the first three layers of the portal node (up to ~50 μm distance) were classified as periportal. Hepatocytes in layers 2–4 from the central vein (~70 μm distance from central vein) were classified as pericentral. We excluded the innermost lobule layer that is directly bordering the central vein, since hepatocytes in that layer exhibited distinct expression levels. All results presented in the paper are for tetraploid (4n) hepatocytes, with the exception of the polyploidy analysis (Figure 6).

Data Analysis

Ploidy classification was based on the 3D nuclear dimension reconstruction, based on DAPI images. Validation of the classification was done with an smFISH probe for Xist on livers of female mice (Figure S1E). TSs were identified as dots that appeared in both the exon and intron probe channels. Exon dot intensity was used to infer the number of nascent mRNA at each TS. To calibrate Pol2 speed, we imaged genes in NIH 3T3 cells before and at different time points following actinomycin D treatment. Burst fraction and transcription rate were measured before treatment, and degradation rates were measured based on the reduction in mRNA levels following treatment and used to fit a single missing variable—Pol2 speed (Supplemental Experimental Procedures).

Fitting Bursting Models to the Measured Distributions

To fit our measured mRNA distributions (Figure 3), we extended the analytical results of Raj et al. (2006) to multiple gene copies by convolving the mRNA distributions predicted for a single gene copy. This extension is based on the independent bursting property observed, where the probability of each promoter to be in a transcriptionally active state is independent of the states of other promoters in the cell (Figure S4B). In our fits, we used the measured burst parameters (f, μ, δ), leaving a single free fit parameter (k_{ON}, since $f = k_{ON}/(k_{ON} + k_{OFF})$). A Poisson distribution was used for the mRNA distributions under the one-state model. For both the one-state and two-state model fits, we included a correction for the broadening effect caused by volume subsampling, namely counting mRNA dots only in a partial volume of the cell rather than the entire cell (Supplemental Experimental Procedures).

Polymerase occupancies for both one-state and two-state models were fit with Poisson distributions. Importantly, while for the two-state model we used the measured mean pol2 occupancy of the active TS as the Poisson parameter $\langle M \rangle$, in the one-state model, the Poisson parameter $\langle M \rangle$ was obtained from the burst fraction $\langle M \rangle = -\log(1-f)$, where f is the fraction of double-labeled dots (TSs), since under a one-state model the fraction of genes where active transcription is not observed is $1-f = e^{-\langle M \rangle}$. The distributions of polymerase occupancies were convolved with a broadening kernel measured based on the intensities of individual mRNA dots (Supplemental Experimental Procedures). Mean squared errors were based on cross validations to avoid over-fitting (Supplemental Experimental Procedures).

Comparing Noise Properties of Different Ploidy Classes

To quantify variability in cytoplasmic concentrations between different ploidy classes, we counted the number of mRNA molecules in ten consecutive z stacks in a cytoplasmic rim surrounding the nucleus with a fixed volume of 500 μm^3. This ensures that variability in concentration will not stem from variability in segmented volumes.

To compare averages and noise of cytoplasmic mRNA concentrations between diploids and tetraploids, we considered only genes and conditions for which at least 15 cells were counted from each ploidy class and for which mean mRNA concentration was not significantly different between the ploidy classes (using Wilcoxon rank-sum tests and false discovery rate of 15%). To ensure comparison of C.V.s of sets of equal lengths of diploids and tetraploids, we sampled 1,000 sets of n cells from the tetraploid class (where n is the number of cells in the diploid class, invariably the smaller class of cells for the tissues studied), recalculated C.V., and computed p values as the fraction of 1,000 resampled tetraploid sets that yielded a higher C.V. than the diploid set. False discovery rate of 15% was used to control for multiple hypothesis testing. Fisher's method was used to obtain a combined p value.

SUPPLEMENTAL INFORMATION

Supplemental Information includes six figures, two tables, and Supplemental Experimental Procedures and can be found with this article online at http://dx.doi.org/10.1016/j.molcel.2015.01.027.

ACKNOWLEDGMENTS

We thank Uri Alon, Ron Milo, Eran Segal, and Jean Hausser for valuable comments. We thank Hagit Shapiro, Sara Weiss, and Yael Kuperman for technical help. S.I. is the incumbent of the Philip Harris and Gerald Ronson Career Development Chair. We acknowledge support from the Henry Chanoch Krenter 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, and the European Research Council under the European Union's Seventh Framework Programme (*FP7/2007-2013*)/ERC grant agreement number 335122.

Received: October 14, 2014
Revised: January 6, 2015
Accepted: January 20, 2015
Published: February 26, 2015

REFERENCES

Bahar, R., Hartmann, C.H., Rodriguez, K.A., Denny, A.D., Busuttil, R.A., Dollé, M.E.T., Calder, R.B., Chisholm, G.B., Pollock, B.H., Klein, C.A., and Vijg, J. (2006). Increased cell-to-cell variation in gene expression in ageing mouse heart. Nature *441*, 1011–1014.

Bar-Even, A., Paulsson, J., Maheshri, N., Carmi, M., O'Shea, E., Pilpel, Y., and Barkai, N. (2006). Noise in protein expression scales with natural protein abundance. Nat. Genet. *38*, 636–643.

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

Blake, W.J., Balázsi, G., Kohanski, M.A., Isaacs, F.J., Murphy, K.F., Kuang, Y., Cantor, C.R., Walt, D.R., and Collins, J.J. (2006). Phenotypic consequences of promoter-mediated transcriptional noise. Mol. Cell *24*, 853–865.

Boettiger, A.N., and Levine, M. (2013). Rapid transcription fosters coordinate snail expression in the Drosophila embryo. Cell Rep. *3*, 8–15.

Braeuning, A., Ittrich, C., Köhle, C., Hailfinger, S., Bonin, M., Buchmann, A., and Schwarz, M. (2006). Differential gene expression in periportal and perivenous mouse hepatocytes. FEBS J. *273*, 5051–5061.

Celton-Morizur, S., and Desdouets, C. (2010). Polyploidization of liver cells. Adv. Exp. Med. Biol. *676*, 123–135.

Chalancon, G., Ravarani, C.N.J., Balaji, S., Martinez-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, X., Chen, Z., Chen, H., Su, Z., Yang, J., Lin, F., Shi, S., and He, X. (2012). Nucleosomes suppress spontaneous mutations base-specifically in eukaryotes. Science *335*, 1235–1238.

Chong, S., Chen, C., Ge, H., and Xie, X.S. (2014). Mechanism of transcriptional bursting in bacteria. Cell *158*, 314–326.

Dar, R.D., Razooky, 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 Turris, 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.

Duncan, A.W. (2013). Aneuploidy, polyploidy and ploidy reversal in the liver. Semin. Cell Dev. Biol. *24*, 347–356.

Duncan, A.W., Taylor, M.H., Hickey, R.D., Hanlon Newell, A.E., Lenzi, M.L., Olson, S.B., Finegold, M.J., and Grompe, M. (2010). The ploidy conveyor of mature hepatocytes as a source of genetic variation. Nature *467*, 707–710.

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

Elowitz, M.B., Levine, A.J., Siggia, E.D., and Swain, P.S. (2002). Stochastic gene expression in a single cell. Science *297*, 1183–1186.

Friedman, N., Cai, L., and Xie, X.S. (2006). Linking stochastic dynamics to population distribution: an analytical framework of gene expression. Phys. Rev. Lett. *97*, 168302.

Gebhardt, R. (1992). Metabolic zonation of the liver: regulation and implications for liver function. Pharmacol. Ther. *53*, 275–354.

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.

Hoehme, S., Brulport, M., Bauer, A., Bedawy, E., Schormann, W., Hermes, M., Puppe, V., Gebhardt, R., Zellmer, S., Schwarz, M., et al. (2010). Prediction and validation of cell alignment along microvessels as order principle to restore tissue architecture in liver regeneration. Proc. Natl. Acad. Sci. USA *107*, 10371–10376.

Itzkovitz, S., Lyubimova, A., Blat, I.C., Maynard, M., van Es, J., Lees, J., Jacks, T., Clevers, H., and van Oudenaarden, A. (2012). Single-molecule transcript counting of stem-cell markers in the mouse intestine. Nat. Cell Biol. *14*, 106–114.

Jungermann, K., and Kietzmann, T. (1996). Zonation of parenchymal and nonparenchymal metabolism in liver. Annu. Rev. Nutr. *16*, 179–203.

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.

Levesque, M.J., and Raj, A. (2013). Single-chromosome transcriptional profiling reveals chromosomal gene expression regulation. Nat. Methods *10*, 246–248.

Little, S.C., Tikhonov, M., and Gregor, T. (2013). Precise developmental gene expression arises from globally stochastic transcriptional activity. Cell *154*, 789–800.

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.

Newman, J.R.S., Ghaemmaghami, S., Ihmels, J., Breslow, D.K., Noble, M., DeRisi, J.L., and Weissman, J.S. (2006). Single-cell proteomic analysis of S. cerevisiae reveals the architecture of biological noise. Nature *441*, 840–846.

Pandit, S.K., Westendorp, B., Nantasanti, S., van Liere, E., Tooten, P.C.J., Cornelissen, P.W.A., Toussaint, M.J.M., Lamers, W.H., and de Bruin, A. (2012). E2F8 is essential for polyploidization in mammalian cells. Nat. Cell Biol. *14*, 1181–1191.

Pandit, S.K., Westendorp, B., and de Bruin, A. (2013). Physiological significance of polyploidization in mammalian cells. Trends Cell Biol. *23*, 556–566.

Pedraza, J.M., and Paulsson, J. (2008). Effects of molecular memory and bursting on fluctuations in gene expression. Science *319*, 339–343.

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

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.

Raj, A., van den Bogaard, P., Rifkin, S.A., van Oudenaarden, A., and Tyagi, S. (2008). Imaging individual mRNA molecules using multiple singly labeled probes. Nat. Methods *5*, 877–879.

Raser, J.M., and O'Shea, E.K. (2004). Control of stochasticity in eukaryotic gene expression. Science *304*, 1811–1814.

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

Sekine, S., Ogawa, R., Mcmanus, M.T., Kanai, Y., and Hebrok, M. (2009). Dicer is required for proper liver zonation. J. Pathol. *219*, 365–372.

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.

Shinar, G., Dekel, E., Tlusty, T., and Alon, U. (2006). Rules for biological regulation based on error minimization. Proc. Natl. Acad. Sci. USA *103*, 3999–4004.

So, L.H., Ghosh, A., Zong, C., Sepúlveda, L.A., Segev, R., and Golding, I. (2011). General properties of transcriptional time series in Escherichia coli. Nat. Genet. *43*, 554–560.

Storchova, Z., and Pellman, D. (2004). From polyploidy to aneuploidy, genome instability and cancer. Nat. Rev. Mol. Cell Biol. *5*, 45–54.

Süel, G.M., Kulkarni, R.P., Dworkin, J., Garcia-Ojalvo, J., and Elowitz, M.B. (2007). Tunability and noise dependence in differentiation dynamics. Science *315*, 1716–1719.

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.

Tang, Y.-C., and Amon, A. (2013). Gene copy-number alterations: a cost-benefit analysis. Cell *152*, 394–405.

Taniguchi, Y., Choi, P.J., Li, G.-W., Chen, H., Babu, M., Hearn, J., Emili, A., and Xie, X.S. (2010). Quantifying E. coli proteome and transcriptome with single-molecule sensitivity in single cells. Science *329*, 533–538.

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.

Vargas, D.Y., Shah, K., Batish, M., Levandoski, M., Sinha, S., Marras, S.A.E., Schedl, P., and Tyagi, S. (2011). Single-molecule imaging of transcriptionally coupled and uncoupled splicing. Cell *147*, 1054–1065.

Zenklusen, D., Larson, D.R., and Singer, R.H. (2008). Single-RNA counting reveals alternative modes of gene expression in yeast. Nat. Struct. Mol. Biol. *15*, 1263–1271.

Molecular Cell, Volume *58*

Supplemental Information

Bursty Gene Expression

in the Intact Mammalian Liver

Keren Bahar Halpern, Sivan Tanami, Shanie Landen, Michal Chapal, Liran Szlak, Anat Hutzler, Anna Nizhberg, and Shalev Itzkovitz

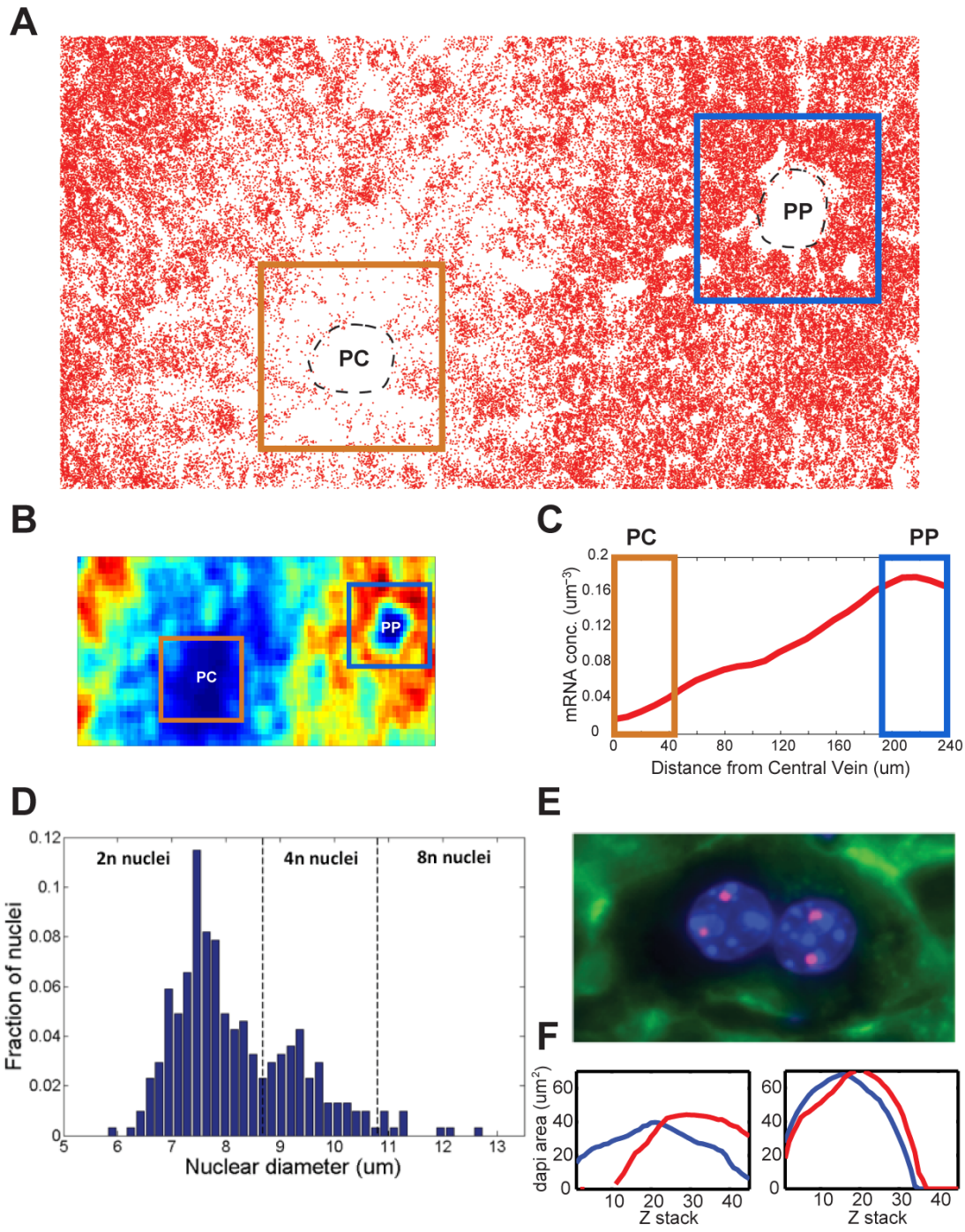


Figure S1 (Related to Figure 1) Single molecule gene expression measurements in the intact liver control for spatial zonation and polyploidy. (**A**) Transcript counting of Pck1 in the intact liver lobule. Each red dot is a mRNA molecule of Pck1, blue square highlights the periportal zone (PP), brown square highlights the pericentral zone (PC). (**B**) Heat map of gene expression in **A**. (**C**) calculation of the concentration of mRNA in different layers around the central vein. Boxes highlight the pericentral and and periportal zones. **(D)** Histograms of nuclear diameters extracted from the

maximal DAPI cross-section. **(E)** smFISH for Xist RNA (red dots). Each dot represents an inactive X-chromosome. Shown is an example of an octoploid hepatocyte, with two tetraploid nuclei, each with 2 inactive X chromosome loci. **(F)** Representative profiles of DAPI area at different optical sections (Z-stack) for bi-nucleated hepatocytes. Blue and Red are the two nuclei profiles. Left – a tetraploid hepatocyte (2 diploid nuclei). Right – an octoploid hepatocyte (2 tetraploid nuclei).

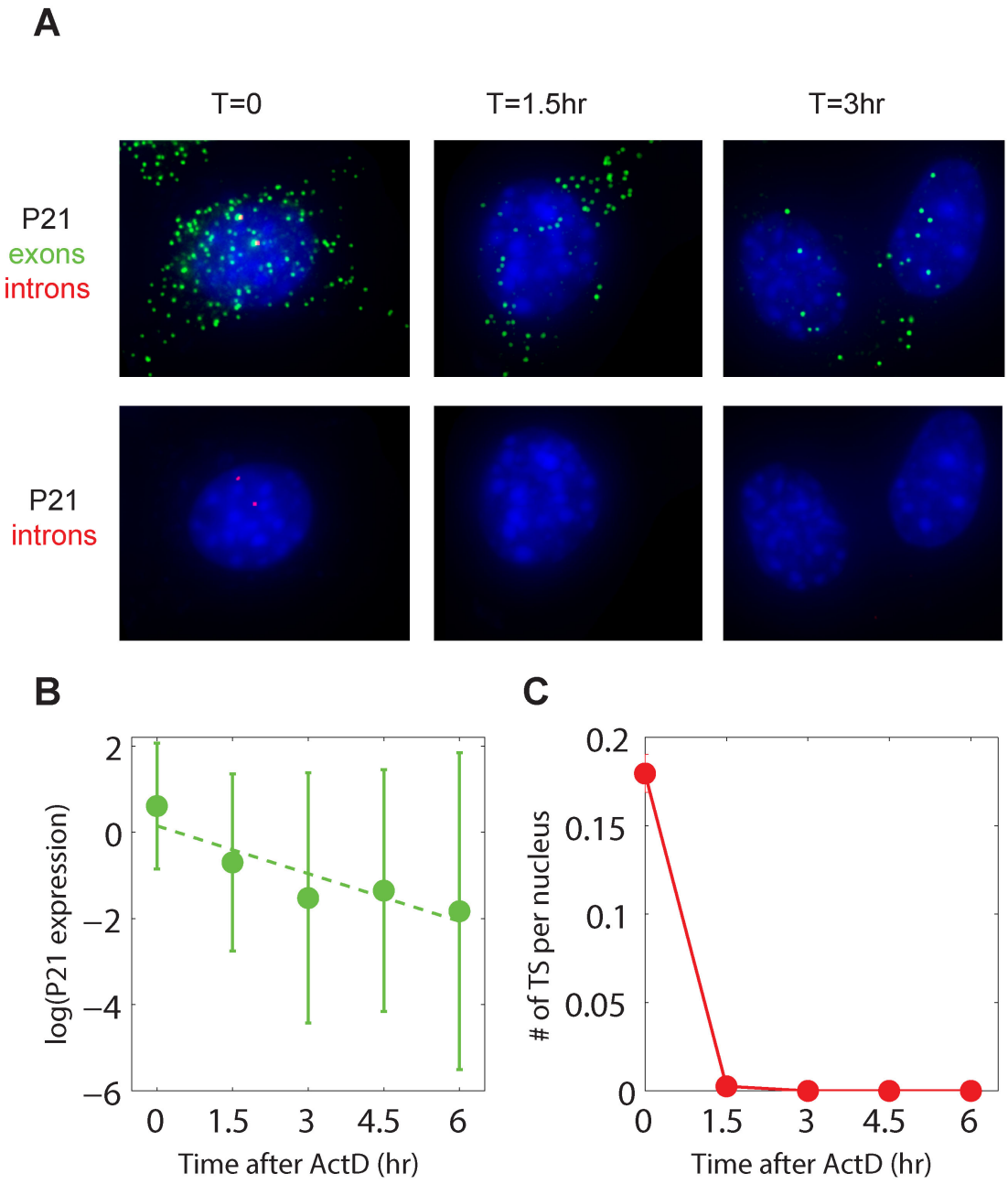
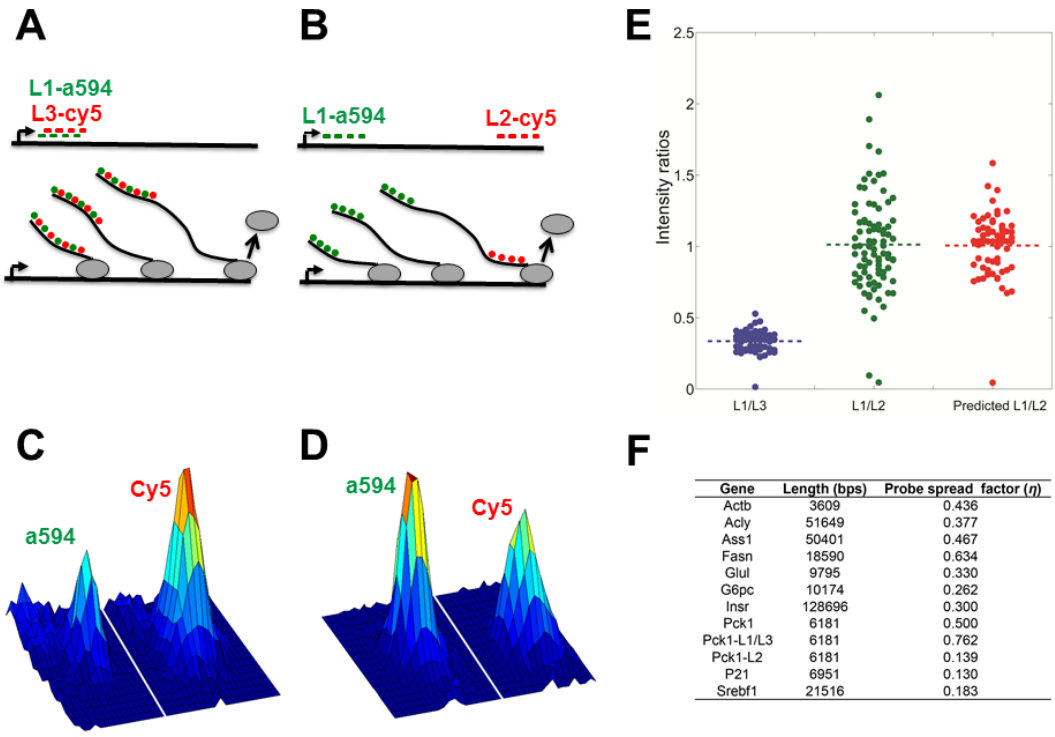


Figure S2 (Related to Figure 2) Actinomycin-D treatment. (**A**) Images of NIH-3T3 cells before Actinomycin-D treatment (left column) and at different time points following 3µg/ml Actinomycin-D (middle and right columns). Red dots are P21 introns, green dots are P21 exons. (**B**) Estimating P21 degradation rate. Shown are the natural logarithms of the number of mRNA dots per cell per stack before and at different time points after Actinomycin-D treatment. The slope of log(P21 expression) vs. time is $0.37 \pm 0.12\ hr^{-1}$. (**C**) Intronic dots of P21 rapidly disappear following actD treatment. Y axis shows the number of TS per nucleus.



Gene	Length (bps)	Probe spread factor (η)
Actb	3609	0.436
Acly	51649	0.377
Ass1	50401	0.467
Fasn	18590	0.634
Glul	9795	0.330
G6pc	10174	0.262
Insr	128696	0.300
Pck1	6181	0.500
Pck1-L1/L3	6181	0.762
Pck1-L2	6181	0.139
P21	6951	0.130
Srebf1	21516	0.183

Figure S3 (Related to Figure 2) Split probe libraries demonstrate that most mRNA molecules at transcription sites are attached to actively transcribing polymerases. (**A-B**) diagrammatic representation of the probe libraries used. (**C**) Intensities of representative TS coupled to L1-a594 and L3-cy5 at the corresponding channels. Cy5 intensity is higher than a594 due to the fluorophore properties. (**D**) Intensities of representative TS coupled to L1-a594 and L2-cy5 at the corresponding channels. Cy5 intensity in (**D**) is lower than in (**C**), indicating that most Pol2 molecules are spread along the gene and not pausing at the 3' end. Surface heights in (**C-D**) are the dot intensities in the optical section of maximal intensity. (**E**) Ratio of the intensities in

the alexa594 channel and the cy5 channel for the L1/L3 experiment (blue), and the L1/L2 experiment (green). Red dots are the intensity ratios of L1/L3 multiplied by the theoretical ratio of equation [15] with F=0.86. (**F**) Gene lengths and Probe spread correction factors for the genes studied.

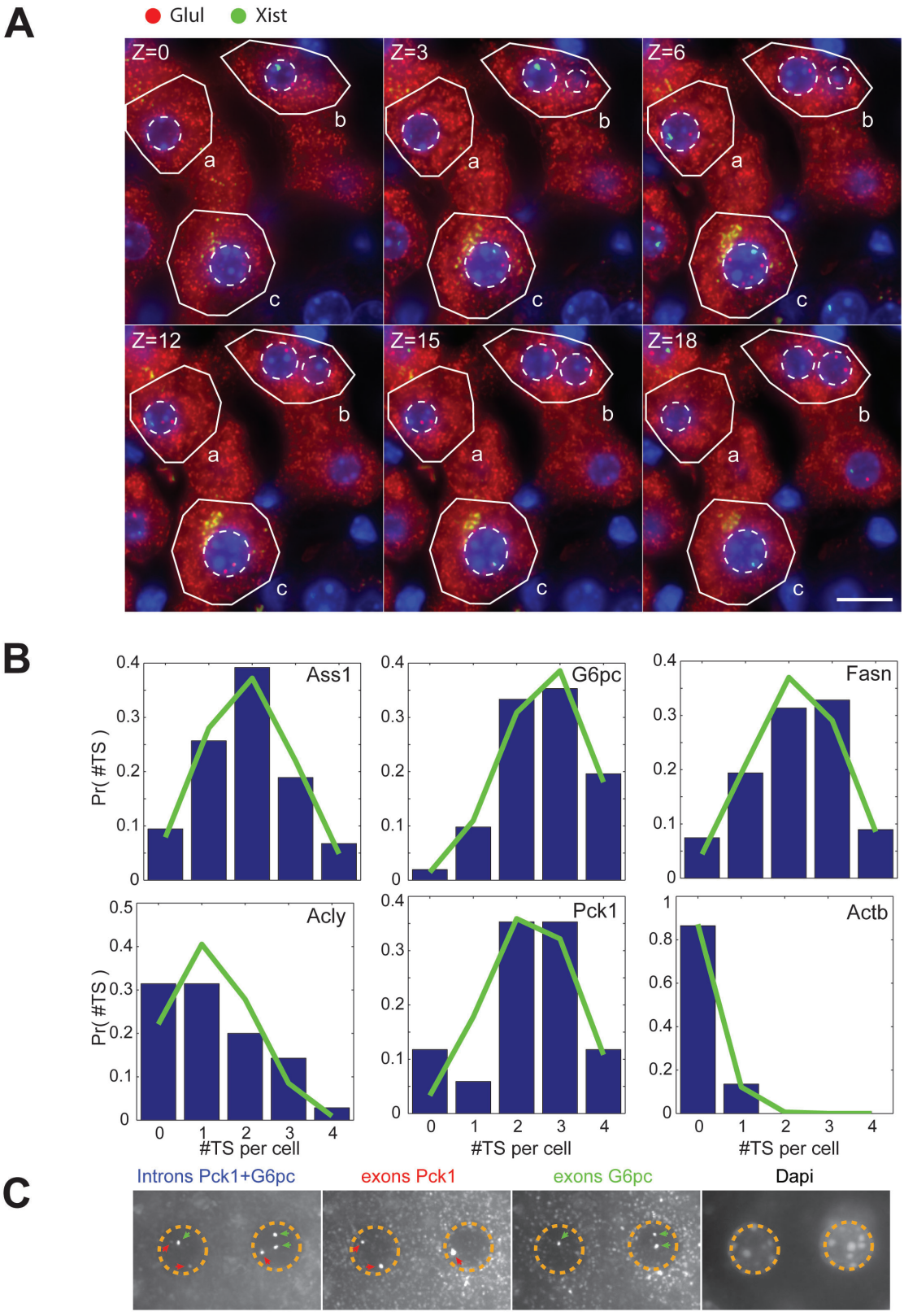
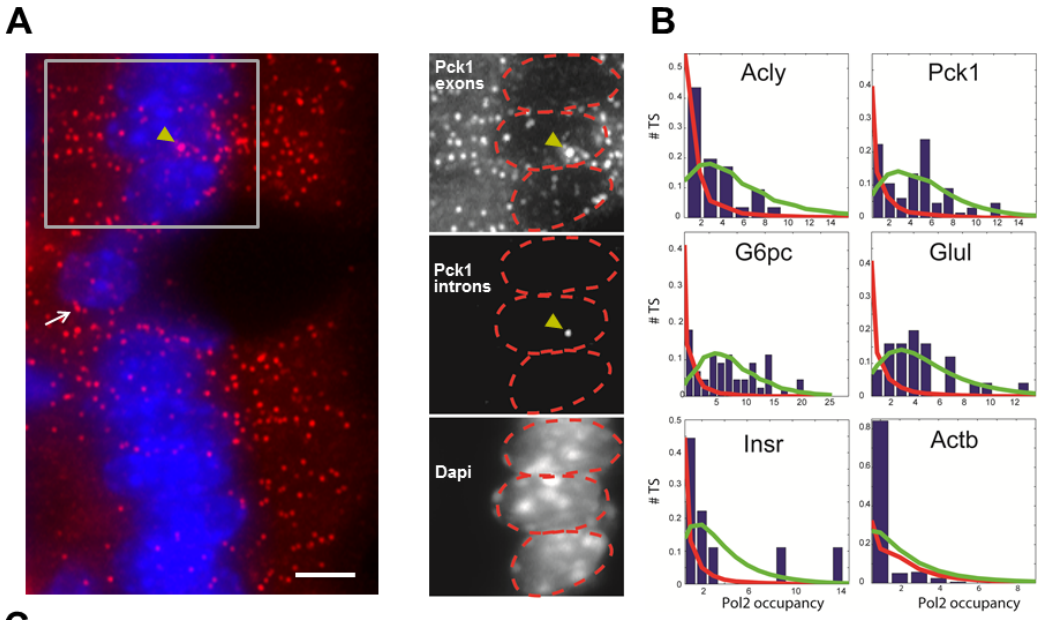


Figure S4 (Related to Figure 3) Liver promoters burst in a non-correlated manner. **(A)** Glutamine Synthetase (Glul) exhibits non-bursty transcription in liver tissues. Shown are 6 representative Z-sections out of a total of 45 imaged stacks, spaced 0.3um apart. Green dots are Xist loci marking inactivated X chromosomes, nuclear

red dots are TS of Glul, blue are nuclei marked with DAPI. Dashed lines mark nuclei, solid lines mark cells. Three cells for which the entire nucleus appeared in the entire Z-stack are shown - cell ***a*** is a mono-nuclear diploid cell showing 1 Xist locus and 2 TS of Glul, cell ***b*** is a bi-nucleated tetraploid cell (two diploid nuclei) in which each nucleus exhibits 1 Xist locus and 2 TS of Glul, cell ***c*** is a mononuclear tetraploid cell exhibiting 2 Xist loci and 4 TS of Glul. **(B)** The numbers of active transcription sites (TS) per cell are binomially distributed. Blue bars are experimental measurements; green is a binomial distribution with parameter *f*, the mean fraction of active TS per site. Data is for tetraploid hepatocytes in the periportal zone. (**C**) Combinatorial smFISH demonstrates independent bursting of Pck1 and G6pc. Transcription sites for Pck1 (red arrows) were detected using probes targeting Pck1 introns labeled with tmr and probes targeting Pck1 exons labeled with alexa594. Transcription sites of G6pc (green arrows) where detected using probes targeting G6pc introns labeled with tmr and probes targeting G6pc exons labeled with cy5. Although hybridization was performed with two distinct libraries labeled with the same fluorophore (Pck1 and G6pc introns, left), TS for these genes could clearly be detected and quantified based on the appearance of double-labeled dots. Circles denote the nuclei of two adjacent cells (differentiated based on co-staining with Phalloidin, not shown). The number of TS per nuclear volume and the summed occupancies of all TS divided by nuclear volume were not significantly correlated between G6pc and Pck1 ($R=0.29$, $p=0.17$ and $R=0.25$, $p=0.24$ respectively). Experiments were done on fasting mice.



C

Gene+zone	M(#Pol2/TS)	S.E.	μ(mRNA/hr)	S.E.	δ(1/hr)	S.E.	f	S.E.	MSE 1-state Pol2 occupancy	MSE 2-state Pol2 occupancy
Pck1 Villus	4.7	0.39	92.2	7.8	0.113	0.0020	0.031	0.0051	0.0674	0.0017
G6pc Villus	8.5	0.86	102.0	10.4	0.772	0.0130	0.027	0.0035	0.096	0.0095
Insr Villus	3.8	1.54	3.6	1.5	0.005	0.0003	0.003	0.0006	0.0396	0.0182
Glul Crypt	4.6	0.58	57.5	7.3	0.062	0.0020	0.007	0.0013	0.0836	0.005
Acly Crypt	3.7	0.29	8.9	0.7	0.037	0.0010	0.049	0.0175	0.0329	0.0027
Actb Crypt	1.4	0.09	47.3	3.1	0.262	0.0050	0.216	0.038	0.0011	0.0058

Figure S5 (Related to Figure 3) Bursty gene expression in the intestinal epithelium. (**A**) Most genes in intestinal epithelial cells exhibit rare transcription sites (TS) with high Pol2 occupancies that cannot be explained by a non-bursty transcription model. Red dots – mRNA for Pck1 detected with an exon library, blue – DAPI nuclei. Boxed region is magnified on the right, showing single TS with occupancy of 7 Pol2 molecules (yellow arrowhead) at the exon (top), intron (middle) and DAPI (bottom) channels. Scale bar is 5um. Arrow points at a nucleus of a goblet cell. (**B**) All genes except Actb fit a 2-state bursty model (green) rather than a 1-state non-bursty model (red). For Actb Pol2 occupancies better fit a 1-state model. (**C**) Burst parameters for the genes studied in the intestinal epithelium.

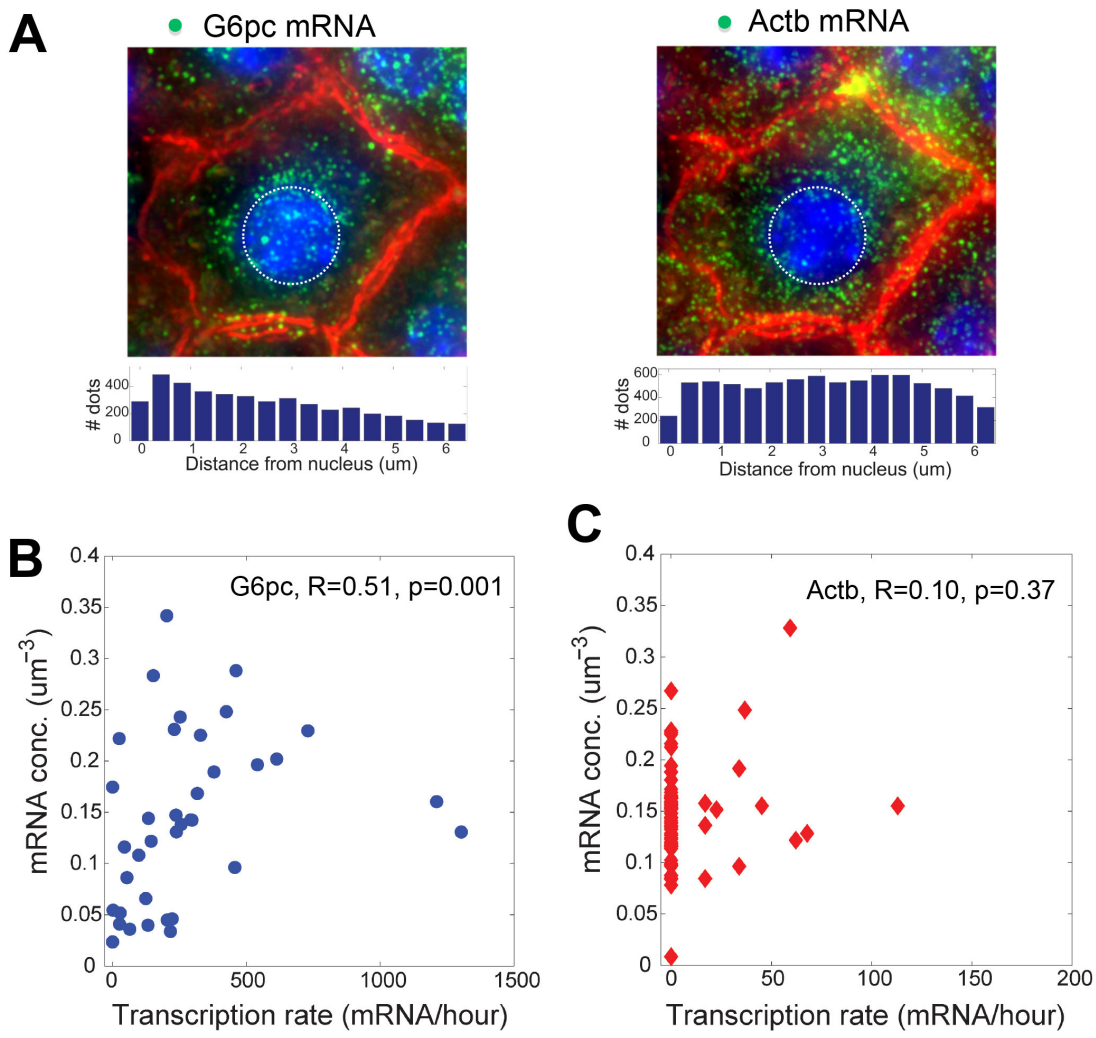


Figure S6 (Related to Figure 5) Gluconeogenic genes exhibit high degradation rates generating non-random intra-cellular mRNA localization and correlations with transcription rates. (**A**) Liver genes with short-lived mRNA are clustered around the nuclear periphery. Green dots are G6pc transcripts (left) or Actb transcripts (right). Bottom plots show the histograms of distances of cytoplasmic mRNA to the nuclear periphery. (**B-C**) Highly unstable genes exhibit positive correlations between cellular transcription rate and mRNA concentrations. (**B**) G6pc in pericentral zone in fasting state, $R_{Spearman}$=0.51, p=0.001 (δ=1.24 hr^{-1}), (**C**) Actb in periportal zone in fed state, $R_{Spearman}$=0.10, p=0.37 (δ=0.02 hr^{-1}). Every dot in panels **B-C** is a cell, x-axis is the summed transcription rate from all active TS divided by the cellular ploidy, y-axis is the cellular concentration of mRNA.

Table S1 Burst parameters for the different genes and conditions studied in the liver. *M* - Polymerase occupancy. μ - Transcription rate. δ - Transcript degradation rate. f - burst fraction. k_{ON}, k_{OFF} - Rates of promoter transitions between ON and OFF states. MSE – mean square errors between the experimental distributions and the model fits (averages of 100 cross-validation runs). MSE are shown for both the distributions of cytoplasmic mRNA per cell and the distributions of polymerase occupancies per transcription site. PP – Periportal, PC – Pericentral.

Table S2 Sequences of the probe libraries.

Extended Experimental Procedures

1. Mice and imaging

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 6AM (high-fed state), 9AM (fed state) and 12 PM (fast state, for these mice food was removed at 8AM). In the re-feeding experiment (Figure 5) mice were housed under reverse phase cycle, and fasted for 5 hours starting at 7AM. Mice were then re-fed ad libitum for an hour and sacrificed either before refeeding (2 mice) or after refeeding (2 mice). Validation of ploidy classification using Xist was performed on female mice of age 2 months. All mice were anesthetized with an intraperitoneal injection of a mixture containing 100 mg/kg ketamine and 10 mg/kg xylazine. Liver 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 cryosections were used for hybridization. To examine bursty gene expression in the intestinal epithelium mouse duodenum was harvested from 5-month old male mice fasted for 5 hours prior to sacrifice. 12 µm cryosections were used for hybridizations. Six out of the nine genes studied in the liver were also expressed in the intestine – Insr, G6pc, Glul, Pck1, Acly and Actb and were quantified either in crypts or villi, depending on the zone in which they were more highly expressed. For the gene Actb expression was too high in the villi to resolve individual mRNA molecules, and therefore quantification was done on its expression in crypts.

Hybridization and imaging

Probe libraries were designed and constructed as previously described (Itzkovitz et al., 2012). Single molecule FISH probe libraries consisted of 44-96 probes (Table S2) of length 20 bps, and were coupled to cy5, alexa594 or tmr. Hybridizations were performed overnight in 30°C. Typical hybridization mixtures included exon library in cy5, intron library in tmr and Actb library in alexa594. DAPI dye for nuclear staining was added during the washes. To detect cell borders alexa fluor 488 conjugated

phalloidin (Rhenium A12379) was added to the GLOX buffer wash. Images were taken with a Nikon Ti-E inverted fluorescence microscope equipped with a ×100 oil-immersion objective and a Photometrics Pixis 1024 CCD camera using MetaMorph software (Molecular Devices, Downington, PA). The image-plane pixel dimension was 0.13 μm. Quantification of transcription sites and nuclear volume were done on stacks of 45 optical sections in the liver and 30 optical sections in the intestine, with *Z* spacing of 0.3 μm, whereas quantification of the mRNA concentrations was done on the first 10 optical sections. Portal node was identified morphologically on DAPI images based on bile ductule, central vein was identified using smFISH for Glutamine Synthetase performed on serial sections. Hepatocytes within the first three layers of the portal node (up to ~50 um distance) were classified as periportal and hepatocytes within the first four layers of the central vein (~60um distance), but excluding the innermost layer were classified as pericentral. This exclusion stems from distinct expression levels which we often observed in the innermost layer directly bordering the central vein. All results presented in the paper are for tetraploid (4n) hepatocytes, with the exception of the polyploidy analysis, which compared diploid and tetraploid cells (Figure 6).

Cell culture and Actinomycin-D treatment

NIH 3T3 cells were grown in Dulbecco's modified Eagle medium (DMEM) supplemented with 10% fetal calf serum (FCS), penicillin (200 I.U./ml) and streptomycin (100 μg/ml) at 37°C with 5% CO2 and split every three days. A day prior to Actinomycin-D treatment 200,000 NIH 3T3 Cells/well were seeded in a 24 well plate. Actinomycin-D (Sigma, A9415-2MG) was resuspended in DMSO with a final concentration of 2 mg/mL. Actinomycin-D was thoroughly mixed with the growth medium (3 μg/ml) and added on the cells. Cells were harvested at 0, 1.5, 3, 4.5 and 6 hours and immediately fixed in 4% paraformaldehyde for 15 minutes. Cells were then washed once with PBS and incubated in ethanol for at least two hours at 4°C. The number of nascent and mature mRNA was determined by smFISH.

Western blot analysis

Liver tissues (~150 mg) were homogenized in ice-cold RIPA buffer (R0278, Sigma) supplemented with protease inhibitors (P8340, Sigma). Protein samples (60 µg) were separated on 12% reducing polyacrylamide gels and electroblotted. Immunoblots were blocked and incubated overnight at 4°C with G6PC antibody (sc-25840, Santa Cruz) or Monoclonal Anti-α-Tubulin (T9026, Sigma) as a loading control. Blots were visualized by chemiluminescence using an ECL kit (Thermo). Quantifications were performed using the ImageQuant Las 4000 mini luminescence image analyzer (GE Healthcare) and Fiji software.

2. Ploidy classification algorithm

In this section we describe our algorithm for hepatocyte ploidy classification. We developed an image-processing algorithm implemented in MATLAB to infer ploidy from nuclear dimensions and validated it with smFISH for Xist, expressed from inactive X chromosomes in female.

Cell borders were manually marked for all hepatocytes in an image field based on FITC-Phalloidin membrane staining and DAPI nuclear staining. For bi-nucleated hepatocytes a line was manually drawn between the two nuclei to assist in automatic nuclear detection. Otsu's method was next used to threshold the three dimensional image stack of each segmented cell. For each optical section within the 3D image stack a spherical structuring element, the size of the thresholded object was convolved with the binary image and the area of the resulting binary object was calculated. The resulting 3D profile of the DAPI cross-sectional areas at each $Z, A(Z)$ was used to extrapolate the maximal cross-sectional area. This facilitated inference of nuclear size even if the 3D stack did not include the entire nucleus. Assuming a nucleus of radius R, the cross-sectional area *A(Z)* at stack *Z* is:

[1] $A(Z) = \pi r(Z)^2 = \pi(R^2 - (R - Z)^2) = -\pi Z^2 + 2\pi RZ$

Where $r(Z)$ is the radius of the cross-sectional circle at stack Z and R is the radius of the nuclear sphere.

We fit a parabola to each DAPI cross section profile and extrapolated the maximum cross-area. We manually selected thresholds that separated the resulting multimodal distribution (Figure S1D). To validate our ploidy classification method we next performed an smFISH experiment on frozen sections of female mouse liver with

an Xist probe. The number of nuclear dots correlated with the nuclear ploidy levels, with diploid nuclei exhibiting 1 dot, tetraploid nuclei exhibiting 2 dots etc. (Figure S1E). The ploidy classification error was <10%.

3. Dot intensity inference

Dots were automatically detected using previously described algorithms (Itzkovitz et al., 2012; Lyubimova et al., 2013). Image stacks were filtered with a 3D Laplacian of Gaussian filter with a standard deviation of 1.5 pixels. Next, a range of thresholds was tested, for each threshold the filtered image was converted to a binary image and the number of connected components recorded. We automatically chose the threshold for which the number of connected component was least sensitive to threshold selection. Each threshold was manually validated and corrected when necessary. To extract dot intensity we examined for each dot a subimage centered on the detected dot centroid, of size 20*20*Z voxels, where Z is the range of z-stacks spanned by the dot. For this subimage Otsu's method was applied sequentially in two thresholding steps. In the first step Otsu's method was used to infer a threshold, T_1 and only voxel values above the threshold were used for the second step. In the second step Otsu's method was used again, this time on the refined voxel values, to extract a threshold T_2. The background of the sub-image, estimated as the median of all original pixel values below T_2, was subtracted from each voxel value (negative pixel values were set to 0). Dot intensity was computed as the integrated intensity of all background-subtracted voxels in the plane in which the intensity was maximal. In situations were the maximal cross-section of a dot was smaller than 4 pixels an expanded 3*3 pixel cross section around the centroid pixel in the maximal intensity stack was used to estimate dot intensity.

4. Algorithm for inferring f, μ and δ

Calculating the fraction of active transcription sites

To infer f, the fraction of active transcription sites (TS), one must unambiguously identify the TS from nuclear mRNA and compare those to the expected number of nuclear sites based on the ploidy classification. To this end we searched for each nuclear intronic dot a nuclear exonic dot that was within 1 um of it and further

removed double-labeled dots with a background dot (a dots appearing in the third fluorescence channel) within 0.5 um of it. Such rare dots were cytoplasmic elements that were directly adjacent to the nucleus. To avoid under-counting the number of TS in a nucleus that only partially appeared in our image stack we only included cells for which both nuclei were entirely included in the image stack in the calculation of f (Figure S1F). For each such cell f was computed as the ratio of observed TS to that expected based on the cell ploidy classification. We performed manual validation of all detected TS.

Calculation of Pol2 occupancy

To extract μ, the rate of mRNA produced from active TS we first calculated for each TS its Pol2 occupancy, M, based on the ratio of the TS intensity and the intensities of non-TS dots:

[2] $M = \lceil I_E/(\eta \cdot \kappa \cdot median(I_{E-nonTS})) \rceil$

Where $\lceil x \rceil$ is the ceiling operator, the lowest integer number larger than x, I_E is the intensity of the TS dot appearing in the exonic channel and the median is taken over non-TS that are in the same optical section as the maximal-intensity section of the TS. η and κ are correction factors described below.

Correction for probe library spread along the gene - η

Unlike mature mRNA for which the entire probe library is bound, Pol2 molecules at distinct positions along the gene are attached to nascent mRNA tails on which a lower number of probes are bound. If the Pol2 molecule is at the beginning of the gene a small number of probes would be bound on the nascent mRNA tail, whereas if it is at the 3' end of the gene most probes would be bound. To account for this effect we introduced a correction factor η representing the average intensity of TS in units of cytoplasmic dot intensity. This incorporates the location of the probes along the target gene:

[3] $\eta = \left(\frac{1}{L \cdot N(L)}\right) \sum_{i=1}^{L} N(i)$

Where *L* is the length of the gene and $N(i)$ is the number of probes that are bound on a nascent mRNA which is attached to a Pol2 molecule that has reached position *i*

on the gene. Equation [3] assumes that a polymerase molecule has an equal probability to reside at each nucleotide position and neglects pausing of the nascent transcripts at the 3' end of the gene. Below we describe experimental support for using this approximation. Figure S3F shows the probe spread correction factors for the genes studied.

Correction for inferred occupancies - κ

Calculating polymerase occupancy as the ratio between TS intensity and the intensity of single mRNA dots yields estimates that are larger than the real occupancies, even when correcting for probe locations. This is a result of the ceiling operator in Equation [2]. To account and correct for this we computed the occupancies of individual mRNA molecules (non-TS) using $\eta = 1$ to obtain a kernel distribution *K(M)*, the mean of which was $\kappa = 1.5$ rather than 1. We subsequently divided by this factor to obtain the estimated polymerase occupancies of all TS.

Calculation of μ and δ

Polymerase occupancy M was next used to compute μ, the average rate of mRNA produced from active TS:

[4] $\mu = M \cdot v/L$

Where v is polymerase speed (calibrated below) and *L* is the length of the gene. Finally, we used f, μ and $\langle X \rangle$, the average number of mRNA copies per cell to compute transcript degradation rates as:

[5] $\delta = n \cdot f \cdot \mu / \langle X \rangle$

Where $\langle X \rangle$ is the average number of cytoplasmic mRNA molecules per hepatocyte and n is the ploidy level (typically 4 in our experiments). To compute cytoplasmic mRNA levels we counted the number of cytoplasmic dots for each segmented cell within the first 3um of the Z-stack. We next divided this number by the segmented cytoplasmic volume to obtain the cytoplasmic concentration, and multiplied by the cytoplasmic volume of the respective ploidy class (Martin et al., 2002).

Calculation of parameters for the intestinal epithelium

In the intestinal epithelium single cell segmentation is challenging as cells are too dense to avoid overlap. We therefore applied a slightly modified method that still enables extracting burst fraction, transcription rates and degradation rates *in-situ*, but not the precise rates k_{ON}, k_{OFF} (the extraction of which requires precise single-cell distributions).

To compute burst fraction we detected all TS in each image and divided by the number of chromosomal copies, in that image, *C,* calculated as:

[6] $C = N \cdot n \cdot \theta$

Where N is the number of nuclei in a single Z section, n is the ploidy and θ is the the average number of nuclei in the entire Z-stack (~1.5 for the intestinal images used). The cell ploidy n was set to $n = 2$ for villus cells which are postmitotic (for the genes Pck1, G6pc and Insr) and $n = 2.5$ for genes that were expressed in the intestinal crypts (Acly, Actb and Glul). This factor takes into account the fact that crypt cells are dividing and therefore contain more than two copies on average. In the mouse small intestinal crypt the kinetic parameters for total cell cycle time, S-phase period, G1 period and G2+M periods are $T_C = 18\ hr$, $T_{G1} = 8.5\ hr$, $T_S = 7.5\ hr, T_{G2+M} = 2\ hr$ respectively (Quastler and Sherman, 1959). Using these parameters the average ploidy for crypt cells is:

[7] $n = 2 \cdot \frac{T_{G1}}{T_C} + 3 \cdot \frac{T_S}{T_C} + 4 \cdot \frac{T_{G2+M}}{T_C} \sim 2.5$

To extract δ we segmented 'meta-cells' - epithelial regions containing multiple cells, and divided the summed transcription rates of all TS in each meta-cell by the total number of mRNA dots contained in it. μ was computed as described above for liver tissue.

5. Calibrating polymerase speed and validating degradation rate estimates

In this section we describe our calibration of v, the polymerase speed, and two experiments to validate our estimated degradation rates – an *ex-vivo* validation on NIH3T3 cells and an *in-vivo* validation on liver from fasted and refed mice.

Calibrating polymerase speed

The polymerase speed, v, is used in equation [4] to translate polymerase occupancy M to transcription rate μ. Estimates for average Pol2 speed in mammals vary widely, ranging from 6 bp/sec (Darzacq et al., 2007) to 60 bp/sec (Singh and Padgett, 2009). Rather than using this range we independently measured speed using our smFISH measurements of bursting dynamics combined with Actinomycin-D treatment on NIH 3T3 cells. We first measured the fraction of active TS, f, and promoter occupancy, M using equation [2]. We then treated the cells with 3 µg/ml Actinomycin-D and subsequently fixed the cells at sequential time points. We used smFISH to count mRNA molecules at each time point (Figure S2) and fit an exponential decay to obtain the degradation rate δ. Using equations [4] and [5] we obtain:

[6] $\delta = \langle S \cdot v/(L \cdot X)\rangle \Rightarrow v = L \cdot \delta \cdot \langle X/S\rangle$

where $S = \sum M$ is the total number of Pol2 residing on all TS in the cell, X is the number of cytoplasmic mRNA molecules and the average $\langle\ \ \rangle$ is taken over all cells. We chose to perform this calibration on P21, a gene with short life-time, since Actinomycin-D experiments are considered more representative for short-lived transcripts (Schwanhäusser et al., 2011). Using this calibration we obtained a value of $v = 34 \pm 11 bp/s$ for polymerase speed.

Ex-vivo validation of estimated degradation rates

To validate our estimated degradation rates we next used the inferred speed v to estimate the degradation rate of Acly, another relatively short life-time gene, using our *in-situ* approach (Equations [2-5]). Using $v = 34 \pm 11 bp/s$ our *in-situ* estimates yielded $\delta_{in-situ} = 0.14 \pm 0.04 hr^{-1}$. We next measured the decline in Acly mRNA levels at different time-points following cessation of transcription with Actinomycin-D to obtain $\delta_{ActD} = 0.12 \pm 0.02 hr^{-1}$. Thus our *in-situ* estimates are within 15% error of the degradation rates extracted from Acly mRNA degradation following Actinomycin-D treatment.

In-vivo validation of estimated degradation rates and protein life-time measurements

In the refeeding experiment (Figure 5) we sought to compare expression in two metabolic states where a sudden drop in transcription can be observed. Measuring the decline in mRNA levels under such a scenario would effectively constitute the *in-vivo* equivalent to an Actinomycin-D experiment. To this end we used our *in-situ* method to measure transcription rates and mRNA concentrations, as well as western blots to measure protein concentrations of G6pc in 5-hour fasted mice before and after 1 hour of free refeeding (2 mice for each time point).

For simplicity we assume that both mRNA and protein degradation rates (δ_m, δ_p) do not change during the relatively short 1-hour refeeding period, that mice commenced refeeding immediately at the beginning of the hour and that transcription rates dropped to the refed levels immediately. As explained below deviations from these assumptions yield error estimates that are even smaller than those observed. Denoting transcription rates, translation rates, mRNA concentrations and protein concentrations before and after refeeding by $\beta_{m0}, \beta_{p0}, x_0, y_0$ and $\beta_{m1}, \beta_{p1}, x_1, y_1$ respectively, we use the following equations to describe the mRNA dynamics after commencement of refeeding:

[7] $dx/dt = \beta_{m1} - \delta_m \cdot x$

Assuming expression is at steady state before refeeding with $x_0 = \beta_0/\delta_m$ we obtain:

[8] $x_1/x_0 = (\beta_{m1}/\beta_{m0}) \cdot (1 - e^{-\delta_m t}) + e^{-\delta_m t}$

Using our measurements of μ *and* f before and after refeeding we can readily calculate the transcription rate as $\beta_m = 4 \cdot f \cdot \mu$. Using Equation [8] with our measured values of $\beta_{m1}/\beta_{m0} = 0.054 \pm 0.03$ and $x_1/x_0 = 0.32 \pm 0.017$ (Figure 5A, B) yields $\delta_m = 1.27 \pm 0.15$. Our *in-situ* method estimate of mRNA degradation rates in the fasted state (before commencement of refeeding) was $\delta_m = 2.08 \pm 0.21$. Thus the degradation rate computed from the drop in mRNA levels over an hour is 40% lower than our *in-situ* estimates. If refed mice do not commence refeeding immediately or if transcription rates drop to their refed state more gradually, the actual degradation rates based on Eqaution [8] are expected to increase, reducing this deviation from our *in-situ* estimates even further.

To obtain bounds on protein degradation rates under these metabolic conditions we consider two scenarios – in the first one translation stops completely upon refeeding, and in the second translation proceeds at the same rate as that during fasting. The equations describing the dynamics of protein concentrations are:

[9] $dy/dt = \beta_{p1}x - \delta_p \cdot y$

Where we consider either $\beta_{p1} = 0$ for the first scenario or $\beta_{p1} = \beta_{p0}$ for the second scenario. Solving equation [9] using $\beta_{p1} = 0$ yields:

[10] $y_1/y_0 = e^{-\delta_p t}$

And for $\beta_{p1} = \beta_{p0}$ (where for simplicity we set $\beta_{m1}/\beta_{m0} = 0$):

[11] $y_1/y_0 = \frac{\delta_p}{\delta_p - \delta_m} e^{-\delta_m t} - \frac{\delta_m}{\delta_p - \delta_m} e^{-\delta_p t}$

We numerically solved Equations [10-11] using the δ_m extracted from Equation [8] and our measurements of $y_1/y_0 = 0.43$ (Figure 5B, C), yielding $0.85 < \delta_p < 4.53$. This translates to a short protein life-time of between 10 and 45 min, considerably less than the typical life-times of 50 hours estimated in whole-genome measurements (Schwanhäusser et al., 2011).

6. Estimating the impact of termination pausing

The estimated Pol2 occupancy *M* using Equation [2] assumes that all mRNA are physically attached to Pol2 that are actively transcribing and are uniformly proceeding along the gene. This estimate neglects potential pausing at the 3' UTR for processing. If nascent mRNAs spend a significant time at the 3' UTR before being extruded from the TS our calculations of transcription rates would be overestimating the true ones.

To assess the extent of termination pausing we performed dual-color experiments with one library consisting of 15 probes targeting the first 2600 bps of Pck1 coupled to alexa594 (L1, Figure S3A) and a second library consisting of 15 probes targeting the last 1000 bps of Pck1 coupled to cy5 (L2, Figure S3B). As a control we used a third library of 15 probes targeting the beginning of the gene and coupled to cy5 that was interleaved with the alexa594 coupled L1 library (L3, Figure S3A). If the time of termination pausing is significantly longer than the time taken for a polymerase molecule to transcribe the gene, most mRNA in the TS will be at the 3'

UTR and should have the entire set of 15 library L1 probes and 15 library L2 probes and hence equal intensities, compared to the experiment with library L1 and library L3. If however most nascent mRNAs at the TS are attached to actively transcribing polymerase molecules only polymerase molecules that have reached the last 1000 bps would have the L2 probes attached to them. A model in which a fraction *F* of the mRNA at the TS are attached to actively transcribing Pol2 molecules and a fraction (*1-F)* are pausing at the 3' end yields the following equations for the TS intensities:

[12] $I_{Li-TS} = \langle M \rangle [\eta_{Li} \cdot F + 1 \cdot (1-F)] \cdot I_{Li-cyto}$

Where $I_{Li-cyto}$ are the average intensities of cytoplasmic mRNA dots labeled with libraries *Li* (*i=1,2,3*), and $\langle M \rangle$ is the average Pol2 occupancy and η_{Li} are the probe spread correction factors (Equation [3]) for each library. Using the probe spread correction factor values in Figure S3F we obtain:

[13] $R_{L1-L3} = \frac{\langle I_{L1-TS} \rangle}{\langle I_{L3-TS} \rangle} = \frac{[0.76 \cdot F + 1 \cdot (1-F)] \cdot I_{L1-cyto}}{[0.14 \cdot F + 1 \cdot (1-F)] \cdot I_{L3-cyto}}$

[14] $R_{L1-L2} = \frac{\langle I_{L1-TS} \rangle}{\langle I_{L2-TS} \rangle} = \frac{[0.76 \cdot F + 1 \cdot (1-F)] \cdot I_{L1-cyto}}{[0.76 \cdot F + 1 \cdot (1-F)] \cdot I_{L2-cyto}}$

By dividing [13] and [14] and using the fact that $I_{L2-cyto} = I_{L3-cyto}$ (both are cy5 labeled libraries consisting of 15 probes which we validated give rise to dots with indistinguishable intensities) we obtain:

[15] $\frac{R_{L1-L3}}{R_{L1-L2}} = \frac{[0.76 \cdot F + 1 \cdot (1-F)]}{[0.14 \cdot F + 1 \cdot (1-F)]}$

Our measurements of the ratio in equation [15] were 3.02 (Figure S3), yielding *F*=0.86. Our experiments thus indicate that >85% of mRNA at TS are attached to actively transcribing polymerase molecules and thus pausing at the 3' UTR can be neglected without substantially influencing our estimates of transcription rates.

7. Fitting mRNA single-cell distributions with promoter models

In this section we describe our approach for fitting 1-state and 2-state promoter models to our single-cell measurements. We fit both the distributions of mRNA per cell and the distributions of polymerase occupancies. When fitting promoter occupancies we assumed the simplest model of polymerase binding and proceeding at constant rates, neglecting polymerase pausing, as well as transient periods at the start and end of bursts, in which the loci are gradually filled or depleted with

polymerase molecules. We used the MATLAB function nlinfit to obtain the best fits and the function nlparci to obtain confidence intervals.

Loci independence

Our measurements indicate that the number of active TS per cell is in general binomially distributed (Figure S4). Thus different loci in the same cell can be considered independent. For a hepatocyte of ploidy *n*, the number of mRNA molecules per cell is the sum of *n* independent variables, each obeying a distribution $P_1(X)$ or $P_2(X)$, the expected distribution of mRNA molecules per cell from a single locus for a 1-state or 2-state model respectively. The resulting distributions are the convolutions of $P_1(X)$ or $P_2(X)$ with themselves *n* times. For tetraploid hepatocytes this yields the distribution: $P_i^4(X) = P_i(X) * P_i(X) * P_i(X) * P_i(X)$ where i=1,2.

Correction for volume subsampling

Due to limitations of the imaging setup, and the large volume of hepatocytes we are only measuring a sub-volume (*1/V_f*) of the entire hepatocyte volume, where $V_f = 14 \pm 3$. This limitation is due to the scattering of single mRNAs that are at a distance of more than 20um from the coverslip, prohibiting counting mRNA in the entire cellular volume *in-situ*. In addition we only segment areas around the nuclei, to ensure exclusion of mRNA appearing in neighboring cells, a phenomenon that is negligible in the central region of the cells.

To determine the distributions of mRNA per hepatocyte we computed the mRNA concentration (#mRNA/segmented cytoplasmic volume) and multiplied by the cytoplasmic volume of a hepatocyte (Martin et al., 2002), as described above. This calculation yields distributions that are wider than those that would be obtained by directly counting all mRNA in the entire hepatocyte volume. To understand this effect, consider a promoter obeying a 1-state model, generating a Poisson distribution of mRNA with an average of 1000 copies per cell. The standard deviation of the number of mRNA per cell is $\sqrt{1000} \sim 30$. However, if one counts only a 1/10 of the hepatocyte volume and corrects for the real number by multiplying the counts by 10, the counts in a 1/10 volume would be Poisson-distributed with an average of

100 ± 10 and the full distribution would be 1000 ± 100 instead of 1000 ± 30. When fitting both the 1-state and 2-state model we controlled for this subsampling broadening of the distributions as follows. We first generated the full distribution, *P(Y)*, then numerically sampled 10,000 times from it a number Y_i, and produced a modified count $Xi=V_f*(Poissrnd(Yi/V_f))$, where $V_f = 14 \pm 3$. The resulting numeric distribution *P(X)* was used to fit the experimental data.

Correction for polymerase occupancy distributions

In addition to measuring the distributions of mRNA per cell we also measured the distribution of polymerase molecules on an active TS. As described above this was computed as the ceiling of the ratio between the intensity of an exonic dot in the TS to the median intensity of all non-TS exonic dots (Equation [2]). This calculation yields occupancies that are larger than the real ones due to the ceiling operator. To translate this to occupancy distributions we computed the occupancies of individual mRNA molecules (non-TS) using $\eta = 1$ to obtain a kernel distribution *K(M)*, the mean of which was $\kappa = 1.5$. We convolved the theoretical distributions of Pol2 occupancies with this kernel.

Fitting procedure

For both models we compared the experimental data and the theoretical prediction for both the cumulative distribution of cellular cytoplasmic mRNA content and that of the Pol2 occupancies per TS. Since the 2-state model has one free parameter, whereas the 1-state model is parameter-less we applied a cross-validation scheme to avoid over-fitting. The data sets for both cellular cytoplasmic mRNA content and TS Pol2 occupancies were divided into a training set containing 70% of the data points and a test set containing the remaining 30%. Model parameters were fit on the training set and mean squared error was computed for the test set. The mean-squared errors reported in Table S1 and Figure S5C are averages of 100 such cross-validation runs.

1-state promoter model

The distribution of mRNA generated from a promoter obeying a 1-state model is a Poisson distribution, and since convolutions of Poisson distributions are also Poisson distributions:

[16] $P_1^4(Y) = e^{-\langle Y\rangle}\langle Y\rangle^Y/Y!$

Where $\langle Y\rangle$ is the average number of mRNA per cell. We used this distribution to generate the convolved and broadened distribution $P_1^4(X)$. Under a 1-state model the distribution of Pol2 occupancies is also Poisson:

[17] $P(M) = e^{-\langle M\rangle}\langle M\rangle^M/M!$

$\langle M\rangle = -\log(1-f)$, where f is the fraction of double-labeled nuclear dots (TS). Under this model all promoters are active but due to low polymerase initiation rates not all loci have active Pol2 molecules occupying them.

2-state promoter model

We used the 2-state model developed by Raj and colleagues (Raj et al., 2006). The model, described in Figure 2A, considers a promoter that transitions from an OFF state to an ON state with rate k_{ON} and from an ON state to an OFF state with rate k_{OFF}, producing transcripts at rate μ when in the ON state. Transcripts are degraded at a rate δ.

[18] $P_2(Y) = \frac{\Gamma\left(\frac{k_{ON}}{\delta}+Y\right)}{\Gamma(Y+1)\Gamma\left(\frac{k_{ON}}{\delta}+\frac{k_{OFF}}{\delta}+Y\right)}\frac{\Gamma\left(\frac{k_{ON}}{\delta}+\frac{k_{OFF}}{\delta}\right)}{\Gamma\left(\frac{k_{ON}}{\delta}\right)}\left(\frac{\mu}{\delta}\right)^Y\ {}_1F_1\left(\frac{k_{ON}}{\delta}+Y, \frac{k_{ON}}{\delta}+\frac{k_{OFF}}{\delta}+Y, -\frac{\mu}{\delta}\right)$

Where ${}_1F_1$ is a confluent hypergeometric function of the first kind. We used this distribution to generate the convolved and broadened distribution $P_2^4(X)$. The distribution of polymerase occupancies was modeled as a Poisson distribution with mean $\langle M\rangle$ determined from our calculations (equation [2]) and convolved with the Kernel *K(M)*. Figure 3 and Figure S5B demonstrate that a 2-state model fits both the distributions of cellular mRNA numbers as well as the distributions of polymerase occupancies. Table S1 provides the estimates of the promoter bursting rates k_{ON}, k_{OFF} and confidence intervals.

8. Effect of polyploidy on reduction of intrinsic variability

Our measurements indicate that liver promoter bursting is independent. Thus the probability that a promoter will switch to an ON state is not dependent on the state

of other promoter copies in the same cell (Figure S4). Denoting the steady state mRNA level that would result from a single promoter by X_1 (with average $E(X_1)$) the random variable describing the number of mRNA in a diploid/tetraploid hepatocyte is:

[19] $X_n = \sum_1^n X_1$

Where n=2 or 4 respectively, with averages $E(X_n) = nE(X_1)$ and standard deviation $\sigma(X_n) = \sqrt{n}\sigma(X_1)$. Since the volumes of hepatocytes scale linearly with ploidy (Martin et al., 2002) the statistics for the mRNA concentrations $C_n = X_n/(nV)$ where V is the typical hepatocyte cell volume divided by ploidy, are:

[20] $E(C_n) = E(X_1)/V$

[21] $\sigma(C_n) = \sigma(X_1)/\sqrt{n}V$

[22] $C.V.(C_n) = \sigma(C_n)/E(C_n) = (1/\sqrt{n})\sigma(X_1)/E(X_1)$

Thus the coefficient of variance of mRNA concentrations reduces as the square root of the ploidy level. This model predicts a decrease of 30% ($\sqrt{2}$) in intrinsic variability in tetraploid hepatocytes (*n=4*) compared to diploid hepatocytes (*n=2).* In our measurements we obtained a highly significant decline of 13% on average for all genes and conditions compared (Experimental Procedures).

Supplemental References

Darzacq, X., Shav-Tal, Y., de Turris, 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.

Itzkovitz, S., Lyubimova, A., Blat, I.C., Maynard, M., van Es, J., Lees, J., Jacks, T., Clevers, H., and van Oudenaarden, A. (2012). Single-molecule transcript counting of stem-cell markers in the mouse intestine. Nat. Cell Biol. *14*, 106–114.

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.

Quastler, H., and Sherman, F.G. (1959). Cell population kinetics in the intestinal epithelium of the mouse. Exp. Cell Res. *17*, 420–438.

Raj, A., Peskin, C.S., Tranchina, D., Vargas, D.Y., and Tyagi, S. (2006). Stochastic mRNA synthesis in mammalian cells. PLoS Biol. *4*, e309.

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

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