bioRxiv preprint doi: https://doi.org/10.1101/2022.09.21.508820; this version posted September 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

RNA degradation heavily impacts mRNA co-expression

Óscar García Blay[1,2], Pieter Verhagen[1,2], Benjamin Martin[1,3], Maike M.K. Hansen[1,4*]

[1] Institute for Molecules and Materials, Radboud University, Heyendaalseweg 135, 6525 AJ Nijmegen, the Netherlands

[2] These authors contributed equally

[3] Present address: Institute of Bioengineering, School of Life Sciences, Ecole Polytechnique Fédérale de Lausanne (EPFL), Lausanne 1015, Switzerland

[4] Lead contact

*Correspondence: maike.hansen@ru.nl

KEYWORDS

Co-expression, transcriptome, promoter activity, RNA degradation, single-cell.

bioRxiv preprint doi: https://doi.org/10.1101/2022.09.21.508820; this version posted September 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

SUMMARY

Co-expression of genes measured with single-cell RNA sequencing is extensively utilized to understand the principles of gene regulation within and across cell types and species. It is assumed that the presence of correlation in gene expression values at the single-cell level demonstrates the existence of common regulatory mechanisms. However, the regulatory mechanisms that should lead to observed co-expression at an mRNA level often remain unexplored. Here we investigate the relationship between processes upstream and downstream of transcription (i.e., promoter architecture and coordination, DNA contact frequencies and mRNA degradation) and pairwise gene expression correlations at an mRNA level. We identify that differences in mRNA degradation (i.e., half-life) is a pivotal source of single-cell correlations in mRNA levels independently of the presence of common regulatory mechanisms. These findings reinforce the necessity of including post-transcriptional regulation mechanisms in the analysis of gene expression in mammalian cells.

bioRxiv preprint doi: https://doi.org/10.1101/2022.09.21.508820; this version posted September 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

INTRODUCTION

The emergence of single cell analyses has unlocked an exciting new chapter for understanding gene regulation in cells. Specifically, single-cell RNA sequencing (scRNA-seq) is a powerful and widely implemented tool that allows for transcriptome-wide quantification of mRNA in thousands of individual cells (Hwang et al., 2018). One feature that can be extracted from scRNA-seq datasets is gene-to-gene co-expression (Eisen et al., 1998), which has been extensively used over the last years in a range of applications to gain quantitative insights into gene regulation of eukaryotic cells. For example, co-expression of genes has been described as a robust tool to identify cell types or states from scRNA-seq datasets (Crow and Gillis, 2018). Specifically, co-expression has been implemented in order to study cellular heterogeneity of tissues and organs in physiological (Aizarani et al., 2019; Andrews et al., 2022; Muraro et al., 2016; Payen et al., 2021; Travaglini et al., 2020) and pathological contexts (Esmaili et al., 2021; Han et al., 2021), as well as dynamic processes such as cell signaling or changes in cell identity during development (Foreman and Wollman, 2020; Qadir et al., 2020; Salehi et al., 2021). On the other hand, gene co-expression in scRNA-seq datasets has been exploited to extrapolate the functional principles of gene expression regulation (Desai et al., 2021) from cellular populations in the form of gene-regulatory networks (Matsumoto et al., 2017). The most common approach is the construction and analysis of co-expression networks from pairwise correlation measurements (Vivian Li and Li, 2021; Wang et al., 2021). Single-cell co-expression network analysis has thus led to the characterization of novel regulatory pathways (Xie et al., 2021), which in turn has improved our understanding of common gene regulation principles between cell types and species (Crow et al., 2022; Harris et al., 2021). Given the expansive application of single-cell co-expression analysis in both current and likely future studies, it is essential to understand the causality and limitations of observed correlations in scRNA-seq datasets.

In order to exploit co-expression networks to study gene regulation, a common assumption is that gene-to-gene correlation or anticorrelation indicates an underlying functional relationship (Eisen et al., 1998; Oliver, 2000). Traditionally, it has been assumed that such co-expression patterns arise from common regulatory mechanisms that are shared between respective genes. Since gene expression is a

bioRxiv preprint doi: https://doi.org/10.1101/2022.09.21.508820; this version posted September 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

multi-step process (e.g., promoter toggling, transcription, RNA processing) (Ronen et al., 2002), there are different scenarios from which functional (anti)correlation between genes may emerge (Munsky et al., 2012). For instance, gene-to-gene (anti)correlation could either be caused by shared transcription factors commonly affecting the transcription of a group of genes, or due to post-transcriptional processes occurring downstream of transcription (Figure S1A). However, in higher eukaryotes, the balance between transcriptional and post-transcriptional events underlying mRNA co-expression remain elusive. Early bulk co-expression experiments have shown that promoter sharing can be a major source of gene co-expression (Gu et al., 2011). Yet, more recent single cell studies have shown that shared target-regulator relationships are unlikely to result in co-expression and that most—but definitely not all—shared transcription factors fail to enforce co-expressed behavior among target genes (Ribeiro et al., 2021; Yin et al., 2021). Solving the interplay between regulatory architectures and effective co-expression is key for novel strategies that apply combined approaches to study transcriptional regulation (Jeong et al., 2021).

Here, we sought to identify the parameters that contribute to correlation and anticorrelation in scRNA-seq data without enforcing previous regulatory structures. To this end, we integrated scRNA-seq data from mouse embryonic stem cells (mESCs) with other existing datasets from mESCs, including gene-to-gene contact frequencies in Hi-C data (Nora et al., 2017), promoter activity in intron seq-FISH data (Shah et al., 2018), and transcript-specific half-life measurements (Sharova et al., 2009). The analyses show that coordination in promoter activity or gene-to-gene contacts do not appear to be the major sources of gene-to-gene correlation in RNA expression levels. Remarkably, mRNA degradation emerges as clear contributor to the (anti)correlations (i.e., co-expression) measured with scRNA-seq.

RESULTS

Promoter coordination and gene-to-gene contacts do not explain mRNA co-expression

In order to identify possible causes for the (anti)correlation in mRNA levels we integrated several datasets. To this end, we combined single-cell mRNA abundance data (STAR METHODS) with 3

bioRxiv preprint doi: https://doi.org/10.1101/2022.09.21.508820; this version posted September 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

additional published datasets providing promoter activity information (Shah et al., 2018), gene-to-gene contact frequencies (Nora et al., 2017), and mRNA half-lives (Sharova et al., 2009) (Figure 1A and 1B), all 4 datasets from mESCs. In order to accurately compare all data modalities, we limited the number of genes analyzed to those included in all datasets, totaling 5277 genes. To extract (anti)correlation regimes we first performed gene-by-gene pairwise Pearson correlation analysis from scaled scRNA-seq read counts, followed by hierarchical clustering (Figure 1A). Pearson correlation is preferred over Spearman because highly expressed genes in a low number of cells display more accurate Pearson correlation values than Spearman correlation values (Figure S1B and S1C), as described in other studies (Vandenbon, 2022). To incorporate the promoter behavior of each gene into our analysis, we assumed two-state promoter toggling as an underlying mechanistic model for the analyzed promoters (Esmaili et al., 2021; Kepler and Elston, 2001; Raj et al., 2006; Weinberger et al., 2005). Eukaryotic promoters have been described to switch between at least two (Harper et al., 2011; Zenklusen et al., 2008) distinct states: one state that allows for mRNA production (referred to as the ON state) and another state that is not permissive for mRNA production (referred to as the OFF state). This toggling results in the discontinuous or bursty behavior that is associated with transcription in eukaryotes (Chong et al., 2014; Coulon et al., 2013; Golding et al., 2005; Suter et al., 2011). The effective ON and OFF states of each promoter can be defined by the presence or absence of intronic signal respectively (Bahar Halpern et al., 2015; Shah et al., 2018). With this information, we computed the likelihood that pairwise promoters are more coordinated (i.e., ON or OFF at the same time) or anti-coordinated (i.e., one promoter being ON while the other is OFF and *vice versa*) compared to what is expected from independent random promoter behavior (Figure 1A and 1B, green and purple respectively).

To determine if mRNA correlation behavior (i.e., mRNA co-expression) originates at the promoter, we integrated the pairwise matrix of promoter coordination with the Pearson correlation matrix from the scRNA-seq dataset (Figure 1A and 1B, top). Surprisingly, the integrated matrices did not show any clear structure in the (anti-)coordination matrix when ranked according to the scRNA-seq r coefficient clustering, likely because promoter (anti-)coordination only deviated minimally from independent promoter behavior (Figure S1D). In order to gain deeper insights into the relationship of these two

bioRxiv preprint doi: https://doi.org/10.1101/2022.09.21.508820; this version posted September 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

features that could be hidden in the visual representation of the data, we performed a quantitative analysis of 4 distinct correlation regimes with decreasing average co-expression values (Figure 1C): i) a small 387 by 387 gene cluster showing the highest and lowest values for positive and negative correlation respectively (Figure 1B and 1C, regime I); ii) an expanded regime including regime I in addition to a distant cluster with relatively high correlation values and all pairwise comparisons between these two clusters (Figure 1A and 1C, regime II); ii) the full 5277 by 5277 gene dataset (Figure 1A and C, regime III); iv) the central 4224 by 4226 gene regime where Pearson coefficients were close to 0 (Figure 1A, regime IV). Per regime (I-IV) we performed sub-sampling (see STAR Methods for more detail) and plotted the relationship between mRNA co-expression (i.e., scRNA-seq r) and promoter coordination (Figure 1D) and quantified the relationship by calculating Pearson's correlation r as a proxy for the relationship strength (Figure 1D and 1E). The analysis showed a subtle positive association between promoter coordination and mRNA co-expression in regimes II and III (Figure 1E, red). Together, these data indicate that promoter coordination or anti-coordination does not generate a respective positive or negative correlation at an mRNA level.

To identify any hidden coregulation at a DNA level that was not captured by our promoter coordination analysis, we sought to discern if DNA contact frequencies were higher in gene clusters that are more (anti)correlated at an mRNA level (de Wit, 2020; Soler-Oliva et al., 2017). To this end, the same integration process (Figure S1E) and binned correlation analysis (Figure S1F and S1G) as for promoter coordination (Figure 1A-B and 1C-E respectively) was performed with Hi-C based DNA contact frequencies from a previously published dataset (Nora et al., 2017). Hi-C contact frequencies did not show any clear structure when plotted with respect to RNA co-expression (Figure S1E), nor did detailed quantitative analysis reveal a clear correlation between the Hi-C contact frequencies and mRNA co-expression (Figure S1F and S1G).

mRNA half-life differences contribute to mRNA co-expression

bioRxiv preprint doi: https://doi.org/10.1101/2022.09.21.508820; this version posted September 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Since neither promoter coordination nor DNA contact frequencies underlie mRNA co-expression behavior (Figure 1 and Figure S1), we considered whether differences in gene-specific kinetic parameters could be causing mRNA co-expression. From the intron seq-FISH dataset (Shah et al., 2018), we obtained the fraction of cells that are present in the ON state (i.e., f_{on}) for each gene, which, assuming ergodicity coincides with the fraction of time a promoter is active (Dattani and Barahona, 2017; Desai et al., 2021). Numerically, f_{on} is a function of toggling rates—i.e., $f_{on} = K_{ON}/[K_{ON}+K_{OFF}]$. On average, each promoter is on 12.5% of the time (i.e., f_{on} = 0.125), and each cell is expressing approximately 8% of the ~10,000 genes analyzed at a given moment in time as previously described (Shah et al., 2018). The data (Shah et al., 2018) represents a population of cells where every cell has approximately the same number of genes active at a given timepoint, and the average promoter is in a bursty toggling regime (Munsky et al., 2012), where $K_{ON}<K_{OFF}$ (on average $K_{OFF}=\sim 7xK_{ON}$). These promoter toggling frequencies are in line with previous observations (Bahar Halpern et al., 2015; Hansen et al., 2018). To expand our analysis to post-transcriptional kinetic parameters, we included the half-life of each transcript ($t_{1/2}$) from a third published dataset (Sharova et al., 2009).

First, in order to verify that these three separate datasets (mRNA co-expression, f_{on}, and $t_{1/2}$) could indeed be accurately compared, we determined if the relationship between mRNA expression, promoter toggling, and mRNA degradation behaved as expected. Overall, the fraction of time the promoter is in the ON state (f_{on}) followed a similar trend as the mean mRNA abundance. This is especially evident when considering the trend across all ~5000 genes (Figure 1A, blue and red respectively). This is expected, since the mean mRNA abundance (μ) is proportional to the time a promoter spends in the ON state ($\mu \propto f_{on}.k_{tx}$; where k_{tx} is the transcription rate) (Munsky et al., 2012). We next quantified the strength of the association between mean mRNA abundance and the fraction of time a promoter spends in the ON state (Figure 2A) or mRNA half-life (Figure 2B) as a function of mean expression. At low mRNA expression there is a strong positive association between mean mRNA abundance and the fraction of time a promoter spends in the ON state (f_{on}). This association exhibits an overall decrease with increased mRNA expression levels. In other words, for genes with higher mRNA abundance, mean mRNA expression displays a much weaker association with promoter toggling frequencies than

bioRxiv preprint doi: https://doi.org/10.1101/2022.09.21.508820; this version posted September 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

for genes with low mRNA abundance, indicating that other processes play a more dominant role. Interestingly, genes that show a decreased association between the fraction of time a promoter spends in the ON state (f_{on}) and mean mRNA abundance, show a corresponding increase in association of mRNA half-life ($t_{1/2}$) with mean mRNA abundance (Figure 2A and 2B, grey shaded area). This is in accordance with what is expected from the relationship of mean mRNA abundance, promoter toggling and mRNA degradation (Figure S2A) (Hansen et al., 2018; Munsky et al., 2012). Furthermore, the time a promoter spends in the ON state (f_{on}) follows a similar trend as noise (σ^2_{res}, equivalent to the Fano factor, see STAR Methods for more details), which is most apparent when considering the inset of ~900 genes (Figure 1B, blue and orange respectively). We therefore quantified the strength of association between mRNA noise and the fraction of time a promoter spends in the ON state (Figure S2B). As expected from known dependence of noise on promoter toggling rates (Dar et al., 2012), mRNA noise is strongly associated with the fraction of time the respective promoter spends in the ON state (Figure S2A and S2B) with this association becoming stronger at higher mean mRNA abundances. Taken together, these results demonstrate that data obtained from different studies can reliably be compared, and relationships quantified.

To identify if pairwise differences in promoter toggling kinetics (f_{on}) or mRNA half-lives ($t_{1/2}$) underlie co-expression of genes at an mRNA level, we followed the same workflow as in Figure 1 and Figure S1. We performed random subsampling of regimes I-IV (Figure 1C), and plotted the relationship between pairwise mRNA co-expression (i.e., scRNA-seq r) and either the pairwise difference in mRNA half-life (Figure 2C and 2D) or the pairwise difference in the ON-fraction of the respective promoters (Figure 2E and 2F). In the regime with the highest mRNA co-expression (i.e., most positive and most negative scRNA-seq r values, regime I), both the promoter toggling and the mRNA half-life differences have a strong relationship with mRNA co-expression (Figure 2C and E). Nevertheless, this relationship becomes less prominent for promoter toggling when we sample regimes with more subtle mRNA co-expression at an mRNA level (Figure 2F, S2E and S2F). Conversely, pairwise differences in mRNA half-lives remain more strongly associated with mRNA co-expression even at lower co-expression levels (Figure 2D, S2C, and S2D). Together, these data show that mRNA degradation is coupled to

bioRxiv preprint doi: https://doi.org/10.1101/2022.09.21.508820; this version posted September 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

both positive and negative co-expression across all ~5000 genes analyzed, where similar half-lives are associated with positive co-expression, while different half-lives are associated with negative co-expression of two genes (Figure 2I).

mRNA half-life and promoter toggling demonstrate compensatory behavior

The finding that both pairwise differences in promoter toggling as well as mRNA half-life are associated with mRNA co-expression for a small subset (<10%) of genes (Figure 2C and E), and seem to have an inverse effect on mean mRNA abundance (Figure 2A and 2B), led us to question the relationship between promoter toggling and mRNA half-life. Therefore, we quantified both the absolute and directional pairwise differences in half-life and promoter toggling (Figure 2G and S2G, see STAR Methods for details). In the regime with the highest co-expression values (regime I), there is a strong positive correlation between absolute differences in promoter toggling and mRNA half-life (Figure 2G and S2G, top). This means that gene-pairs with a large difference in promoter toggling kinetics, also display a large difference in mRNA half-life. Across all sampled regimes (I-IV) half-life and promoter toggling showed compensatory (i.e., genes with higher f_{on} have lower $t_{1/2}$) rather than synergistic (i.e., the groups of genes with a higher f_{on} also demonstrate a higher $t_{1/2}$) behavior (Figure 2G and S2G, bottom). Intuitively, this compensation can be explained by the inherent relationship between promoter toggling and mRNA half-life ($\mu = f_{on}.k_{tx}/k_d$, where k_d is mRNA degradation rate). Therefore, at similar mean mRNA expression levels we expect mRNA half-life to be inversely proportional to promoter toggling (Figure S2H). This relationship between two seemingly distal kinetic parameters emphasizes the balance between transcriptional and post-transcriptional processes that together orchestrate the gene expression landscapes, to the extent that the one cannot be understood without the other.

DISCUSSION

In this brief report we integrated gene co-expression data from scRNA-seq analysis with other data modalities (Nora et al., 2017; Shah et al., 2018; Sharova et al., 2009) that provide context from various processes involved in gene expression to assess whether specific kinetic steps influence scRNA-seq data. Our goal was to interrogate the relationships between the observed correlations in single-cell

bioRxiv preprint doi: https://doi.org/10.1101/2022.09.21.508820; this version posted September 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

mRNA levels and the processes upstream and downstream of transcription without assuming specific regulatory architectures. The analysis showed that promoter coordination was not playing a discernible role in gene co-expression especially when (anti)correlation of mRNA is the strongest (Figure 1). Yet, it is important to note that promoter behavior might not be properly captured by a single snapshot, as promoter coordination could emerge from more complex temporal dynamics where promoters are, for example, likely to be ON in close temporal proximity but not necessarily at the same time. Unfortunately, the techniques for assessing promoter activity with temporal resolution (e.g. MS2 tagging) are limited to lower throughput applications (Hocine et al., 2013; Wan et al., 2021) and require genetic modification, as a specific sequence has to be added to a gene for the mRNA to be tracked (Tantale et al., 2016).

We next included other kinetic parameters in the analysis. Strikingly, mRNA half-life demonstrated a strong negative association with scRNA-seq co-expression r values (Figure 2). Additionally, for specific regimes of very evident (anti)correlation in mRNA levels, differences in promoter ON ratio also seemed to play an important role. In fact, in these regions the differences in f_{on} and half-life demonstrate clear compensatory behavior. It is possible that there is a mechanistic reason why this occurs — i.e., is it a requirement for genes to have extreme correlation values to show this specific behavior — or that this relationship is an evolutionary consequence. Many elegant studies on gene expression regulation focus on chromatin and transcriptional events (Larson et al., 2013; Lenstra et al., 2016; Zinani et al., 2022), and downstream processes should not be underestimated. This work together with other recent publications (Aizarani et al., 2019; Gilbertson et al., 2018; Hansen et al., 2018; Hansen and Weinberger, 2019; Matkovic et al., 2022) is enforcing the necessity of including post-transcriptional events.

bioRxiv preprint doi: https://doi.org/10.1101/2022.09.21.508820; this version posted September 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

ACKNOWLEDGEMENTS:

We thank all members of the Hansen lab for the thoughtful discussion and suggestions. We also thank Hendrik Marks for his generous gift of mESC-E14 cell line. M.M.K.H. acknowledges support from the Dutch Research Council (NWO) ENW-XS awards (OCENW.XS3.055 and OCENW.XS21.2.050).

AUTHOR CONTRIBUTIONS:

Ó.G.B. and M.M.K.H. conceived the study. Ó.G.B., P.V., B.M, and M.M.K.H. designed the study. Ó.G.B. performed single-cell RNA-seq experiments. Ó.G.B., P.V., and M.M.K.H. designed and performed the data analyses. Ó.G.B. and M.M.K.H. wrote and P.V. and B.M. edited the manuscript

DECLARATION OF INTERESTS: Authors declare no competing interests.

DATA AND MATERIALS AVAILABILITY: Sequencing data from single-cell RNA-seq will be deposited onto GEO. Custom code for analysis of sequencing data and mathematical modeling will be made available on GitHub.

bioRxiv preprint doi: https://doi.org/10.1101/2022.09.21.508820; this version posted September 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

References:

Aizarani, N., Saviano, A., Sagar, Mailly, L., Durand, S., Herman, J.S., Pessaux, P., Baumert, T.F., and Grun, D. (2019). A human liver cell atlas reveals heterogeneity and epithelial progenitors. Nature *572*, 199-204.

Andrews, T.S., Atif, J., Liu, J.C., Perciani, C.T., Ma, X.Z., Thoeni, C., Slyper, M., Eraslan, G., Segerstolpe, A., Manuel, J.*, et al.* (2022). Single-Cell, Single-Nucleus, and Spatial RNA Sequencing of the Human Liver Identifies Cholangiocyte and Mesenchymal Heterogeneity. Hepatol Commun *6*, 821-840.

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

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

Coulon, A., Chow, C.C., Singer, R.H., and Larson, D.R. (2013). Eukaryotic transcriptional dynamics: from single molecules to cell populations. Nat Rev Genet *14*, 572-584.

Crow, M., and Gillis, J. (2018). Co-expression in Single-Cell Analysis: Saving Grace or Original Sin? Trends Genet *34*, 823-831.

Crow, M., Suresh, H., Lee, J., and Gillis, J. (2022). Coexpression reveals conserved gene programs that co-vary with cell type across kingdoms. Nucleic Acids Res.

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 U S A *109*, 17454-17459.

Dattani, J., and Barahona, M. (2017). Stochastic models of gene transcription with upstream drives: exact solution and sample path characterization. J R Soc Interface *14*.

de Wit, E. (2020). TADs as the Caller Calls Them. J Mol Biol *432*, 638-642.

Desai, R.V., Chen, X., Martin, B., Chaturvedi, S., Hwang, D.W., Li, W., Yu, C., Ding, S., Thomson, M., Singer, R.H.*, et al.* (2021). A DNA repair pathway can regulate transcriptional noise to promote cell fate transitions. Science *373*, eabc6506.

Eisen, M.B., Spellman, P.T., Brown, P.O., and Botstein, D. (1998). Cluster analysis and display of genome-wide expression patterns. Proc Natl Acad Sci *95*, 14863-14868.

Esmaili, S., Langfelder, P., Belgard, T.G., Vitale, D., Azardaryany, M.K., Alipour Talesh, G., Ramezani-Moghadam, M., Ho, V., Dvorkin, D., Dervish, S.*, et al.* (2021). Core liver homeostatic co-expression networks are preserved but respond to perturbations in an organism- and disease-specific manner. Cell Syst *12*, 432-445 e437.

Foreman, R., and Wollman, R. (2020). Mammalian gene expression variability is explained by underlying cell state. Molecular Systems Biology *16*, e9146.

Gilbertson, S., Federspiel, J.D., Hartenian, E., Cristea, I.M., and Glaunsinger, B. (2018). Changes in mRNA abundance drive shuttling of RNA binding proteins, linking cytoplasmic RNA degradation to transcription. Elife *7*.

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.

bioRxiv preprint doi: https://doi.org/10.1101/2022.09.21.508820; this version posted September 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

290 Gu, Q., Nagaraj, S.H., Hudson, N.J., Dalrymple, B.P., and Reverter, A. (2011). Genome-wide patterns
291 of promoter sharing and co-expression in bovine skeletal muscle. BMC Genomics *12*, 23.

292 Hafemeister, C., and Satija, R. (2019). Normalization and variance stabilization of single-cell RNA-seq
293 data using regularized negative binomial regression. Genome Biol *20*, 296.

294 Han, J., DePinho, R.A., and Maitra, A. (2021). Single-cell RNA sequencing in pancreatic cancer. Nat
295 Rev Gastroenterol Hepatol *18*, 451-452.

296 Hansen, M.M.K., Desai, R.V., Simpson, M.L., and Weinberger, L.S. (2018). Cytoplasmic
297 Amplification of Transcriptional Noise Generates Substantial Cell-to-Cell Variability. Cell Syst *7*, 384-
298 397 e386.

299 Hansen, M.M.K., and Weinberger, L.S. (2019). Post-Transcriptional Noise Control. Bioessays *41*,
300 e1900044.

301 Harper, C.V., Finkenstadt, B., Woodcock, D.J., Friedrichsen, S., Semprini, S., Ashall, L., Spiller, D.G.,
302 Mullins, J.J., Rand, D.A., Davis, J.R.*, et al.* (2011). Dynamic analysis of stochastic transcription cycles.
303 PLoS Biol *9*, e1000607.

304 Harris, B.D., Crow, M., Fischer, S., and Gillis, J. (2021). Single-cell co-expression analysis reveals that
305 transcriptional modules are shared across cell types in the brain. Cell Syst *12*, 748-756 e743.

306 Hocine, S., Raymond, P., Zenklusen, D., Chao, J.A., and Singer, R.H. (2013). Single-molecule analysis
307 of gene expression using two-color RNA labeling in live yeast. Nat Methods *10*, 119-121.

308 Hwang, B., Lee, J.H., and Bang, D. (2018). Single-cell RNA sequencing technologies and
309 bioinformaticspipelines. Experimental & Molecular Medicine *50*, 1-14.

310 Jeong, D., Lim, S., Lee, S., Oh, M., Cho, C., Seong, H., Jung, W., and Kim, S. (2021). Construction of
311 Condition-Specific Gene Regulatory Network Using Kernel Canonical Correlation Analysis. Front
312 Genet *12*, 652623.

313 Kepler, T.B., and Elston, T.C. (2001). Stochasticity in transcriptional regulation: origins, consequences,
314 and mathematical representations. Biophys J *81*, 3116-3136.

315 Kowalczyk, M.S., Tirosh, I., Heckl, D., Rao, T.N., Dixit, A., Haas, B.J., Schneider, R.K., Wagers, A.J.,
316 Ebert, B.L., and Regev, A. (2015). Single-cell RNA-seq reveals changes in cell cycle and differentiation
317 programs upon aging of hematopoietic stem cells. Genome Res *25*, 1860-1872.

318 Larson, D.R., Fritzsch, C., Sun, L., Meng, X., Lawrence, D.S., and Singer, R.H. (2013). Direct
319 observation of frequency modulated transcription in single cells using light activation. Elife *2*, e00750.

320 Lause, J., Berens, P., and Kobak, D. (2021). Analytic Pearson residuals for normalization of single-cell
321 RNA-seq UMI data. Genome Biol *22*, 258.

322 Lenstra, T.L., Rodriguez, J., Chen, H., and Larson, D.R. (2016). Transcription Dynamics in Living
323 Cells. Annu Rev Biophys *45*, 25-47.

324 Matkovic, R., Morel, M., Lanciano, S., Larrous, P., Martin, B., Bejjani, F., Vauthier, V., Hansen,
325 M.M.K., Emiliani, S., Cristofari, G.*, et al.* (2022). TASOR epigenetic repressor cooperates with a
326 CNOT1 RNA degradation pathway to repress HIV. Nature Communications *13*, 66.

bioRxiv preprint doi: https://doi.org/10.1101/2022.09.21.508820; this version posted September 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Matsumoto, H., Kiryu, H., Furusawa, C., Ko, M.S.H., Ko, S.B.H., Gouda, N., Hayashi, T., and Nikaido, I. (2017). SCODE: an efficient regulatory network inference algorithm from single-cell RNA-Seq during differentiation. Bioinformatics *33*, 2314-2321.

Munsky, B., Neuert, G., and van Oudenaarden, A. (2012). Using Gene Expression Noise to Understand Gene Regulation. Science *336*, 183-187.

Muraro, M.J., Dharmadhikari, G., Grün, D., Groen, N., Dielen, T., Jansen, E., van Gurp, L., Engelse, M.A., Carlotti, F., de Koning, E.J.*, et al.* (2016). A Single-Cell Transcriptome Atlas of the Human Pancreas. Cell Syst *3*, 385-394.e383.

Nora, E.P., Goloborodko, A., Valton, A.-L., Gibcus, J.H., Uebersohn, A., Abdennur, N., Dekker, J., Mirny, L.A., and Bruneau, B.G. (2017). Targeted Degradation of CTCF Decouples Local Insulation of Chromosome Domains from Genomic Compartmentalization. Cell *169*, 930-944.e922.

Oliver, S. (2000). Guilt-by-association goes global. Nature *403*, 601-602.

Payen, V.L., Lavergne, A., Alevra Sarika, N., Colonval, M., Karim, L., Deckers, M., Najimi, M., Coppieters, W., Charloteaux, B., Sokal, E.M.*, et al.* (2021). Single-cell RNA sequencing of human liver reveals hepatic stellate cell heterogeneity. JHEP Rep *3*, 100278.

Qadir, M.M.F., Alvarez-Cubela, S., Klein, D., van Dijk, J., Muniz-Anquela, R., Moreno-Hernandez, Y.B., Lanzoni, G., Sadiq, S., Navarro-Rubio, B., Garcia, M.T.*, et al.* (2020). Single-cell resolution analysis of the human pancreatic ductal progenitor cell niche. Proc Natl Acad Sci U S A *117*, 10876-10887.

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

Ribeiro, D.M., Ziyani, C., and Delaneau, O. (2021). Shared regulation and functional relevance of local gene co-expression revealed by single cell analysis

. bioRxiv.

Ronen, M., Rosenberg, R., Shraiman, B.I., and Alon, U. (2002). Assigning numbers to the arrows: parameterizing a gene regulation network by using accurate expression kinetics. Proc Natl Acad Sci U S A *99*, 10555-10560.

Salehi, N., Karimi-Jafari, M.H., Totonchi, M., and Amiri-Yekta, A. (2021). Integration and gene co-expression network analysis of scRNA-seq transcriptomes reveal heterogeneity and key functional genes in human spermatogenesis. Sci Rep *11*, 19089.

Shah, S., Takei, Y., Zhou, W., Lubeck, E., Yun, J., Eng, C.L., Koulena, N., Cronin, C., Karp, C., Liaw, E.J.*, et al.* (2018). Dynamics and Spatial Genomics of the Nascent Transcriptome by Intron seqFISH. Cell *174*, 363-376 e316.

Sharova, L.V., Sharov, A.A., Nedorezov, T., Piao, Y., Shaik, N., and Ko, M.S. (2009). Database for mRNA half-life of 19 977 genes obtained by DNA microarray analysis of pluripotent and differentiating mouse embryonic stem cells. DNA Res *16*, 45-58.

Soler-Oliva, M.E., Guerrero-Martinez, J.A., Bachetti, V., and Reyes, J.C. (2017). Analysis of the relationship between coexpression domains and chromatin 3D organization. PLoS Comput Biol *13*, e1005708.

bioRxiv preprint doi: https://doi.org/10.1101/2022.09.21.508820; this version posted September 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

366 Stuart, T., Butler, A., Hoffman, P., Hafemeister, C., Papalexi, E., Mauck, W.M., 3rd, Hao, Y., Stoeckius,
367 M., Smibert, P., and Satija, R. (2019). Comprehensive Integration of Single-Cell Data. Cell *177*, 1888-
368 1902 e1821.

369 Suter, D.M., Molina, N., Gatfield, D., Schneider, K., Schibler, U., and Naef, F. (2011). Mammalian
370 Genes Are Transcribed with Widely Different Bursting Kinetics. Science *332*, 472-474.

371 Tantale, K., Mueller, F., Kozulic-Pirher, A., Lesne, A., Victor, J.-M., Robert, M.-C., Capozi, S.,
372 Chouaib, R., Bäcker, V., Mateos-Langerak, J.*, et al.* (2016). A single-molecule view of transcription
373 reveals convoys of RNA polymerases and multi-scale bursting. Nature Communications *7*, 12248.

374 Travaglini, K.J., Nabhan, A.N., Penland, L., Sinha, R., Gillich, A., Sit, R.V., Chang, S., Conley, S.D.,
375 Mori, Y., Seita, J.*, et al.* (2020). A molecular cell atlas of the human lung from single-cell RNA
376 sequencing. Nature *587*, 619-625.

377 Vandenbon, A. (2022). Evaluation of critical data processing steps for reliable prediction of gene co-
378 expression from large collections of RNA-seq data. PLoS One *17*, e0263344.

379 Vivian Li, W., and Li, Y. (2021). scLink: Inferring Sparse Gene Co-expression Networks from Single-
380 cell Expression Data. Genomics Proteomics Bioinformatics *19*, 475-492.

381 Wan, Y., Anastasakis, D.G., Rodriguez, J., Palangat, M., Gudla, P., Zaki, G., Tandon, M., Pegoraro,
382 G., Chow, C.C., Hafner, M.*, et al.* (2021). Dynamic imaging of nascent RNA reveals general principles
383 of transcription dynamics and stochastic splice site selection. Cell *184*, 2878-2895.e2820.

384 Wang, X., Choi, D., and Roeder, K. (2021). Constructing local cell-specific networks from single-cell
385 data. Proc Natl Acad Sci U S A *118*.

386 Weinberger, L.S., Burnett, J.C., Toettcher, J.E., Arkin, A.P., and Schaffer, D.V. (2005). Stochastic gene
387 expression in a lentiviral positive-feedback loop: HIV-1 Tat fluctuations drive phenotypic diversity.
388 Cell *122*, 169-182.

389 Xie, J., Chen, L., Cao, Y., Wu, D., Xiong, W., Zhang, K., Shi, J., and Wang, M. (2021). Single-Cell
390 Sequencing Analysis and Weighted Co-Expression Network Analysis Based on Public Databases
391 Identified That TNC Is a Novel Biomarker for Keloid. Front Immunol *12*, 783907.

392 Yin, W., Mendoza, L., Monzon-Sandoval, J., Urrutia, A.O., and Gutierrez, H. (2021). Emergence of
393 co-expression in gene regulatory networks. PLoS One *16*, e0247671.

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

396 Zinani, O.Q.H., Keseroglu, K., and Ozbudak, E.M. (2022). Regulatory mechanisms ensuring
397 coordinated expression of functionally related genes. Trends Genet *38*, 73-81.

398

399

bioRxiv preprint doi: https://doi.org/10.1101/2022.09.21.508820; this version posted September 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

STAR METHODS

Key resources table

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Cell lines		
mESC E14 type (129/Ola background)	(van Mierlo et al., 2019)	
Cell culture reagents / utensils		
6-well Clear TC-treated Multiple Well Plates	Costar	REF 3516
Gelatin	Sigma-Aldrich	48723/500g
Dulbecco's Modified Eagle Media, high glucose, pyruvate (Gibco)	Fisher Scientific	#11594486
Beta-mercaptoethanol	Fisher Scientific	#11528926
Sodium pyruvate (Gibco)	Fisher Scientific	#11530396
Penicillin/Streptomycin (Gibco)	Fisher Scientific	#15140122
ESGRO recombinant Mouse Leukemia Inhibitory Factor	Merckmillipore	ESG1107
Fetal Bovine Serum	Fisher Scientific	#A3840002
Phosphate-buffered saline (PBS) (Gibco)	Fisher Sientific	#10728775
0.05% Trypsin-EDTA (Gibco)	Fisher Scientific	#11500636
DMSO	Sigma-Aldrich	D5879-1L
Acridine Orange/Propidium Iodide Stain	Logos Biosystems	F23001
LUNA Cell Counting Slides	Logos Biosystems	L12001
LUNA-FL Dual Fluorescence Cell Counter	Logos Biosystems	L20001-LG
Deposited data		
Intron-seq FISH mESC	(Shah et al., 2018)	
Hi-C contact frequencies mESC	(Nora et al., 2017)	
Half-life mRNA mESC	(Harova et al., 2009)	
scRNA-seq mESC	Produced in house	
Software and algorithms		
Seurat V3	(Stuart et al., 2019)	

bioRxiv preprint doi: https://doi.org/10.1101/2022.09.21.508820; this version posted September 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

RESOURCE AVAILABILITY

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contact, M.M.K.H (maike.hansen@ru.nl).

METHODS DETAILS

Single-cell RNA sequencing in mESCs

mESC-E14 (mESCs) in this study were obtained from Hendrik Marks's group at Radboud University. mESC were seeded per well in gelatin coated Costar® 6-well Clear TC-treated Multiple Well Plates (REF 3516), in serum/LIF culture media consisting of Dulbecco's Modified Eagle Media, supplemented with 0.1mM beta-mercaptoethanol, 1000 U/mL of Penicillin, 0.1 mg/mL of Streptomycin, 1mM of sodium pyruvate, 1000 U/mL of ESGRO recombinant Mouse Leukemia Inhibitory Factor and 15% of ES-qualified heat inactivated Fetal Bovine Serum. Cells were grown for 30 hours in a 37°C incubator with 5% CO_2 atmosphere. Cells were exposed to 0.007% EtOH for the last 6 hours. Cells were then washed once with PBS and detached with 0.05% trypsin-EDTA for 4 minutes at 37°C. Detached cells were pelleted and resuspended in 1mL of freezing media consisting of 80% culture media, 10% of extra heat inactivated FBS and 10% of DMSO. Proper viability of the cells prior to freezing was quantified with propidium iodide/acridine orange staining in a LUNA-FL Dual Fluorescence Cell Counter. Frozen cells were delivered to the commercial company Single Cell Discoveries (https://www.scdiscoveries.com) in dry ice. Single cell barcoding was performed in a 10x microfluidic genomic chip in order to encapsulate individual cells in water droplets in oil, containing cell specific barcoded beads. Labeled RNA molecules were pooled and subjected to a poly-A specific reverse transcription. cDNA molecules where linearly amplified using an *in vitro* transcription reaction and the final sequencing library was obtained through a second step of reverse transcription. 3'-end sequencing was performed, followed by quality control and genomic mapping. Final read count per gene and per cell matrices where generated and used in posterior analysis steps.

Calculation of gene-to-gene correlation in scRNAseq data and gene clustering

Preprocessing of raw counts was performed with SeuratV3 (Stuart et al., 2019). Genes detected in less than 5 cells and cells with less than 200 detected genes, less than 24000 total RNA count and cells with more than 5% mitochondrial RNAs were discarded. Cells were given a score reflecting their cell cycle stage based on the expression of specific cell cycle markers (Kowalczyk et al., 2015). A normalized and scaled count matrix was generated by using the SCTransform method (Hafemeister and Satija, 2019) with an offset of θ=100 (Lause et al., 2021). The scaling method accounted for the total RNA count per cell and utilized the percent of mitochondrial RNA and the cell cycle scoring as variables to regress. The scaled count matrix was used to calculate the full correlation matrix using the Pearson or the Spearman methods. Pairwise gene-to-gene Pearson correlation matrices were clustered using complete-linkage hierarchical agglomerative clustering.

Gene-to-gene promoter coordination calculation

From the intron-seq FISH data (Shah et al., 2018), the status of the promoter was assessed according to the presence or absence of intronic signal. Promoter ON states are defined by the presence of intronic signals (i.e., signal ≥1), while OFF states are defined by the lack thereof (i.e., signal = 0). Therefore, per gene:

$$f_{on} = \frac{Count_{signal\geq 1}}{Count_{signal\geq 1} + Count_{signal=0}}$$ Equation 1

bioRxiv preprint doi: https://doi.org/10.1101/2022.09.21.508820; this version posted September 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

For each pair of genes in the dataset, an observed anti-coordination score (O) was defined as the proportion of cell pairs where the genes show an opposite behavior (ON/OFF or OFF/ON) over the total number of cell-to-cell comparisons as described in equation 1. The observed anti-coordination scored was corrected by the chance of these behavior appearing from independently behaving promoters. To this extent, the proportion of cells in which each gene is ON was defined as its f_{on}, and the expected anti-coordination score (E) was calculated as described in equation 2, where f_{on_1} and f_{on_2} represent the ON ratios (f_{on}) of each of the two genes included in the calculation. Observed anti-correlation score was corrected by calculating the percentual fold change of the observed anti-coordination score with respect to the expected score by chance as described in equation 3, where O is the observed anti-coordination score and E is the expected anti-coordination score. In order to present this data in a more intuitive way, we transformed the corrected anti-coordination score in a coordination score as described in equation 5.

$$O = (Count_{ON/OFF} + Count_{OFF/ON})/(Count_{ON/ON} + Count_{ON/OFF} + Count_{OFF/ON} + Count_{OFF/OFF}) \quad \text{Equation 2}$$

$$E = f_{on_1}(1 - f_{on_2}) + f_{on_2}(1 - f_{on_1}) \quad \text{Equation 3}$$

$$Corrected\ anticoordination\ score = (\frac{O-E}{E} * 100) \quad \text{Equation 4}$$

$$Corrected\ coordination\ score = -1 * Corrected\ anticoordination\ score \quad \text{Equation 5}$$

Estimation of the relationship between promoter toggling rates

The effective ratio between K_{on} and K_{off} is calculated from the f_{on} quantification (Equation 1) where:

$$f_{on} = (\frac{K_{on}}{K_{on}+K_{off}}) \quad \text{Equation 6}$$

And the average $f_{on} = 0.125$. Therefore,

$$0.125 = (\frac{K_{on}}{K_{on} + K_{off}})$$

and,

$$K_{on} + K_{off} = 8 * K_{on}$$

$$K_{off} = 7 * K_{on} \quad \text{Equation 7}$$

Calculation of gene expression mean and noise from scRNA-seq dataset

Single gene mean and noise values were obtained through the analysis of the raw read count matrix with SeuratV3 (Stuart et al., 2019). Next, genes detected in less than 5 cells and cells with less than 200 detected genes, less than 24000 total RNA count and cells with more than 5% mitochondrial RNAs were discarded. Cells were given a score reflecting their cell cycle stage based on the expression of specific cell cycle markers (Kowalczyk et al., 2015). Mean gene expression values and residual variance values (σ^2_{res}) were calculated using the SCTransform method with an offset value of θ=100, using the percent of mitochondrial RNA and the cell cycle score as variables to regress. In short, biological noise is obtained through the fitting of the read counts to a negative binomial distribution modeling technical noise and obtaining the residual variance not explained by the model of technical noise (Hafemeister and Satija, 2019). These residual variance values (σ^2_{res}) are comparable to the Fano factor (Lause et al., 2021). Mean gene expression was then calculated as the $\log_2(\mu)$ denoted as μ_{log2}.

bioRxiv preprint doi: https://doi.org/10.1101/2022.09.21.508820; this version posted September 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Rolling averages and analysis of correlation for gene expression parameters

Subsampling

For the four sampling regimes, we randomly sub-sampled 10000 sets of 30 by 30 genes (sampling regimes II, III and IV) or 3 by 3 genes (sampling regime I), within each defined region. Sampling regime I is a much smaller region so we scaled down the subsampling accordingly.

Rolling averages

Genes were sorted according to the matrix resulting from the clustering of the pairwise scRNA-seq Pearson's coefficients. Starting from the first ranked gene, groups of successive 30 genes were made in an iterative fashion moving 1 position forward in the gene list with each iteration. For each group of 30 genes the average of fon, μ_{log2}, σ^2_{std} and $t_{1/2}$ was obtained. This process generated four sequential vectors of averages in which each position corresponded to the same group of genes from the scRNA-seq clustering order.

Correlation for gene expression parameters.

To analyze the correlation between the kinetic parameters (fon and $t_{1/2}$) and the gene expression output parameters (μ_{log2} and σ^2_{std}), Pearson coefficients as well as the 95% confidence intervals for the Pearson coefficients were calculated between the corresponding pairwise values for each parameter in the previously calculated rolling mean vectors. In order to classify gene groups in different regimes, they were grouped in bins according to their μ_{log2} average value by using a rolling threshold for this value.

Binned analysis of correlations between differences in kinetic parameters

Two random groups of 30 or 3 contiguous genes in the clustered scRNAseq r matrix were selected in order to form a 30x30 or 3x3 comparative space. Restrictions in the selection of gene groups were applied in order to create the 4 sampling regions described previously. For each selected space, the average r coefficient was calculated. By matrix symmetry the average promoter coordinated behavior for the corresponding group of genes was obtained from the processed intron-seqFISH data and the average Hi-C contact frequency was obtained with the same method from the Hi-C data. The average difference in mean, noise, ON ratio and half-life (i.e., μ_{log2}, σ^2_{std}, fon, and $t_{1/2}$) was obtained by averaging these parameters for each of the two groups in the comparison and obtaining the directional or absolute difference as described in equations 8 and 9 respectively.

$$directional\ difference = log_2\left(\frac{\langle . \rangle_{group\ 1}}{\langle . \rangle_{group\ 2}}\right) \quad \text{Equation 8}$$

$$absolute\ difference = |directional\ difference| \quad \text{Equation 9}$$

bioRxiv preprint doi: https://doi.org/10.1101/2022.09.21.508820; this version posted September 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

FIGURE LEGENDS

Figure 1: mRNA co-expression is not explained by promoter behavior.

(A-B) Top: Combined matrix showing clustered pairwise scRNA-seq Pearson's r coefficients (bottom half) and corresponding promoter (anti-)coordination scores (top). Red is positive Pearson's r and blue is negative Pearson's r. Green is more coordination and purple is more anti-coordination of promoters compared to what is expected from random promoter behavior. Bottom: Rolling averages for ON-ratio (f_{on}), relative mean mRNA expression (μ_{log2}), noise (~Fano factor, σ^2_{res}), and mRNA half-life ($t_{1/2}$, hours) ranked in the same order as the clustered matrix, with the background color representing the respective values.

(C) Four sampling regimes (I-IV) corresponding to decreasing average co-expression (i.e., scRNA-seq Pearson's r) from left to right.

(D-E) mRNA co-expression (i.e., scRNAseq Pearson's r coefficient) versus average (anti-)coordination of pairwise promoters for regimes I, II, III and IV. **(D)** Color (blue to red) represent the average scRNAseq Pearson r coefficient. **(E)** Correlation between mRNA co-expression and promoter coordination, with linear regression line plotted in yellow and the background color corresponding to Pearson's r correlation coefficient as a proxy for relationship strength.

bioRxiv preprint doi: https://doi.org/10.1101/2022.09.21.508820; this version posted September 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 2: Differences in mRNA half-lives are most strongly associated with mRNA co-expression.

(A) Quantification (Pearson's r coefficient) of association between promoter ON ratio (f_{on}) and relative mean mRNA expression (μ_{log2}) calculated by imposing specific expression mean thresholds. Color bar represents the average promoter ON ratio (f_{on}) for genes in each corresponding bin (error bar is 95% confidence interval).

(B) Quantification (Pearson's r coefficient) of association between transcript half-life ($t_{1/2}$) and mean mRNA expression (μ_{log2}) calculated by imposing specific expression mean thresholds. Color bar represents the average transcript half-life ($t_{1/2}$) for genes in each corresponding bin (error bar is 95% confidence interval).

(C-D) Scatter plot (left) and correlation (right) between average mRNA co-expression (i.e., scRNAseq Pearson's r coefficient) and average difference in transcript half-life (C) or promoter ON ratio (D) for regime I.

(E-F) Scatter plot (left) and correlation (right) between average mRNA co-expression (i.e., scRNAseq Pearson's r coefficient) and average difference in transcript half-life (E) or promoter ON ratio (F) for regime II.

(G) Top: correlation between average absolute difference in transcript half-life and average absolute difference in promoter ON ratio. Bottom: correlation between average directional difference in transcript half-life and average directional difference in promoter ON ratio. Sampling region corresponds to regime I.

(H) Scatter plot showing correlation between average ON ration and half-lives for genes where $\log_2(\mu)>1$. Dots are colored according to $\log_2(\mu)$ to compare with Figure S2H.

(I) Schematic illustrating that sampled region with decreasing average mRNA co-expression behavior (i.e., regimes I to IV, left to right) show a decreasing correlation between mRNA co-expression and mRNA half-life.

bioRxiv preprint doi: https://doi.org/10.1101/2022.09.21.508820; this version posted September 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure S1: Gene-to-gene contacts do not explain mRNA co-expression (Related to Figure 1).

(A) Schematic illustration of two genes that can exhibit mRNA co-expression by a shared transcriptional (upstream) kinetic step or a shared post-transcriptional (downstream) kinetic step.

(B) Scatter plots representing scaled single cell mRNA expression values for pairs of genes where both Pearson and Spearman correlation coefficients give the same r values (left), and where Spearman coefficients are unreliable (right).

(C) Combined matrix showing clustered gene-to-gene scRNAseq expression Pearson r coefficients (bottom half) and corresponding Spearman coefficients (top half).

(D) Frequency of pairwise promoter interactions that are either more coordinated (green) or more anti-coordinated (purple) than expected from random promoter behavior. Shaded area represents 90% of pairwise comparisons.

(E) Combined matrix showing clustered gene-to-gene scRNAseq expression Pearson's r coefficients (bottom half) and corresponding Hi-C contact frequencies (top half).

(F-G) mRNA co-expression (i.e., scRNAseq Pearson's r coefficient) versus average Hi-C contact frequencies for regions I, II, III and IV. **(F)** Color (blue to red) represent the average scRNAseq Pearson r coefficient. **(G)** Correlation between mRNA co-expression and Hi-C contact frequencies, with linear regression line plotted in yellow and the background color corresponding to Pearson's r correlation coefficient.

bioRxiv preprint doi: https://doi.org/10.1101/2022.09.21.508820; this version posted September 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure S2: Differences in promoter toggling does not associate with mRNA co-expression and shows compensation behavior with mRNA degradation (Related to Figure 2).

(A) Schematic illustrating how promoter toggling can be impacted by altering either the switching ON of a promoter (K_{on}) or the switching OFF of a promoter (K_{off}).

(B) Quantification (Pearson's r coefficient) of association between promoter toggling (f_{on}) and relative mRNA noise (Fano, μ_{log2}) calculated by imposing specific expression mean thresholds. Color bar represents the average promoter toggling (f_{on}) for genes in each corresponding bin (error bar is 95% confidence interval).

(C-D) Scatter plot (left) and correlation (right) between average mRNA co-expression (i.e., scRNAseq Pearson's r coefficient) and average difference in transcript half-life (C) or promoter ON ratio (D) for regime III.

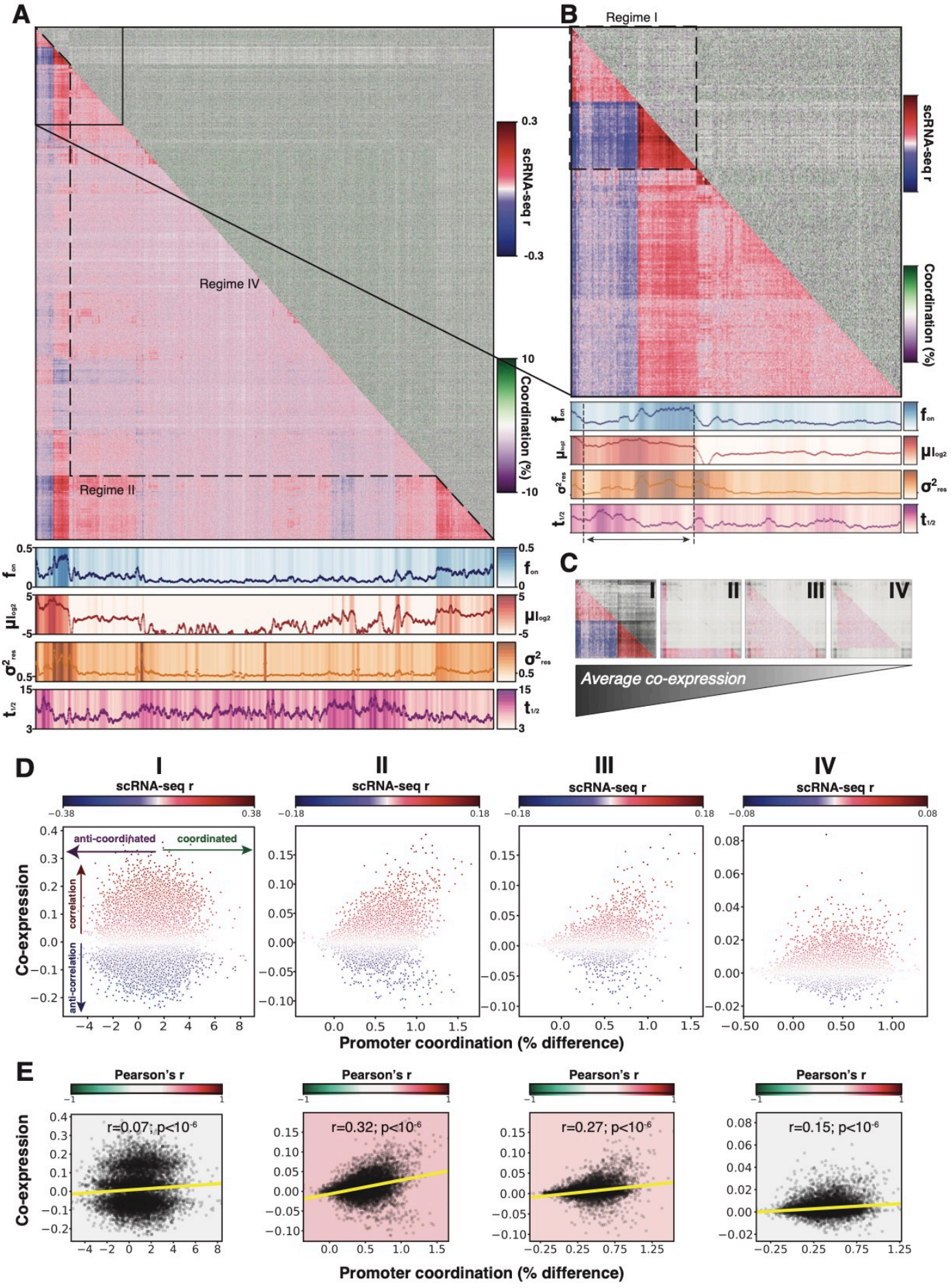
(E-F) Scatter plot (left) and correlation (right) between average mRNA co-expression (i.e., scRNAseq Pearson's r coefficient) and average difference in transcript half-life (E) or promoter ON ratio (F) for regime IV.

(G) Top: correlation between average absolute difference in transcript half-life and average absolute difference in promoter ON ratio. Bottom: correlation between average directional difference in transcript half-life and average directional difference in promoter ON ratio. Sampling region corresponds to regime II-IV from left to right.

(H) Scatter plot showing correlation between average ON ration and half-lives per gene. Dots colored according to $\log_2(\mu)$ (left) or noise (right). Inset: Scatter plot showing correlation between average ON ration and half-lives for genes where $\log_2(\mu)>1$. Dots colored according to noise.

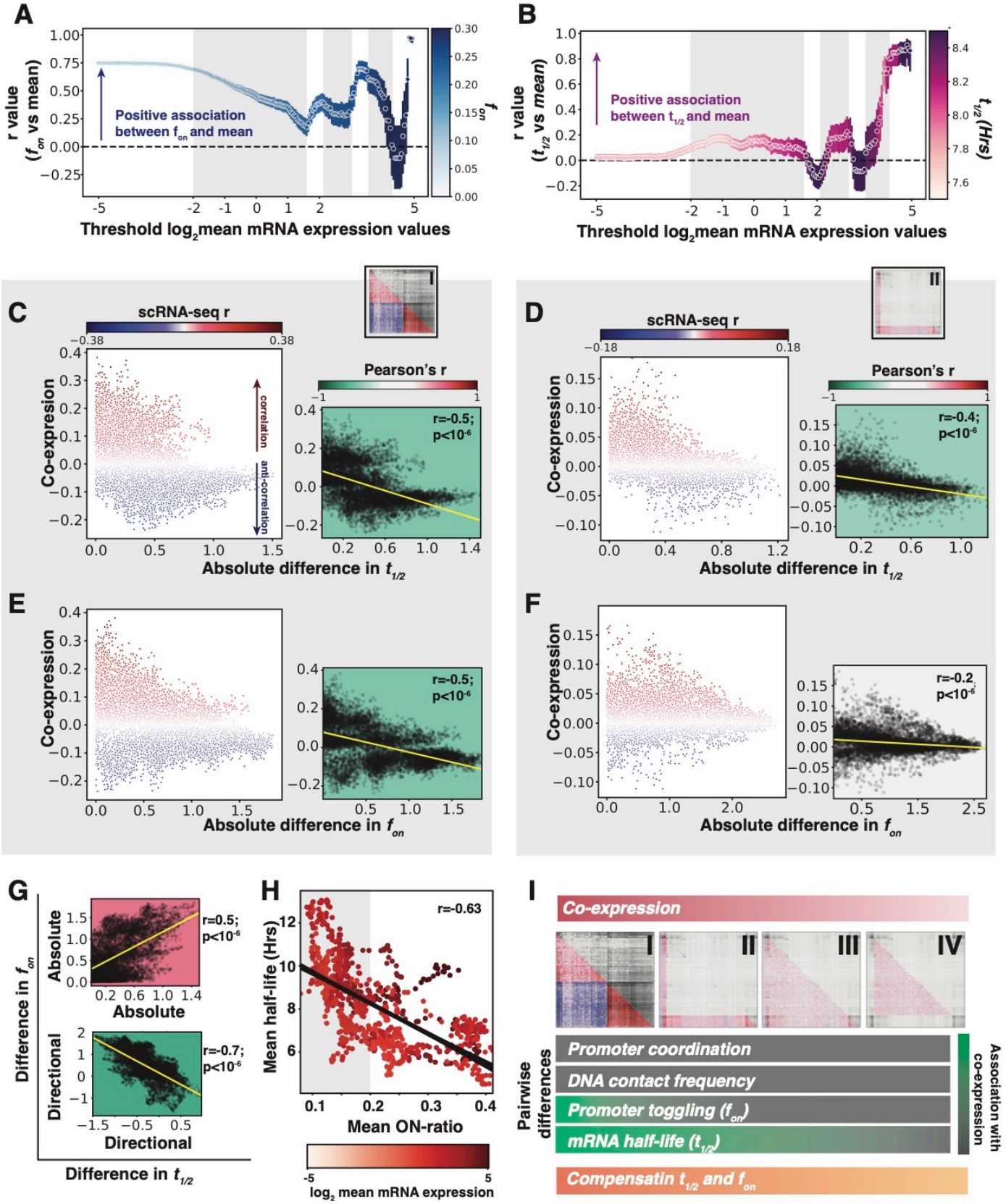
bioRxiv preprint doi: https://doi.org/10.1101/2022.09.21.508820; this version posted September 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 1.



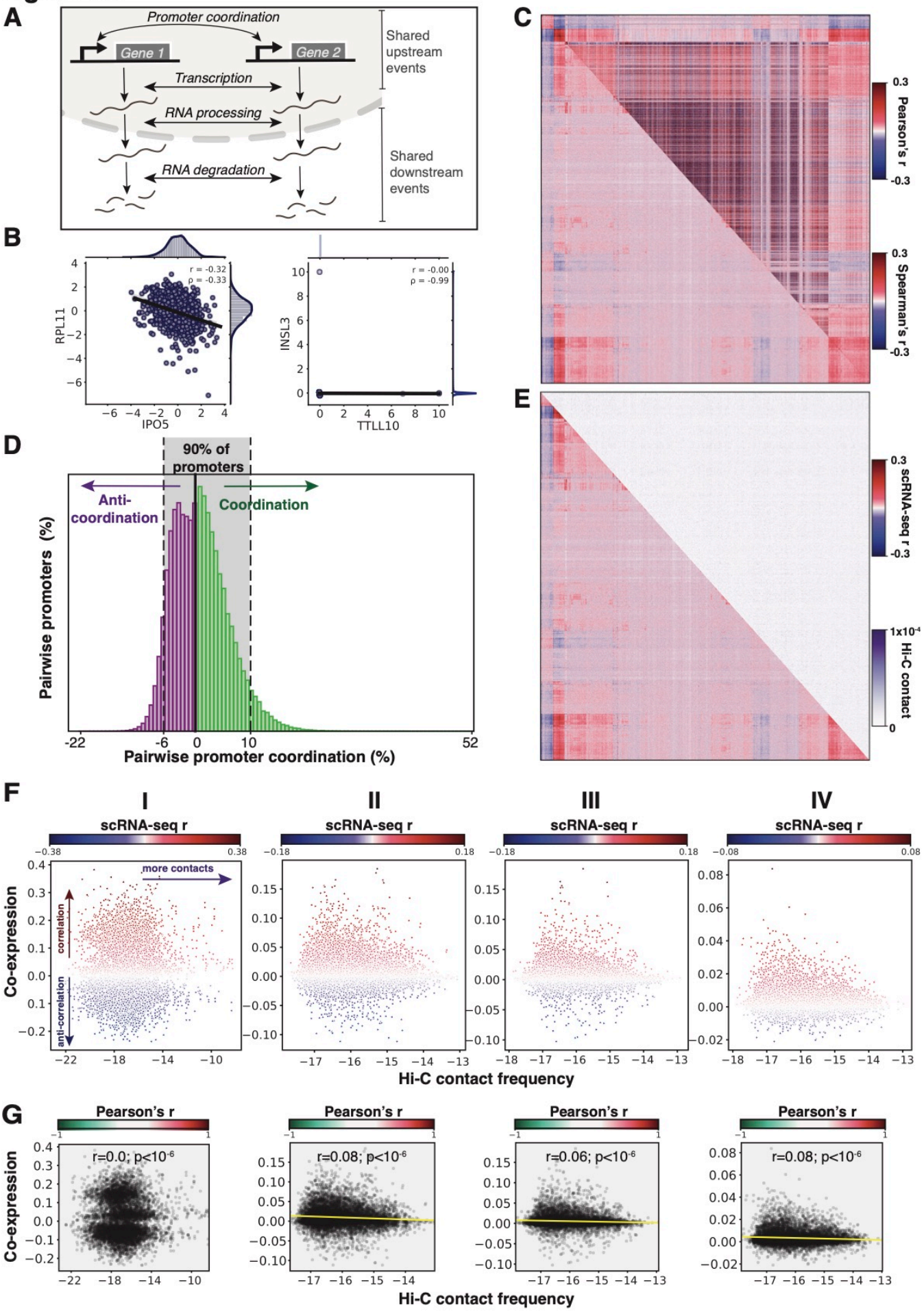
bioRxiv preprint doi: https://doi.org/10.1101/2022.09.21.508820; this version posted September 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure 2.



bioRxiv preprint doi: https://doi.org/10.1101/2022.09.21.508820; this version posted September 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

Figure S1.



bioRxiv preprint doi: https://doi.org/10.1101/2022.09.21.508820; this version posted September 22, 2022. The copyright holder for this preprint (which was not certified by peer review) is the author/funder. All rights reserved. No reuse allowed without permission.

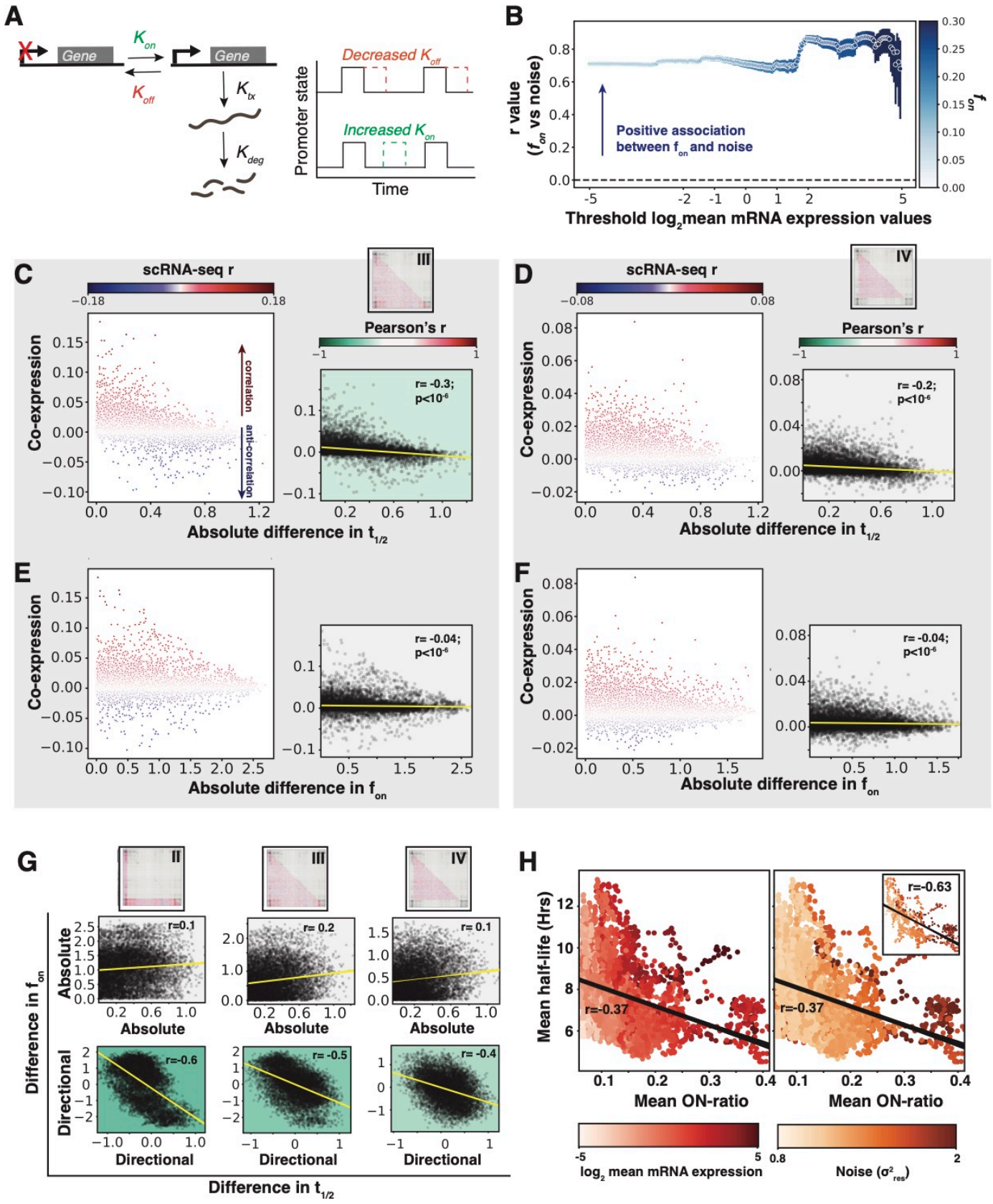



Figure S2.