

UCSF

UC San Francisco Electronic Theses and Dissertations

Title

Developments of Two Imaging based Technologies for Cell Biology Researches

Permalink

<https://escholarship.org/uc/item/3ph012sf>

Author

Yan, Xiaowei

Publication Date

2021

Peer reviewed|Thesis/dissertation

Developments of Two Imaging based Technologies for Cell Biology Researches

by
Xiaowei Yan

DISSERTATION
Submitted in partial satisfaction of the requirements for degree of
DOCTOR OF PHILOSOPHY

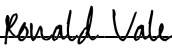
in

Biochemistry and Molecular Biology

in the

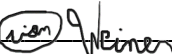
GRADUATE DIVISION
of the
UNIVERSITY OF CALIFORNIA, SAN FRANCISCO

Approved:

DocuSigned by:

DCB8EA4803FF4AA... Ronald Vale
Chair

DocuSigned by:

DC0B9413C4A0416... Jonathan Weissman

DocuSigned by:

DC0B9413C4A0416... Orion Weiner

Committee Members

Copyright 2021

By

Xiaowei Yan

DEDICATION

Everything happens for the best.

To my family, who supported me with all their love.

ACKNOWLEDGEMENTS

The greatest joy of my PhD has been joining UCSF, working and learning with such a fantastic group of scientists. I am extremely grateful for all the support and mentorship I received and would like to thank:

My mentor, Ron Vale, who is such a great and generous person. Thank you for showing me that science is so much fun and thank you for always giving me the freedom in pursuing my interest. I am grateful for all the guidance from you and thank you for always supporting me whenever I needed. You are a person full of wisdom, and I have been learning so much from you and your attitude to science, science community and even life will continue inspire me. Thank you for being my mentor and thank you for being such a great mentor.

Everyone else in Vale lab, past and present, for making our lab a sweet home. I would like to give my special thank to Marvin (Marvin Tanenbaum) and Nico (Nico Stuurman), two other mentors for me in the lab. I would like to thank them for helping me adapt to our lab, for all the valuable advice and for all the happiness during the time that we work together. Without Marvin, I wouldn't have the chance to dig into the wonderful single molecule world and I just cannot imagine my PhD life without Nico's microscopy expertise and his exceptional ability and willingness to teach. I would also like to thank Kara (Kara McKinley) and Jess (Jess Sheu-Gruttadauria) especially for their generous guidance and always support. You guys are so sweet and always there for me whenever I have scientific questions or life concerns. Especially Jess, thank you for taking the time to go through all my documents again and again, for my paper and for my application. Thank you for your each single sweet note and this valuable collaboration experience during my last period of time in the lab and I am really happy to have this chance to work with you. Moreover, I would also like to thank Nan (Nan Zhang), Soraya (Soraya Pedemonte)

and Phoebe (Phoebe Grigg) for all their effort in keeping the lab running; thank Courtney (Courtney Schroeder) for introducing me such a wonderful lab; thank Walter (Walter Huynh) for all your selfless support and each graduate student gather-together lunch; thank Kate (Kate Carbone) for guiding me through my early PhD years as an elder student in the lab and thank you Kate and Stefan (Stefan Niekamp) for being such great neighbors; thank Taylor (Taylor Skokan) for all your valuable advice and patience in reading/editing all my paper drafts; thank Ankur (Ankur Jain) for helping me adapt to the lab during my early rotation; thank Susana (Susana Ribeiro) for your generous and timely guidance even after you have left the lab; thank Adam (Adam Williamson) for your every single cheerful face and your passion; thank Yuxiao (Yuxiao Wang) for organizing the crosstalk subgroup which gave our non-dynein, non-signaling people a tiny home; thank Enfu (Enfu Hui) and Xiaolei (Xiaolei Su) for your guys' generous help whenever I needed and thank Rick (Rick McKenney), Gira (Gira Bhabha), Damian (Damian Ekiert), Jongmin (Jongmin Sung), Meghan (Meghan Morrissey), Iris (Iris Grossman), Nadja (Nadja Kern), Garrett (Garrett Greenan), Christina (Christina Gladkova), Zhen (Zhen Chen), Rui (Rui Dong), Patrick (Patrick Chitwood) for all your accompany, it is you guys who made my PhD journey so much more cheerful and memorable.

Outstanding scientists and my collaborators I worked with and learned from outside of the lab: Elizabeth Costa, Calvin Jan, Luke Gilbert, Max Horlbeck, Ryan Pak, Owen Chen, Marco Jost, Christina Liem from Weissman lab and Guiping Wang from Zhuang lab. Thank you for showing me science could be so much fun when working in a group and thank you for letting me know that interacting with others is actually easier and less scary than I initially thought.

My high school teacher Limin Chang and my collage teachers Ming Tian and Qing Li. Thank teacher Chang for fostering my love to biology, without you, I might not have the chance to step into this cheerful scientific world; thank Dr. Tian, who I received my earliest scientific

training, it is you who showed me what is true scientific research and your rigorous attitude to each single experiment and passion to science will continue motivate me; thank Dr. Li (Qing Li) for your support, mentorship and all the joy that we share.

My thesis committee members, Jonathan Weissman and Orion Weiner, who provided me valuable advice and guidance. I would also like to thank Jonathan for not only being my committee, but my mentor during my rotation and all the generous support and help I received afterwards.

My classmates, Aditya Anand, Allison Cohen, Athena Lin, Benjamin Barsi-Rhyne, Efren Reyes, Joseph Choe, Justin Salat, Karen Cheng, Karen Ruiz, Karina Perlaza, Kelsey Haas, Kyle Fowler, Nairi Hartooni, Nnejiuwa Ibe, Nick Sanchez, Rachel Greenstein, Sumitra Tatapudy and Valentina Garcia. Thank you all for accepting me and helping me adapt to this entirely different life. It is such a pleasure for me to have you guys as my classmates and every single minute with you is my unforgettable treasure.

My roommate Yan Zhang and friends : Zao Li, Xiaohua Hu, Huadong Wang, Wen Lu, Ran You, Tongchao Li, Yi Wang, Yu Bai, Qiunan Chen *etc.* I am so grateful to have such a group of incredible friends and thank you so much for all your accompany and support, for sharing great games and wonderful food. Thank you for filling my PhD journey with so many wonderful moments.

My family. Thank you mom (Ning Wang), dad (Xiongli Yan), grandma (Fengyan Zhou), grandpa (Shaorong Wang) and my aunt (Lin Wang) for all your love. I am really grateful for the freedom that I have been giving throughout my life, thank you for your understanding and support for whatever I decided to do.

My boyfriend, Hao Xu, for all your love and support. Even though you are not physically around, you are always there whenever I needed. Thank you for cheering me up whenever I am down and thank you for filling my life with joy.

STATEMENT REGARDING AUTHOR CONTRIBUTIONS

Chapter 2 of this dissertation contains partial reprints of previously published material:

Yan, X., T.A. Hoek, R.D. Vale, and M.E. Tanenbaum. 2016. Dynamics of Translation of Single mRNA Molecules In Vivo. *Cell*. 165:976–989. doi:10.1016/j.cell.2016.04.034.

Chapter 3 of this dissertation contains partial reprints of previously published material:

Yan, X., N. Stuurman, S.A. Ribeiro, M.E. Tanenbaum, M.A. Horlbeck, C.R. Liem, M. Jost, J.S. Weissman, and R.D. Vale. 2021. High-content imaging-based pooled CRISPR screens in mammalian cells. *J Cell Biol*. 220. doi:10.1083/jcb.202008158.

ABSTRACT

Developments of Two Imaging based Technologies for Cell Biology Researches

Xiaowei Yan

New technology developments foster new biology. Beginning from the early observation of cells by Robert Hooke and Anton van Leeuwenhoek, who figured out how to use optical lenses to achieve high magnification, microscopy has made huge advancement and each single discovery in cell biology has been accompanied with the development of microscopy innovations. Recent advances in microscopy have reached a previously unachievable resolution and acquisition speed, allowing people study the heterogenous cellular phenotypes accurately and efficiently. However, not all questions can be addressed with the existing approaches. For example, what is the single molecule basis of translation, how single ribosome translates; can we combine imaging based technologies with other existing approaches like CRISPR to further speed up cell biology discoveries? All of these questions require further effort to develop suitable imaging based technologies to address and this is what I aim to do during my PhD. In Chapter 2 and Chapter 3, I will introduce two newly developed imaging based technologies to address the above mentioned questions and together, these technologies provide new opportunities to further facilitate cell biology researches.

TABLE OF CONTENTS

CHAPTER 1	1
Introduction	1
CHAPTER 2	3
Dynamics of Translation of Single mRNA Molecules In Vivo.	3
Abstract	3
Introduction	4
Results	6
Discussion	17
Figures and figure legends	22
Supplemental figures and figure legends	31
Materials and Methods	41
Acknowledgements	54
CHAPTER 3	55
High-content imaging-based pooled CRISPR screens in mammalian cells.	55
Abstract	55
Introduction	56
Results	58
Discussion	66
Figures and figure legends	71
Supplemental figures and tables with legends	80
Materials and Methods	98
Acknowledgements	107
REFERENCES	108

LIST OF FIGURES

CHAPTER 2

Figure 2.1	22
Figure 2.2	24
Figure 2.3	25
Figure 2.4	27
Figure 2.5	28
Figure 2.6	29
Figure 2.7	30
Figure S2.1	31
Figure S2.2	33
Figure S2.3	35
Figure S2.4	37
Figure S2.5	38
Figure S2.6	39
Figure S2.7	40

CHAPTER 3

Figure 3.1	71
Figure 3.2	72
Figure 3.3	74
Figure 3.4	76
Figure 3.5	78
Figure S3.1	80
Figure S3.2	81

Figure S3.3	82
Figure S3.4	84
Figure S3.5	86
Figure S3.6	88

LIST OF TABLES

CHAPTER 3

Table S3.1	91
Table S3.2	94
Table S3.3	96
Table S3.4	97

CHAPTER 1

Introduction

Recent advances in single molecule technologies have reached an unachievable resolution, allowing people study the activity of single molecules. However, these technologies cannot be easily applied to study molecules *in vivo* since the signal from a single fluorophore is not sufficient to overcome the relative high background in real cellular environment. To overcome this, all sorts of signal enhancing approaches have been developed. For example, tandem labeling from MS2-MCP system was successfully applied to increase the signal coming from single mRNA which enables live cell single mRNA detection (Park et al., 2014). However, such system can only be utilized for nucleic acids, which cannot be directly applied for protein labeling. The development of SunTag system overcome this challenge, which enables single protein labeling in live cells (Tanenbaum et al., 2014). This is also our basis for *in vivo* detection of translation, which will be further described in Chapter 2.

During the time when single molecule technologies are rapidly developed, microscopy tools have also been used in combination with other technologies, such as CRISPR (clustered regularly interspaced short palindromic repeats) technology. CRISPR-based gene inactivation provides a powerful means for linking genes to particular phenotypes and microscopy further expands its application to morphological based phenotypes that can only be detected under microscopes. Benefit from the rapid development of automatic imaging acquisition and image analysis, arrayed based CRISPR optical screen have been widely applied to study previously inaccessible cellular phenotypes. However, this approach cannot be easily scaled up due to the utilization of the arrayed format. In addition, generation and maintenance of the screening library, set up and performing the screen itself is also time consuming. To overcome this, we decided to

develop a platform which enables CRISPR optical screen to be performed in a pooled format. This technology will be further described in Chapter 3, where we demonstrate our approach with a screen aiming for nuclear size regulators. In addition to the biological findings, I anticipate our screening approach can be widely applied to study different cellular morphological phenotypes, which could further advance cell biology researches.

CHAPTER 2

Dynamics of Translation of Single mRNA Molecules In Vivo

Abstract

Regulation of mRNA translation, the process by which ribosomes decode mRNAs into polypeptides, is used to tune cellular protein levels. Currently, methods for observing the complete process of translation from single mRNAs *in vivo* are unavailable. Here, we report the long-term (>1 hr) imaging of single mRNAs undergoing hundreds of rounds of translation in live cells, enabling quantitative measurements of ribosome initiation, elongation and stalling. This approach reveals a surprising heterogeneity in the translation of individual mRNAs within the same cell, including rapid and reversible transitions between a translating and non-translating state. Applying this method to the cell cycle gene *Emi1*, we find strong overall repression of translation initiation by specific 5'UTR sequences, but individual mRNA molecules in the same cell can exhibit dramatically different translational efficiencies. The ability to observe translation of single mRNA molecules in live cells provides a powerful tool to study translation regulation.

Introduction

Precise tuning of the expression of each gene in the genome is critical for many aspects of cell function. The level of gene expression is regulated at multiple distinct steps, including transcription, mRNA degradation and translation (Schwanhäusser et al., 2011). Regulation of all of these steps in gene expression is important, though the relative contribution of each control mechanism varies for different biological processes (Brar et al., 2012; Jovanovic et al., 2015; Peshkin et al., 2015; Tanenbaum et al., 2015; Vardy and Orr-Weaver, 2007).

Measuring the translation rate from individual mRNAs over time provides valuable information on the mechanisms of translation and translational regulation. *In vitro* experiments, mainly using bacterial ribosomes, have revealed exquisite information on ribosome translocation dynamics at the single molecule level (Blanchard, 2009; Chen et al., 2012; Cornish et al., 2008; Fei et al., 2008; Wen et al., 2008; Zaher and Green, 2009), but such methods have not yet been applied *in vivo*. In contrast, a genome-wide snapshot of the translational efficiency of endogenous mRNAs *in vivo* can be obtained through the method of ribosomal profiling (Ingolia et al., 2009, 2011). However, this method requires averaging of many cells and provides limited temporal information because of the requirement to lyse cells to make these measurements. Single cell imaging studies have succeeded in measuring average protein synthesis rates (Aakalu et al., 2001; Brittis et al., 2002; Han et al., 2014; Leung et al., 2006; Tanenbaum et al., 2015; Yu et al., 2006), observing the first translation event of an mRNA (Halstead et al., 2015), localizing sub-cellular sites of translation by co-localizing mRNAs and ribosomes (Katz et al., 2016; Wu et al., 2015), and staining nascent polypeptides with small molecule dyes (Rodriguez et al., 2006).

While ribosomal profiling and other recently developed methods have provided many important new insights into the regulation of translation, many questions cannot be addressed using current technologies. For example, it is unclear to what extent different mRNA molecules produced in a single cell from the same gene behave similarly. Many methods to study translation *in vivo* require averaging of many mRNAs, masking potential differences between individual

mRNA molecules. Such differences could arise from differential post-transcriptional regulation, such as nucleotide modifications (Choi et al., 2016; Wang et al., 2015), differential transcript lengths through use of alternative transcriptional start sites (Rojas-Duran and Gilbert, 2012) or polyadenylation site selection (Elkon et al., 2013; Gupta et al., 2014), differences in ribonucleic protein (RNP) composition (Wu et al., 2015), distinct intracellular localization (Hüttelmaier et al., 2005), or different states of RNA secondary structure (Babendure et al., 2006; Kertesz et al., 2010). Heterogeneity among mRNA molecules could have a profound impact on the total amount of polypeptide produced, as well as the localization of protein synthesis, but remains poorly studied. Furthermore, the extent to which translation on single mRNA molecules varies over time also is largely unknown. For example, translation may occur in bursts, rather than continuously (Tatavarty et al., 2012; Yu et al., 2006), and regulation of protein synthesis may occur by modulating burst size and/or frequency, which could occur either globally or on each mRNA molecule individually. In addition, the ability of an mRNA molecule to initiate translation may vary with time or spatial location, for example as cells progress through the cell cycle (Stumpf et al., 2013; Tanenbaum et al., 2015) or undergo active microtubule-based transport to particular cellular destinations (Holt and Schuman, 2013). Such regulation could involve changes in the rates of translation initiation and/or the ribosome elongation. To address these questions, new methods are required for visualizing translation on single mRNA molecules in live cells over time.

Here, we present a new method, based on the SunTag fluorescence tagging system that we recently developed (Tanenbaum et al., 2014), for measuring the translation of single mRNA molecules over long periods of time. Using this system, we have measured initiation, elongation and stalling on individual mRNA molecules and have uncovered unexpected heterogeneity among different mRNAs molecules encoded by the same gene within a single cell. Our system will be widely applicable to the study of mRNA translation in live cells.

Results

An assay for long term observation of translation of individual mRNAs

Observing the synthesis of a genetically-encoded fluorescent protein, such as GFP, *in vivo* is difficult because of the relatively long maturation time required to achieve a fluorescent state. Thus, a GFP-fusion protein typically will not fluoresce until after its translation is completed. To overcome this temporal challenge and to create a sufficiently bright signal to observe protein synthesis from single mRNAs *in vivo*, we used our recently developed SunTag system (Tanenbaum et al., 2014). In this assay, cells are co-transfected with a reporter transcript containing an array of 24 SunTag peptides followed by a gene of interest, along with a second construct expressing a GFP-tagged single-chain intracellular antibody (scFv-GFP) that binds to the SunTag peptide with high affinity. As the SunTag peptides are translated and emerge from the ribosome exit tunnel, they are rapidly bound by the soluble and already fluorescent scFv-GFP (**Fig. 2.1a**). Importantly, labeling of nascent chains using the SunTag antibody did not detectably alter protein synthesis rates of a reporter mRNA in human U2OS cells, as determined by FACS analysis (**Fig. 2.1b**). At the same time, the mRNA was fluorescently labeled by introducing 24 copies of a short hairpin sequence into the 3' UTR, and co-expressing the PP7 bacteriophage coat protein (Chao et al., 2008), which binds with high affinity to the hairpin sequence, fused to three copies of mCherry (PP7-mCherry) (**Fig. 2.1a**).

When observed by spinning disk confocal microscopy, the co-expression of a reporter construct (SunTag_{24x}-Kif18b-PP7_{24x}, with Kif18b being a kinesin motor with a 2.5 kb coding sequence (Tanenbaum et al., 2011)), scFv-GFP and PP7-mCherry resulted in the appearance of a small number (10-50) of very bright green and red fluorescent spots per cell that co-migrated in time-lapse movies (**Fig. 2.1c**). Spot tracking revealed that these spots diffused with a diffusion coefficient of 0.047 $\mu\text{m}^2/\text{s}$, which is slightly slower than previous measurements of mRNA diffusion (0.1 – 0.4 $\mu\text{m}^2/\text{s}$) (Katz et al., 2016), consistent with the fact that our reporter mRNA contains a larger open reading frame (4.4 kb versus 1.1 kb) and thus more associated ribosomes. In addition,

we observed many dim GFP spots that did not co-migrate with a mCherry signal in time-lapse movies. The bright spots rapidly disappeared upon terminating translation by addition of a protein synthesis inhibitor puromycin, which dissociates nascent polypeptides and ribosomes from mRNA (**Fig. 2.1c**), indicating that they are sites of active translation where multiple ribosomes are engaged on a single mRNA molecule. The dim spots were unaffected by puromycin treatment, suggesting that they represent individual, fully synthesized SunTag_{24x}-Kif18b proteins that had already been released from the ribosome. Thus, this translation imaging assay allows visualization of ongoing translation of single mRNA molecules.

Rapid 3-D diffusion of mRNAs makes it difficult to track single mRNAs for >1 min, as mRNAs continuously diffuse in and out of the z-plane of observation, and mRNAs regularly cross paths, complicating identification and tracking of individual mRNA molecules over time. To track mRNAs unambiguously for long periods of time, we added a CAAX sequence, a prenylation sequence that gets inserted into the inner leaflet of the plasma membrane, to the PP7-mCherry protein which served to tether mRNAs to the 2-D plane of the plasma membrane (**Fig. 2.1d and e**). As a result of many PP7-mCherry molecules clustering through their interaction with the multiple recognition sites on a single mRNA, bright red dots appeared on the plasma membrane at the bottom on cell, representing a tethered mRNA molecule (**Fig. 2.1e**). Tethered mRNA molecules co-migrated with scFv-GFP foci, indicating that they are sites of active translation (**Fig. 2.1e**). The membrane tethering of the mRNA had minimal effects on the protein expression of a GFP reporter construct as analyzed by FACS (**Fig. 2.1f**). While membrane tethering greatly improves the ability to visualize translation on single mRNA molecules over long periods of time and does not appear to grossly perturb mRNA translation, it is important to note that some aspects of translation, especially localized translation, may be altered due to tethering.

We first analyzed the PP7-mCherry spots observed on the plasma membrane to confirm that they contained only a single mRNA molecule. The fluorescence intensities of PP7-mCherry foci were very homogeneous (**Fig. S2.1a**); their absolute intensity was ~1.4-fold brighter, on

average, than single, membrane-tethered SunTag_{24x}-CAAX proteins bound with scFv-mCherry, which is expected to contain 24 mCherry molecules (**Fig. S2.1b**). PP7 binds as a dimer to the RNA hairpin and each PP7 was tagged with two tandem copies of mCherry. Thus mRNAs spots could be expected to be four times as bright as single scFv-mCherry-SunTag_{24x}-CAAX spots, but previous studies suggested that only about half of PP7 binding sites may be occupied (Wu et al., 2015); thus, mRNA spots would be about 2-fold brighter than single mCherry-SunTag_{24x} spots if they contain a single mRNA molecule, but 4- or more fold brighter if they contained 2 or more mRNAs. These results are therefore most consistent with the mCherry-PP7 foci being single mRNA molecules rather than multiple copies of mRNAs. Further supporting this idea, we tracked 63 single mRNA foci for 30-45 min and did not find a single case in which one spot split into two, which would have been indicative of more than one mRNA molecule being present in a single spot.

Because single mRNAs were tethered to the plasma membrane through multiple PP7 molecules, and thus through many CAAX membrane insertion domains, the 2-D diffusion of mRNAs was extremely slow ($1.06 \times 10^{-3} \mu\text{m}^2/\text{s}$, $n = 211$). This slow diffusion made it possible to track individual mRNAs and their associated translation sites for extended periods of time (mean tracking time >1 hr, **Fig. S2.1c**). Furthermore, the very slow diffusion rate of tethered mRNAs allowed us to image tethered translation sites using long exposure times (500-1000 ms); during this time interval, rapidly diffusing, non-tethered fully synthesized polypeptides only produced a blurred, diffuse image on the camera sensor, which enabled sites of translation to be easily distinguished from fully synthesized molecules (**Fig. S2.1d**). Finally, to confirm that the scFv-GFP was binding to nascent SunTag peptides, we replaced the SunTag epitope peptides in our reporter mRNA with an unrelated nucleotide sequence (encoding BFP) and found no GFP foci formation near mRNAs (**Fig. S2.1e**).

In conclusion, we have developed assays that enable both single mRNAs and their associated nascent translating polypeptides to be imaged over time. This general SunTag-based

method can be performed with either freely diffusing mRNAs or mRNAs tethered to the plasma membrane, each of which has unique advantages depending on the specific biological question. For further experiments in this study, we used the membrane-tethered system to follow translation for long periods of time.

Measurement of ribosome number, initiation rate, and elongation rate on single mRNAs

To estimate the number of ribosomes translating each mRNA, we compared the scFv-GFP fluorescence intensity of translation sites with that of the single, fully synthesized SunTag_{24x}-Kif18b molecules present in the same cell (**Fig. S2.2a and b**). Several considerations need to be taken into account to calculate ribosome number from the fluorescence intensities of translation sites and fully synthesized single SunTag proteins. First, ribosomes present at the 5' end of the reporter transcript have translated only a subset of the 24 SunTag peptides, so the nascent polypeptide associated with these ribosomes will have lower fluorescence intensity due to fewer bound scFv-GFPs. We generated a mathematical model to correct for the difference in fluorescence intensity for ribosomes at different positions along the transcript. Second, if scFv-GFP peptide has a slow on rate for the epitope *in vivo*, a lag time could exist between the synthesis of a SunTag peptide and binding of a scFv-GFP, which could result in the underestimation of the number of ribosomes per mRNA. To test this, cells were treated with the translation inhibitor cycloheximide (CHX), which blocks ribosome elongation by locking ribosomes on the mRNA and prevents the synthesis of new SunTag peptides, while allowing binding of scFv-GFP to existing peptides to reach equilibrium. The translation site scFv-GFP signal did not substantially increase after CHX treatment (**Fig. S2.2c**), indicating that under our experimental conditions, the lag time between peptide synthesis and scFv-GFP binding does not detectably affect translation-site intensity. Based on the above controls and our mathematical model, we could estimate the ribosome number per mRNA from the fluorescence intensity of the translation site. Approximately 30% of the mRNAs did not have a corresponding GFP signal, suggesting that they were not

actively translating. For the remaining 70% of the mRNAs that were translating, the majority (76%) had between 10–25 ribosomes (**Fig. 2.2a**), corresponding to an average inter-ribosome distance of ~200–400 nucleotides (nt). We also compared translation-site intensity of two additional reporter mRNAs with either 5x or 10x SunTag peptides with the 24x peptide reporter. This analysis revealed that ribosome density was very similar on the 5x and 10x reporter (1.26-fold and 1-fold, respectively) (**Fig. S2.2d**), indicating that the long 24x SunTag array does not grossly perturb ribosome loading on the reporter mRNA.

Next, we measured the translocation speed of ribosomes on single mRNAs by treating cells with harringtonine, a small molecule inhibitor of translation that stalls new ribosomes at the start of the mRNA coding sequence without affecting ribosomes further downstream (Ingolia et al., 2011). As mRNA-bound ribosomes complete translation one-by-one after harringtonine treatment, the GFP signal on mRNAs decreases (**Fig. 2.2b-d**). Using a simple mathematical model to fit the decay in fluorescence of a cumulative curve from many mRNAs (**Fig. S2.7**), we estimate a ribosome translocation rate of 3.5 ± 1.1 codons/s. In a parallel approach, we also measured the total time required for runoff of all ribosomes from individual mRNAs (**Fig. S2.2e**), from which we calculated a similar translation elongation rate (3.1 ± 0.14 codons/s) as the one obtained through our model. A reporter with only 5 instead of 24 SunTag peptides showed similar elongation kinetics (3.1 ± 0.4 codons/s) (**Fig. S2.2f**), indicating that translocation rates are likely not affected by SunTag labeling of the nascent chain. Finally, we measured elongation rates of a shorter and codon-optimized reporter gene, which revealed a somewhat faster elongation rate of 4.9 codons/s (**Fig. S2.2g**), indicating that elongation rates may differ on different transcripts. Using the elongation rate and ribosome density described above, we were able to estimate the translation initiation rate to be between $1.4\text{--}3.6 \text{ min}^{-1}$ on the Kif18b reporter.

Together, these results provide the first in vivo measurements of the rates of ribosome initiation and translocation on single mRNA molecules in live cells.

Temporal changes in translation of single mRNA molecules

To study translation over time, we imaged cells for 2 hr and quantified the scFv-GFP signal from single mRNA molecules that could be tracked for >1 hr (**Fig. 2.3a, b and Fig. S2.3a**). The results show considerable fluctuations in the translational state of individual mRNAs over time (**Fig. 2.3a, b and Fig. S2.3a**). Such large fluctuations were not observed when cells were treated with the translation inhibitor CHX (**Fig. S2.3b**), indicating they were due to changes in translation initiation and/or elongation rather than measurement noise. We also observed heterogeneity of behavior between different mRNAs. Some remained in a high translating state for >1 hr (e.g., **Fig. S2.3a 12 and 13**). Others shut down translation initiation and lost their scFv-GFP signal (e.g., **Fig. 2.3a, b and Fig. S2.3a 1, 3-11, and 14**), which may account for the population of non-translating mRNAs observed in steady-state measurements (**Fig. 2.2a**). From the progressive decline in scFv-GFP fluorescence (**Fig. 2.3c**), we could estimate a ribosome run-off rate of 3 codons/s (**Fig. 2.3c**), which is similar to that measured after addition of harringtonine (3.5 ± 1.1 codons/s) (**Fig. 2.2**). Interestingly, a subset (67 of 104 mRNAs, three independent experiments, 19 cells) of these mRNAs later reinitiated translation and largely recovered their original scFv-GFP fluorescence (**Fig. 2.3a, b, d and Fig. S2.3a 1, 3, 5, and 8-10**). Individual mRNAs even showed repeated cycling between non-translating and translating states (**Fig. 2.3a, yellow line, and Fig. S2.3a 3, 5 and 8**). Such cycles of complete translational shutdown and re-initiation occurred 0.29 ± 0.10 times per mRNA per hour ($n = 4$ independent experiments, 27 cells, 106 mRNAs), suggesting that most mRNAs will undergo one or more translational shutdown and re-initiation events in their lifetime. Thus, single mRNA imaging reveals reversible switching between translational shutdown and polysome formation.

After synchronized expression of the reporter construct using an inducible promoter, we often observed the initial binding events of newly transcribed mRNAs to the PP7-mCherry at the membrane (**Fig. 2.4a and b**). Of these initial binding events, 44% of the mRNAs were associated with scFv-GFP fluorescence, indicating that they had already begun translation. However, the

majority, 56% of mRNAs, initially appeared at the membrane in a non-translating state and subsequently converted to a translating state, usually within 1–5 min (**Fig. 2.4c**). These mRNAs are likely newly transcribed mRNAs that are translating for the first time, rather than mRNAs that have already undergone translation but transitioned temporarily to a non-translating state. In support of this argument, long-term (>1 hr) imaging of single mRNAs reveals that mRNAs spend on average only 2.5% of their lifetime in such a temporary non-translating state (n = 4 independent experiments, 27 cells, 106 mRNAs), which is not sufficient to explain the 56% non-translating mRNAs that appeared at the membrane after synchronized transcription of the reporter. Rapid initiation of translation on newly transcribed mRNAs was described recently (Halstead et al., 2015), but our assay additionally allows an analysis of polysome buildup on new mRNAs (**Fig. 2.4b**). Our analysis of the increase in scFv-GFP fluorescence indicates that, once the first ribosome begins chain elongation, additional ribosomes initiate translation with a rate indistinguishable from that on polysomes at steady state. We also examined the rate of fluorescence recovery (corresponding to polysome buildup) after complete shutdown of translation and subsequent re-initiation (**Fig. 2.4d**). The polysome buildup on new transcripts was comparable to that observed for mRNAs that were cycling between translating and non-translating states (**Fig. 2.4d**).

Ribosome stalling

Several studies reported that ribosomes can pause or stall at a defined nucleic acid sequence with a regulatory function (Walter and Blobel, 1981; Yanagitani et al., 2011), at chemically modified or damaged nucleotides (Simms et al., 2014), or at regions in the RNA with a strong secondary structure (Tholstrup et al., 2012; Wen et al., 2008). We found that a subset (~5%–10%) of mRNAs retained a bright scFv-GFP signal 15 min after harringtonine treatment (**Fig. 2.2b and d**), a time at which ribosomes translocating at ~3 codons/s should have finished translating the reporter. A similar percentage of stalled ribosomes was observed on two additional reporter transcripts, both of which were designed using optimal codon usage (**Fig. S2.2g and**

S2.4a). Ribosome stalling also was observed using hippuristanol (**Fig. S2.4b**), a translation initiation inhibitor with a different mechanism of inhibition (Bordeleau et al., 2006), indicating that the stalling was not caused by harringtonine. We also observed stalls when examining ribosome runoff from non-tethered cytosolic mRNAs lacking PP7 binding sites (**Fig. S2.4c**). Importantly, stalls were not observed after puromycin treatment (**Fig. S2.4d and e**) and the prolonged (>15 min) scFv-GFP signal on mRNAs from harringtonine-treated cells rapidly disappeared upon the addition of puromycin, confirming that the observed signal indeed represents stalled ribosomes (**Fig. S2.4f**). The majority of mRNAs with stalled ribosomes (33 of 43) could be tracked for >40 min, the typical duration of our harringtonine runoff experiments, indicating that they were not readily targeted by the no-go mRNA decay machinery within this time frame.

Ribosome stalls could be due to defective ribosomes causing roadblocks on the mRNA or due to defects in the mRNA. These models can potentially be distinguished by examining how such stalls are resolved. A single defective ribosome will inhibit ribosome runoff until the stalled ribosome is removed, after which, the remaining ribosomes will run off at a normal rate. In contrast, if the stalls are caused by defects to the mRNA, such as chemical damage, then each ribosome passing over the damaged nucleotide will be delayed, resulting in an overall slower scFv-GFP decay rate (**Fig. 2.5a**). Long-term tracking of stalled ribosomes on single mRNAs was consistent with the latter model, indicating that ribosome stalling is likely caused by defective mRNA (**Fig. 2.5b**). Consistent with the hypothesis that chemical damage to mRNA causes ribosome stalling, treatment of cells with 4-nitroquilone-1-oxide (4NQO), a potent nucleic-acid-damaging agent that causes 8-oxoguanine modifications and stalls ribosomes *in vitro* (Simms et al., 2014), resulted in a slow runoff on the majority of mRNAs, indicating widespread ribosome stalling (**Fig. 2.5c**). Thus, chemical damage to mRNAs stalls ribosome elongation *in vivo*.

Regulated ribosome pausing occurs both *in vitro* and *in vivo* at asparagine 256 in the stress-related transcription factor Xbp1 (Ingolia et al., 2011; Yanagitani et al., 2011), and this ribosome pausing is important for membrane targeting of the mRNA (Yanagitani et al., 2011). To

test whether our translation imaging system could recapitulate such translation pausing, we introduced a strong ribosome-pausing sequence (a point mutant of the wild-type Xbp1-pausing sequence that shows enhanced ribosome pausing (Yanagitani et al., 2011)) into the 3' region of the coding sequence of our reporter (hereafter referred to as Xbp1 reporter). Harringtonine ribosome runoff experiments on the Xbp1 reporter revealed a delay in ribosome runoff (**Fig. 2.5d**), confirming that our reporter faithfully reproduced the ribosome-pausing phenotype. To study the behavior of individual ribosomes on the Xbp1 ribosome-pausing sequence, we tracked single mRNAs during ribosome runoff. Surprisingly, the fluorescence decay was not linear, as would be expected if each ribosome paused a similar amount of time on the pause site. Rather, fluorescence decay occurred in bursts interspaced with periods in which no decay was detectable (**Fig. 2.4 and 2.5e**, representative traces shown out of 25 analyzed). These results indicate that most ribosomes are only briefly delayed at the Xbp1 pause site, but a small subset of ribosomes remain stalled for an extended (>10 min) period of time, explaining the strong ribosome stalling phenotype observed in ensemble experiments.

Translational regulation of the cell-cycle regulator Emi1

We also applied our assay to study the transcript-specific translational regulation of Emi1, a key cell-cycle regulatory protein. Our recent work reported strong translational repression of Emi1 during mitosis and found that the 3' UTR of Emi1 is involved in this regulation (Tanenbaum et al., 2015), but a role of its 5' UTR in translational regulation was not established. Interestingly, Emi1 has at least two splicing isoforms that differ in their 5' UTR sequence: NM_001142522.1 (hereafter referred to as 5' UTR_long) and NM_012177.3 (hereafter referred to as 5' UTR_short) (**Fig. 2.6a**). We found that a GFP protein fused downstream of the 5' UTR_long was expressed at 40-fold lower levels than a GFP fused to the 5' UTR_short (**Fig. 2.6b**). Such difference in protein expression could be due to a difference in transcription rate, mRNA stability, or reduced translation initiation or elongation rates. To distinguish between these possibilities, we prepared

translation reporter constructs bearing either the short or long 5' UTR of Emi1. Robust translation was observed on ~50% of mRNAs encoding the short 5' UTR (**Fig. 2.6c**). In contrast, the majority (~80%) of transcripts encoding the Emi1 5' UTR_long showed no detectable translation (not shown), and of the translating mRNAs, only very weak scFv-GFP fluorescence was usually detected (**Fig. 2.6c**). Surprisingly, however, a very small fraction of mRNAs containing the 5' UTR_long (~2%) was associated with a bright scFv-GFP signal (**Fig. 2.6c**, >92 bin), indicating that they are bound to many ribosomes. This was not due to ribosome stalling and subsequent (slow) accumulation of ribosomes on a subset of mRNAs, as this bright scFv-GFP signal rapidly dissipated upon harringtonine treatment (**Fig. S2.5**), indicating that these mRNAs were translated at high levels. Calculation of the total number of ribosomes associated with the mRNAs, based upon scFv-GFP fluorescence intensity, revealed that 52% of all ribosomes translating the Emi1 5' UTR_long reporter were associated with the minor (2%) fraction associated with the highest scFv-GFP intensity. These results indicate that the great majority of 5' UTR_long transcripts are strongly translationally repressed but that a small subset of these mRNAs escape repression and undergo robust translation. Thus, substantial heterogeneity in translational efficiency can exist among different mRNA molecules within the same cell.

Observation of translation by single ribosomes

Interestingly, with the Emi1 5' UTR_long reporter, we often observed the abrupt appearance of a weak scFv-GFP signal on a transcript that was previously translationally silent. The GFP signal initially increased over time, plateaued, and then was abruptly lost after 6–8 min (**Fig. 2.7a-c**). This type of signal is best explained by a single ribosome sequentially decoding the 24 SunTag peptides on the mRNA, followed by the release of the newly synthesized polypeptide upon completion of translation. Consistent with this hypothesis, the absolute fluorescence intensity of such translation events at the plateau phase (when all 24 SunTag peptides have been synthesized) was very similar to the intensity of a single fully synthesized SunTag_{24x}-Kif18b

protein (**Fig. S2.6a and b**). The duration of the scFv-GFP signal per translation event could be converted to a translocation speed of single ribosomes, which revealed an average elongation rate of 3 codons/s (**Fig. 2.7d**). This value is similar to that determined from our bulk measurements of harringtonine-induced ribosome runoff or natural translational initiation shutdown and runoff (3–3.5 codons/s), indicating that ribosome elongation was not affected by the Emi1 5' UTR_long. Comparison of translocation rates obtained from single ribosome translation events also revealed heterogeneity in the decoding speed of individual ribosomes *in vivo* (**Fig. 2.7d**).

Discussion

Using the SunTag system, we have developed an imaging method that measures the translation of individual mRNAs in living cells. Immobilization of mRNAs on the plasma membrane allows the long-term (>1 hr) observation of translation of single mRNA molecules, which enables analyses of translational initiation, elongation, and stalling in live cells for the first time. Under conditions of infrequent translational initiation, we can even observe a single ribosome decoding an entire mRNA molecule. Our observations reveal considerable and unexpected heterogeneity in the translation properties of different mRNA molecules derived from the same gene in a single cell, with some not translating, others actively translating with many ribosomes, and others bound to stalled ribosomes. The SunTag translation imaging assay should be applicable to many different cell types, including neurons and embryos, in which the localization and control of protein translation is thought to play an important role in cell function.

Comparison of methods to study translation in vivo

Ribosome profiling, a method in which fragments of mRNAs that are protected by the ribosome are analyzed by deep sequencing (Ingolia et al., 2009), has found widespread use in measuring translation. The strength of ribosomal profiling lies in its ability to measure translation on a genome-wide scale of endogenous mRNAs. However, a limitation of ribosome profiling is the need to pool mRNAs from many thousands of cells for a single measurement. Thus, ribosome profiling in its present form cannot be used to study translation heterogeneity between different cells in a population or among different mRNA molecules in the same cell. Furthermore, since ribosome profiling requires cell lysis, only a single measurement can be made for each sample, limiting studies of temporal changes.

A number of single-cell translation reporters have been developed based on fluorescent proteins (Aakalu et al., 2001; Brittis et al., 2002; Han et al., 2014; Raab-Graham et al., 2006; Tanenbaum et al., 2015; Tatavarty et al., 2012; Yu et al., 2006). Such reporters generally rely on

the accumulation of new fluorescence after the assay is initiated. Advantages of these systems are that they are generally easy to use and have single-cell sensitivity. However, they do not provide single mRNA resolution, often do not allow continuous measurement of translation, and do not report on ribosome initiation and elongation rates.

Finally, two methods were developed recently to image translation on single mRNAs *in vivo*. In one approach, the first round of translation is visualized (Halstead et al., 2015). This method, however, does not allow continuous measurements of translation. The second approach involves measurements of the number of ribosomes bound to an mRNA using fluorescence fluctuation spectroscopy (Wu et al., 2015). The advantage of this method is that it can detect binding of a single fluorescent protein to an mRNA and different subcellular sites can be probed to study spatial differences in translation. The limitation of this method though is the inability to follow translation of single mRNAs over time, as these mRNAs cannot be tracked in the cell.

SunTag-based translation imaging assays are unique thus far in their ability to follow translation of individual mRNAs over time. This translation assay can be employed with either freely diffusing or tethered mRNAs, the choice of which will depend on the biological question to be addressed. In the study by Wang et al. (2016) (Wang et al., 2016), translation is observed in distinct spatial compartments in neurons using a similar SunTag-based translation imaging method with non-tethered mRNAs. In contrast, for studying ribosome translocation dynamics, the tethering assay provides the ability to track a single mRNA throughout the duration of the ribosome elongation cycle. Using this assay, we could measure polysome buildup rates over time, observe mRNAs cycling between translating and non-translating states, uncover heterogeneity in translation initiation rates (*e.g.*, with the Emi1 5' UTR) and even observe a single ribosome translating an entire transcript. These measurements were aided by the vastly improved signal-to-noise of the tethered assay and the ability to easily track slowly diffusing tethered mRNAs for an hour or more. These long-term observations allowed us to discover that mRNAs can reversibly switch between a translating and non-translating state and have a high variability in pause

duration at the Xbp1 site. Thus, the untethered and tethered SunTag assays provide means to study translation of single mRNA molecules, which will be applicable to a wide variety of biological questions and will be complementary to existing methods of studying translation.

A drawback of our assay is the need to insert an array of SunTag peptide repeats into the mRNA of interest to fluorescently label the nascent polypeptide and the need to insert an array of PP7 binding sites in the 3' UTR to label the mRNA. As is true of any tagging strategy, these modifications could interfere with translation and/or mRNA stability under certain conditions. We have performed a number of control experiments to ensure that binding the scFv-GFP to the nascent chain and tethering of the transcript to the membrane do not grossly perturb translation (**Fig. 2.1b and f**). We have also shown that ribosome translocation rates and ribosome density are similar when using a reporter with a very short (5x) or long (24x) SunTag peptide array and comparing tethered and non-tethered mRNAs (**Fig. 2.2d, S2.2f and S2.4c**), indicating that many aspects of translation are not perturbed in our assay. Nevertheless, tethering of certain mRNAs to the plasma membrane may influence translation, especially for those mRNAs that undergo local translation in a specific compartment of the cell. Thus, our assay has unique advantages for certain types of measurements of translation, but appropriate controls should be performed for each experimental system or objective.

Heterogeneity in translation of single mRNAs: possible molecular mechanisms

Using our system, we measured the ribosome translocation speed on single mRNA molecules. Ribosome translocation rates have been measured in bulk previously in mouse embryonic stem cells (Ingolia et al., 2011), which yielded a translocation rate of 5.6 codons/s. Our values of 3–5 codons/s (**Fig. S2.7**) are in general agreement with those published values and very similar to those measured by Wang et al. (2016) (4 codons/s). Our experiments, and those of Wang et al. (2016), are the first to measure ribosome translocation rates for a single mRNA

species, in single cells and on single mRNAs, which provides new opportunities to study regulation of translation elongation.

We also found that translation initiation can shut down temporarily on individual mRNAs and rapidly restart (**Fig. 2.3**). Such shutdown of translation initiation could be due to transient loss of eIF4E binding to the mRNA cap, mRNA decapping followed by recapping (Mukherjee et al., 2012), or transient binding of regulatory proteins. Using our mRNA tethering assay, binding and unbinding of single proteins to translating mRNA could potentially be observed using total internal reflection fluorescence (TIRF), which could open up many additional possibilities for studying translational regulation at the single-molecule level.

The pioneer round of translation, the first ribosome to initiate translation on a newly transcribed mRNA, may be especially important, as it is thought to detect defects in the mRNA, including premature stop codons (Ishigaki et al., 2001). A recently developed translation biosensor can detect the location of this pioneer round of translation (Halstead et al., 2015). However, what happens after the first ribosome initiates translation is unknown. We found that the translation initiation rate on our reporter mRNA was similar on newly transcribed, recently shut down, and re-initiating mRNAs and polysomal mRNAs (**Fig. 2.4**), indicating that the initiation rate is independent of the number of ribosomes bound to the mRNA. The presence of introns in a gene may also affect translation initiation on newly transcribed mRNAs (Hir et al., 2016), which could be tested in future studies.

A subset of ribosomes stall on mRNAs in a sequence-independent fashion (**Fig. 2.2d, S2.2g and S2.4a**). One possible explanation for this is that ribosome stalling is caused by naturally occurring mRNA “damage” (*i.e.*, chemical modifications of the nucleotides). Previous studies have found that the 8-oxoguanine modification occurs on mRNA *in vivo*, and such modifications cause ribosome stalling *in vitro* (Simms et al., 2014) and *in vivo* (**Fig. 2.5c**). Alternatively, while we have performed numerous control experiments (**Fig. 2.5 and S2.4**), we cannot completely exclude that the observed stalling on a small subset of mRNAs is an artifact of

our construct or assay. We also observe ribosome pausing in a sequence-dependent fashion on the pause site of the Xbp1 transcription factor. Such pausing had been observed previously in bulk measurements (Ingolia et al., 2011; Yanagitani et al., 2011), but our quantitative analysis of single mRNAs revealed a high degree of variability in ribosome pausing at this site.

Finally, we show that the 5' UTR sequence of one Emi1 transcript isoform severely inhibits translation initiation. A likely explanation for this effect is the presence of several upstream open reading frames (uORFs) in this sequence. Surprisingly, a small number of mRNA molecules encoding this 5' UTR do undergo high levels of translation. It is possible that highly translating mRNAs are generated through alternative downstream transcription start site selection, which generates an mRNA that lacks the repressive sequence (for example, the uORFs). Alternatively, translation could occur if the 5' UTR repressive sequence is cleaved off, followed by recapping after transcription, if a repressive protein factor dissociates, or if an inhibitory RNA secondary structure unfolds. Further studies will be required to distinguish between these possibilities.

In summary, here we have developed an imaging method that enables the measurement of ribosome initiation and translocation rates on single mRNA molecules in live cells. Future developments of this technology could include simultaneous observation of single translation factors or other regulatory molecules together with mRNAs and nascent polypeptides, which would provide a very powerful system to dissect the molecular mechanisms of translational control.

Figure and figure legends

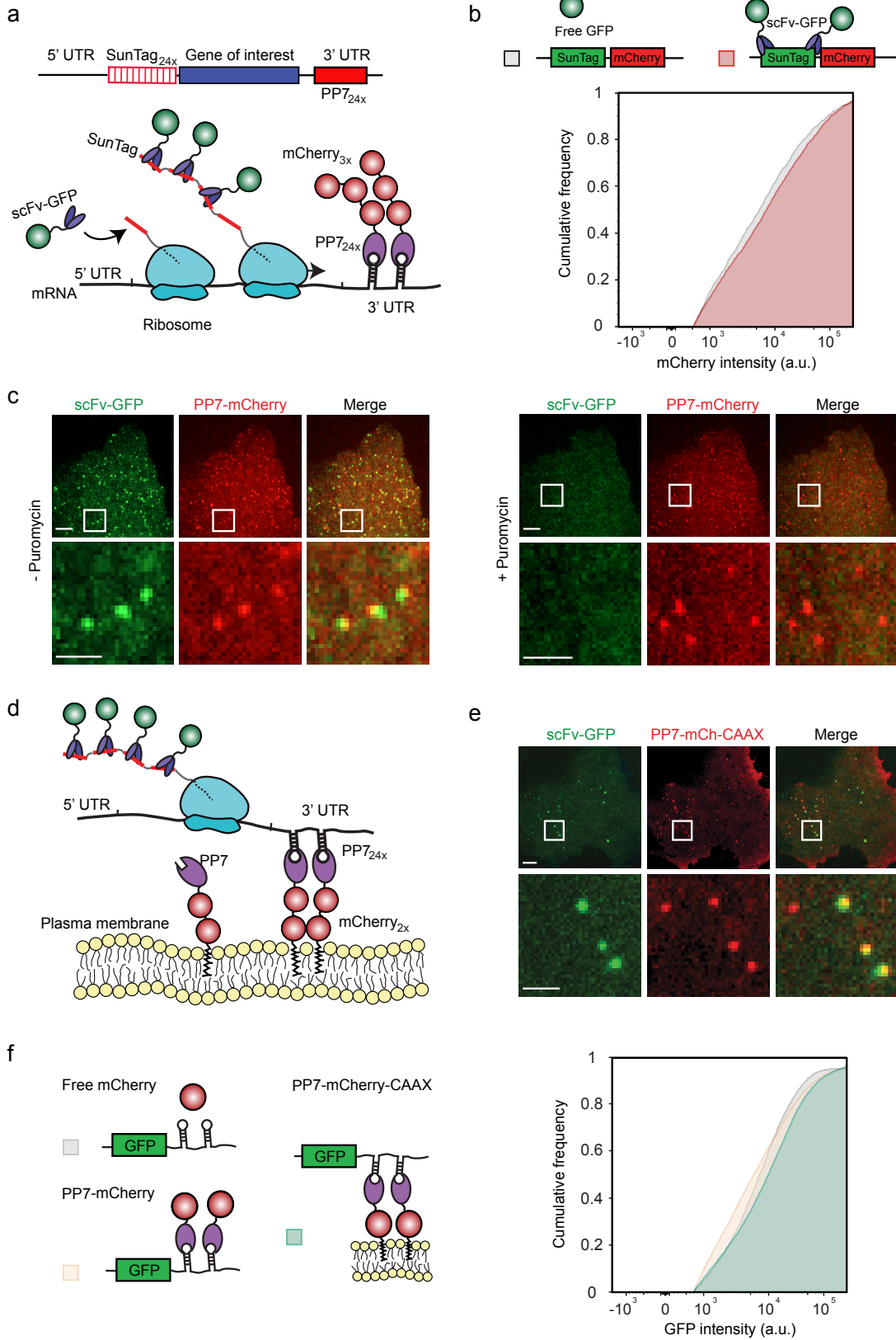


Figure 2.1 | Fluorescence labeling of nascent chains to visualize translation of single mRNA molecules. (a) Schematic of nascent polypeptide labeling using the SunTag system and mRNA labeling (a) and membrane tethering (d) using the PP7 system. (b) A mCherry-SunTag_{24x} reporter gene was co-transfected with either GFP or scFv-GFP, and the expression of the SunTag_{24x}-mCherry reporter was determined by FACS. Binding of the scFv-GFP to the SunTag nascent chain did not detectably alter protein expression. (c) A representative U2OS cell is shown expressing scFv-GFP, PP7-3xmCherry, and the translation reporter (SunTag_{24x}-Kif18b-PP7_{24x}). Cytosolic translation sites (scFv-GFP) co-localize with mRNAs (PP7-3xmCherry). Ribosomes were dissociated from mRNA by addition of puromycin (right panel). Note that translation sites and mRNA do not perfectly overlap because of the brief time difference in acquiring GFP and mCherry images. (d) Schematic of nascent polypeptide labeling and membrane tethering of the mRNA using the PP7 system. (e) U2OS cells expressing scFv-GFP (green), PP7-2xmCherry-CAAX (red), and the translation reporter (SunTag_{24x}-Kif18b-PP7_{24x}). A single time point of the cell (top panel) and a zoomed-in view from the white-boxed area containing a few mRNAs (lower) are shown. (f) U2OS cells were transfected with mCherry, PP7-mCherry, or PP7-mCherry-CAAX together with a GFP reporter transcript with 24 PP7 binding sites in the 3' UTR, and GFP expression was analyzed by FACS. Cumulative distribution of GFP expression levels from GFP-mCherry double positive cells are shown in (b) and (f) (n = 3 independent experiments). Scale bars, 5 mm (upper) and 2 mm (lower).

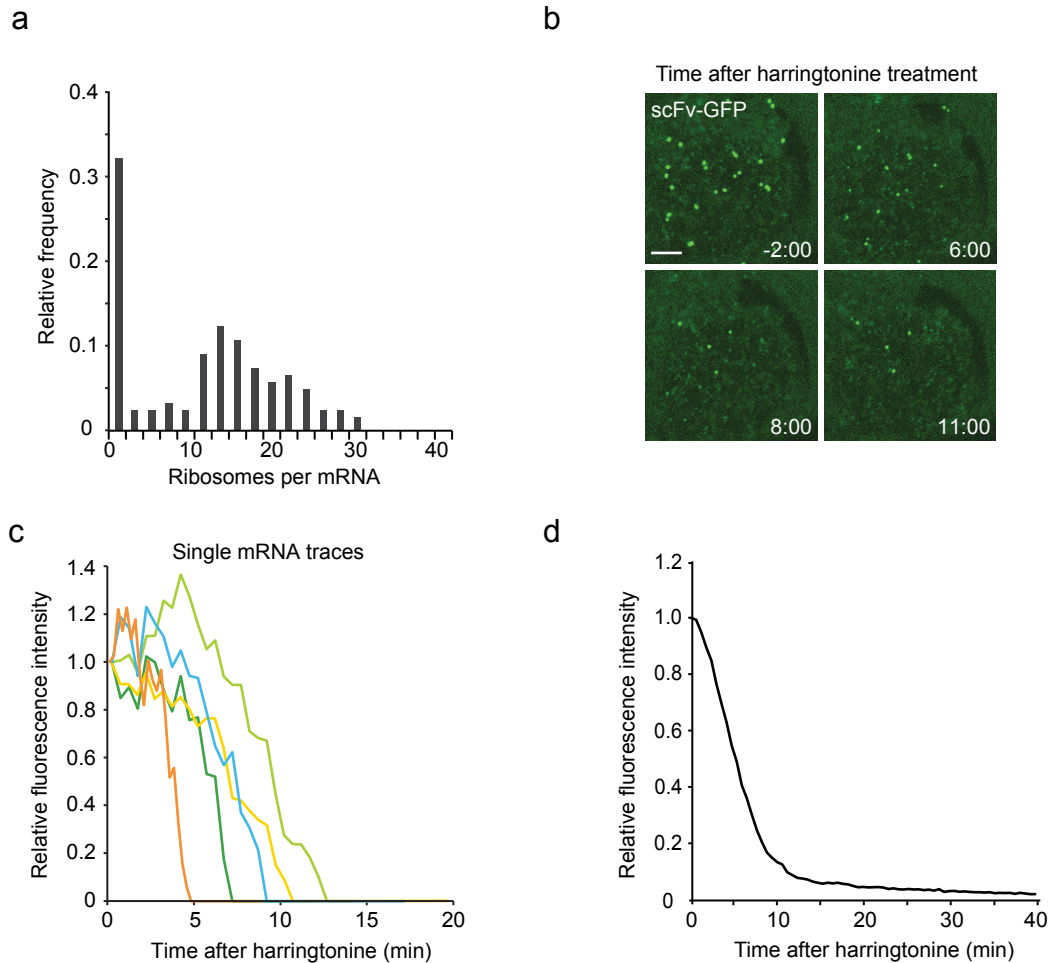


Figure 2.2 | Measurements of ribosome initiation and elongation rates on single mRNA molecules. U2OS cells expressing scFv-GFP, PP7-2xmCherry-CAAX, and the translation reporter (SunTag_{24x}-Kif18b-PP7_{24x}). (a) Distribution of the number of ribosomes bound to single mRNAs of the translation reporter (SunTag_{24x}-Kif18b-PP7_{24x}) (n = 2 independent experiments, 16 cells, 124 mRNAs). (b-d) U2OS cells expressing the translation reporter (SunTag_{24x}-Kif18b-PP7_{24x}) were treated with harringtonine at t = 0. (b) Representative images from a time-lapse movie. (c) Five representative traces of fluorescence decay on single mRNAs (of >100 analyzed). (d) Normalized quantification of the decrease in fluorescence over time from many translation sites (n = 4 independent experiments, 37 cells, 536 mRNAs). Scale bars, 5 μm.

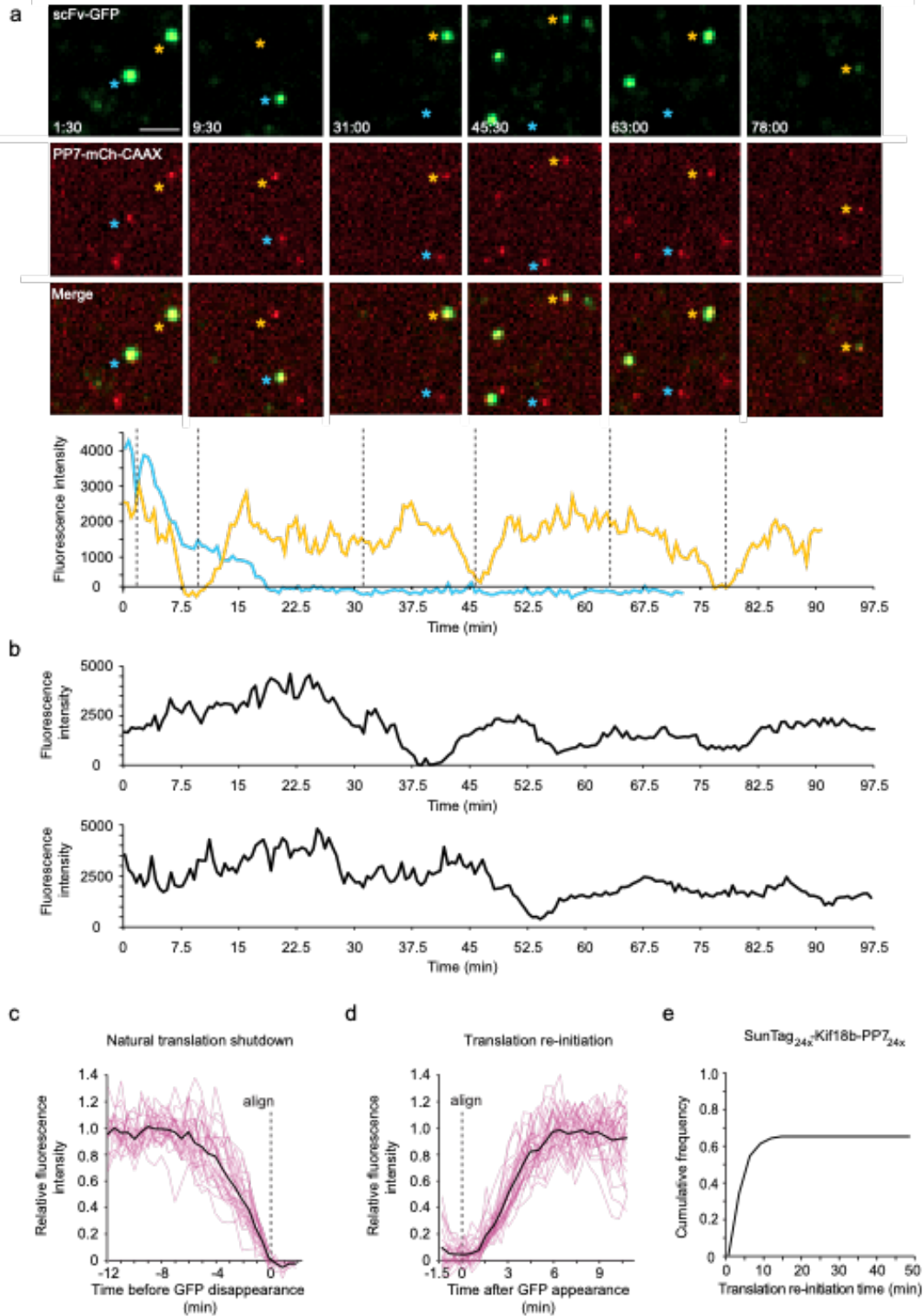


Figure 2.3 | Long-term dynamics of translation of single mRNA molecules. U2OS cells expressing scFv-GFP, PP7-2xmCherry-CAAX, and the translation reporter (SunTag_{24x}-Kif18b-PP7_{24x}). (a) U2OS cell expressing the SunTag_{24x}-Kif18b-PP7_{24x} reporter was imaged by time-lapse microscopy. Blue and yellow asterisks mark two different mRNAs undergoing changes in translation over time (upper). Intensity of scFv-GFP was measured over time for the two mRNAs

(lower). Colors of lines correspond to scFv-GFP intensity of translation sites marked by asterisk with the same color. (b) ScFv-GFP intensity traces of two additional mRNA molecules. (c) mRNAs undergoing permanent translation shutdown. Fluorescence intensity quantification is shown (n = 24 mRNAs). Average (black line) and single traces (pink lines) are shown. (d) mRNAs undergoing translation re-activation after shutdown. Average (black line) and single traces (pink lines) are shown (n = 30 mRNAs). (e) Time to reappearance of the first scFv-GFP fluorescence from translation sites that underwent complete translational shutdown. ~60% of the mRNAs re-initiated translation after complete shutdown and did so within 10 min (n = 104 translational sites analyzed). Scale bar, 2 μ m.

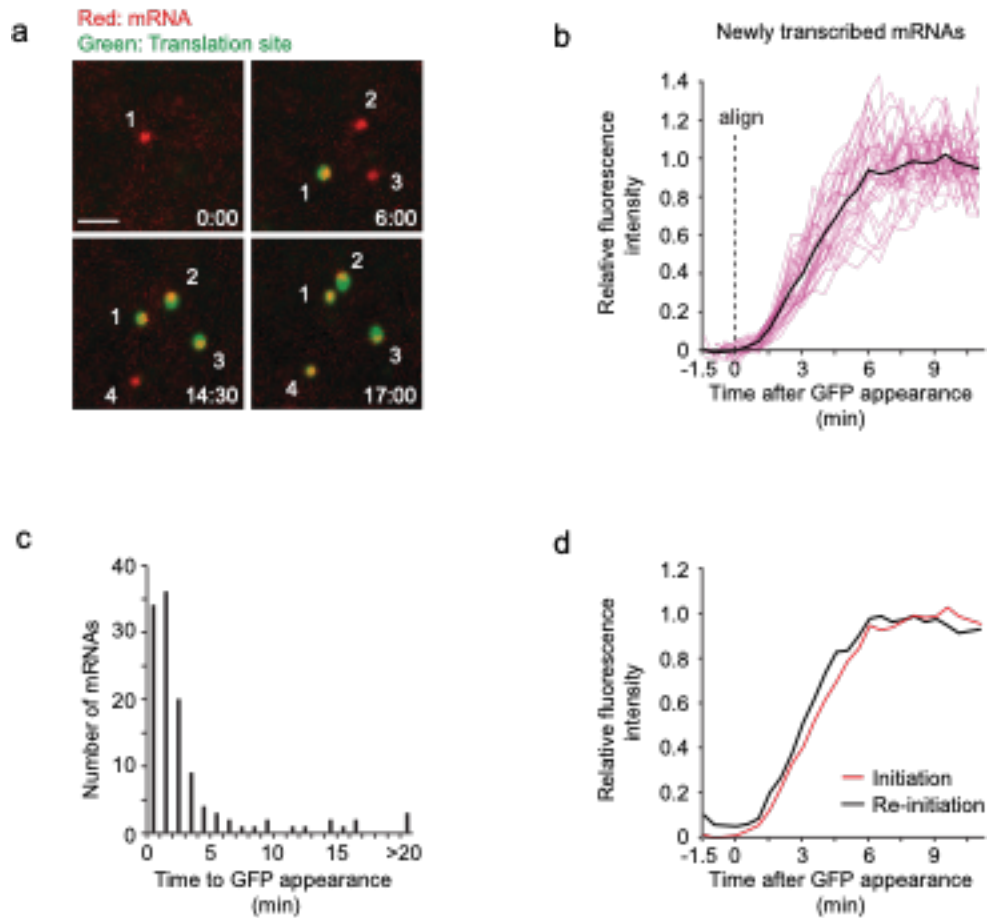


Figure 2.4 | Analysis of polysome build up on newly transcribed mRNAs. U2OS cells expressing scFv-GFP, PP7-2xmCherry-CAAX, and the translation reporter (SunTag_{24x}-Kif18b-PP7_{24x}). (a) Images from a time-lapse movie of newly transcribed mRNAs undergoing the first rounds of translation. (b) Quantification of the fluorescence intensity increase, aligned at the first time point at which scFv-GFP signal was detected ($n = 30$ individual mRNAs [pink lines], and average [black line] is shown). (c) Quantification of the time between mRNA appearance and the first detection of translation by scFv-GFP fluorescence. (d) Comparison of scFv-GFP fluorescence buildup on either new transcripts (red line) or on re-initiating mRNAs (black line). Data are re-plotted from **Fig. 2.3d** and **2.4b**. Scale bar, 2 μ m.

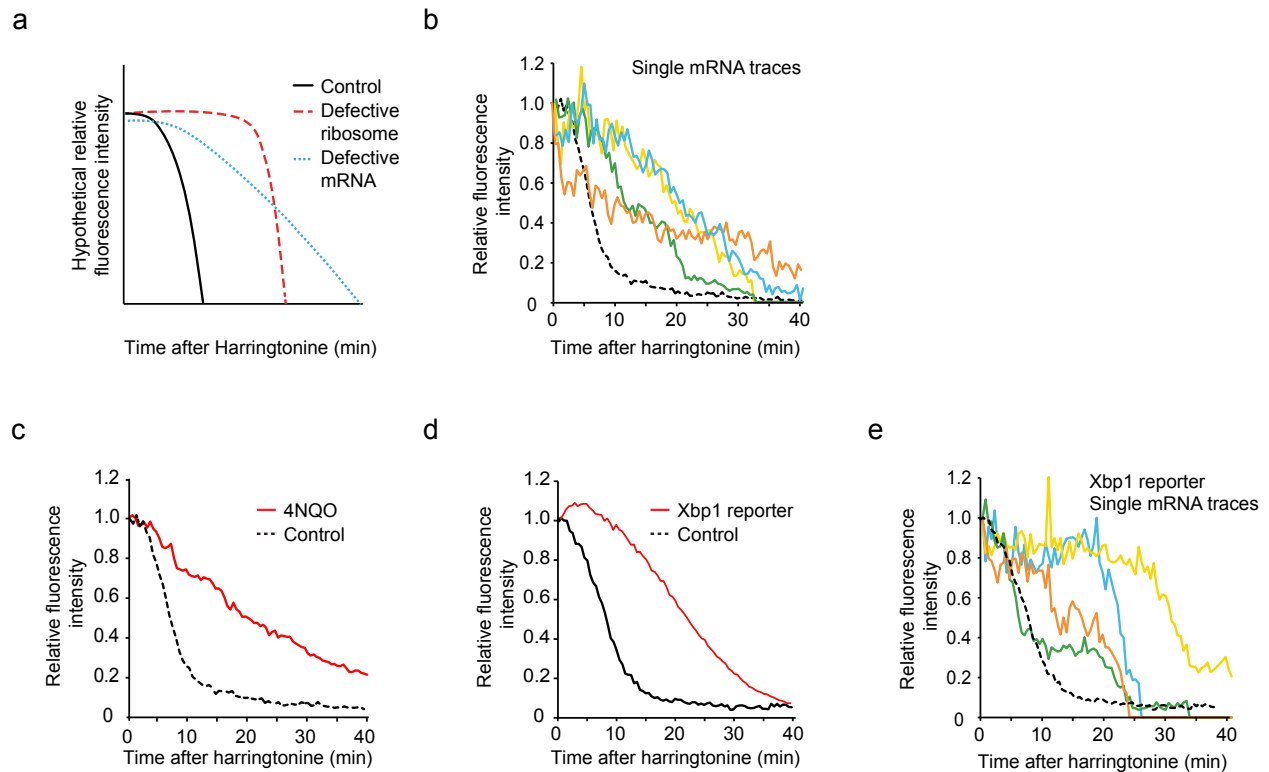


Figure 2.5 | Dynamics of ribosome stalling. U2OS cells expressing scFv-GFP, PP7-2xmCherry-CAAX, and the SunTag_{24x}-Kif18b-PP7_{24x} translation reporter (a-c) or the Xbp1 translation reporter (d-e). (a and b) Ribosome stalling likely results from mRNA defects, model (a) and experiment (b). (b) Fluorescence intensity over time is shown for four representative stalled translation sites (colors; of 20 analyzed). Since intensity values of single mRNAs were derived from the experiments presented in **Fig. 2.2d**, the average fluorescence decay presented in **Fig. 2.2d** is re-plotted here for comparison (dashed black line). (c) Nucleic acid damage through 4NQO treatment (red line) induces ribosome stalling (n = 3 independent experiments, 40 cells, 455 mRNAs). For comparison, the harringtonine runoff from control cells with the SunTag_{24x}-Kif18b-PP7_{24x} reporter from **Fig. 2.2d** is re-plotted, as these experiments were performed in parallel. (d and e) Harringtonine runoff for the Xbp1 pause site (red line, n = 3 independent experiments, 31 cells, 990 mRNAs) (d) and control reporter (black dashed lines, n = 3 independent experiments, 27 cells, 437 mRNAs) (e).

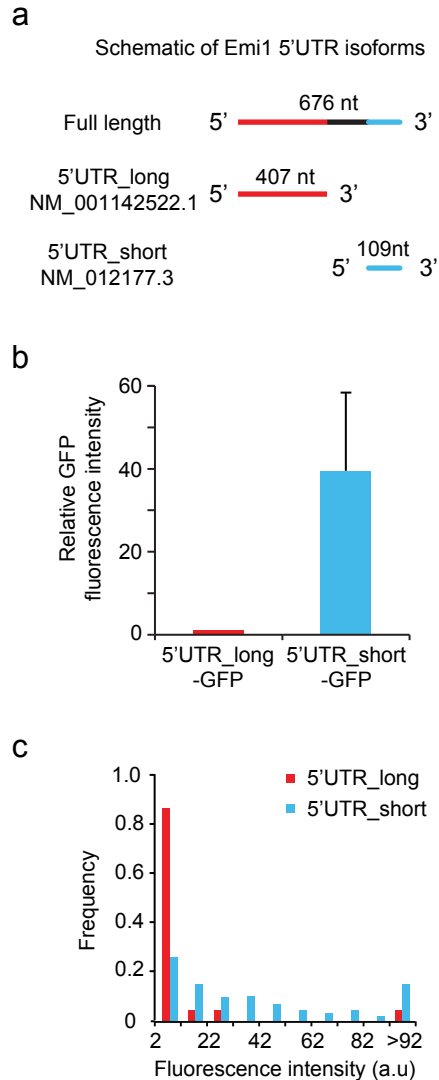


Figure 2.6 | Differential control of translation initiation by two Emi1 splicing isoforms. (a) Schematic of the 5' UTR of two Emi1 splicing isoforms. (b) Fluorescence intensity of a GFP reporter under control of the two Emi1 isoforms (5' UTR_long and 5' UTR_short) expressed in HEK293 cells was measured by microscopy for single cells. Mean intensities were determined, which was corrected for background fluorescence in untransfected cells. At least 20 cells were measured per experiment per condition. Error bars, SD between experiments. (c) Fluorescence intensity distributions of single translation sites of indicated reporters, n = 3 independent experiments, 283 mRNAs, 14 cells (5' UTR_long) and n = 3 independent experiments, 433 mRNAs, 16 cells (5' UTR_short). Background from adjacent regions was subtracted. Only mRNAs are plotted that had translation signal above background (with an intensity value >2; 16% and 53% of mRNAs for 5' UTR_long and 5' UTR_short, respectively).

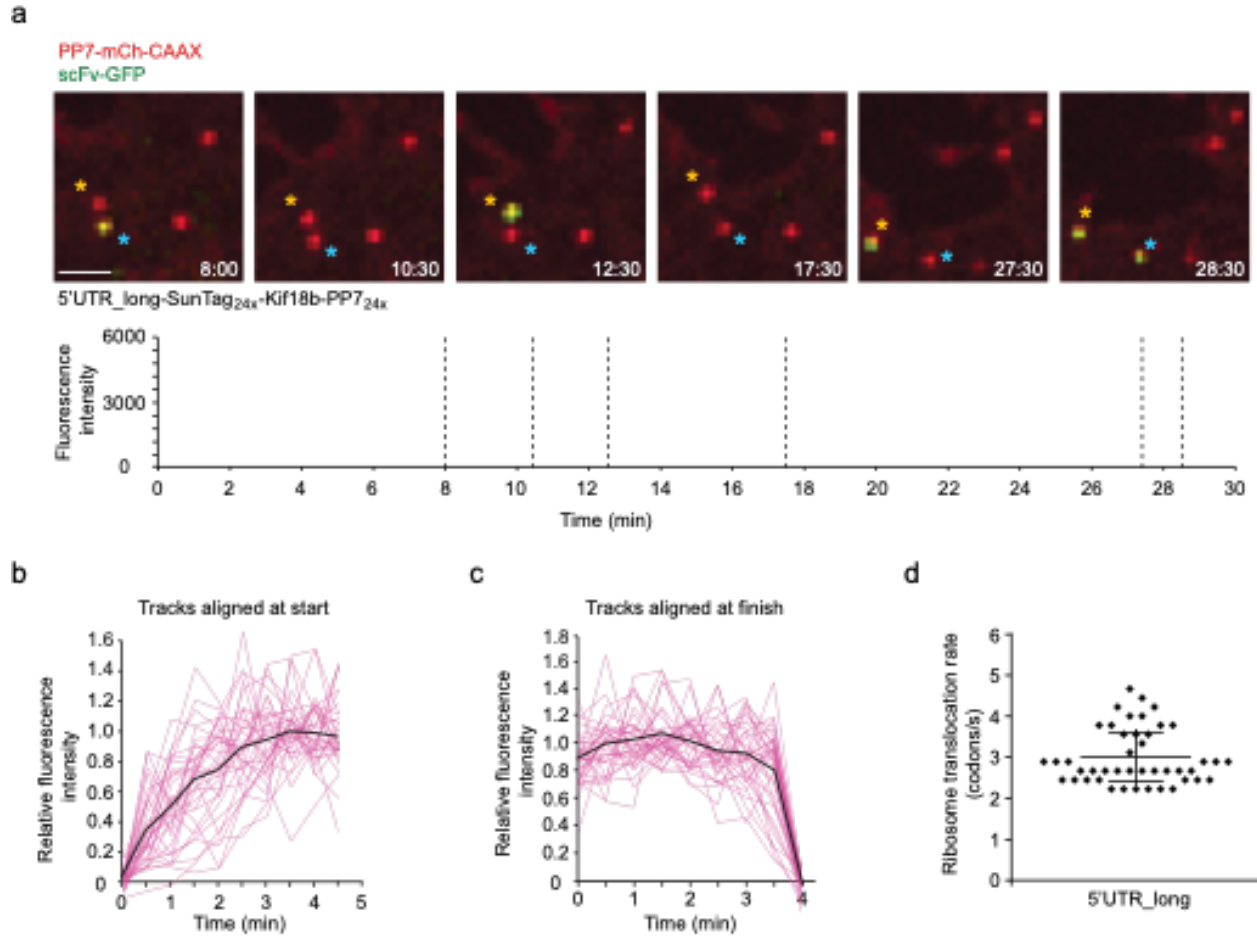


Figure 2.7 | Visualizing single ribosomes decoding an mRNA molecule. (a-d) Analysis of single ribosomes on the Emi1 5' UTR_long reporter mRNA. (a) Representative images of multiple single ribosome translation events of individual mRNAs (upper). ScFv-GFP intensity was quantified over time for the two mRNAs marked by asterisks with the same color (lower). (b) Increase in scFv-GFP fluorescence from single ribosome translation events aligned at the first detectable scFv-GFP signal ($n = 35$ individual mRNAs in pink and average in black). (c) Steady-state and then abrupt decrease in scFv-GFP fluorescence from single translating ribosomes ($n = 35$ individual mRNAs [pink] and average [black]). (d) Single ribosome elongation rates ($n = 44$). Mean \pm SD is shown in (d). Scale bar, 2 μ m.

Supplemental figures and figure legends

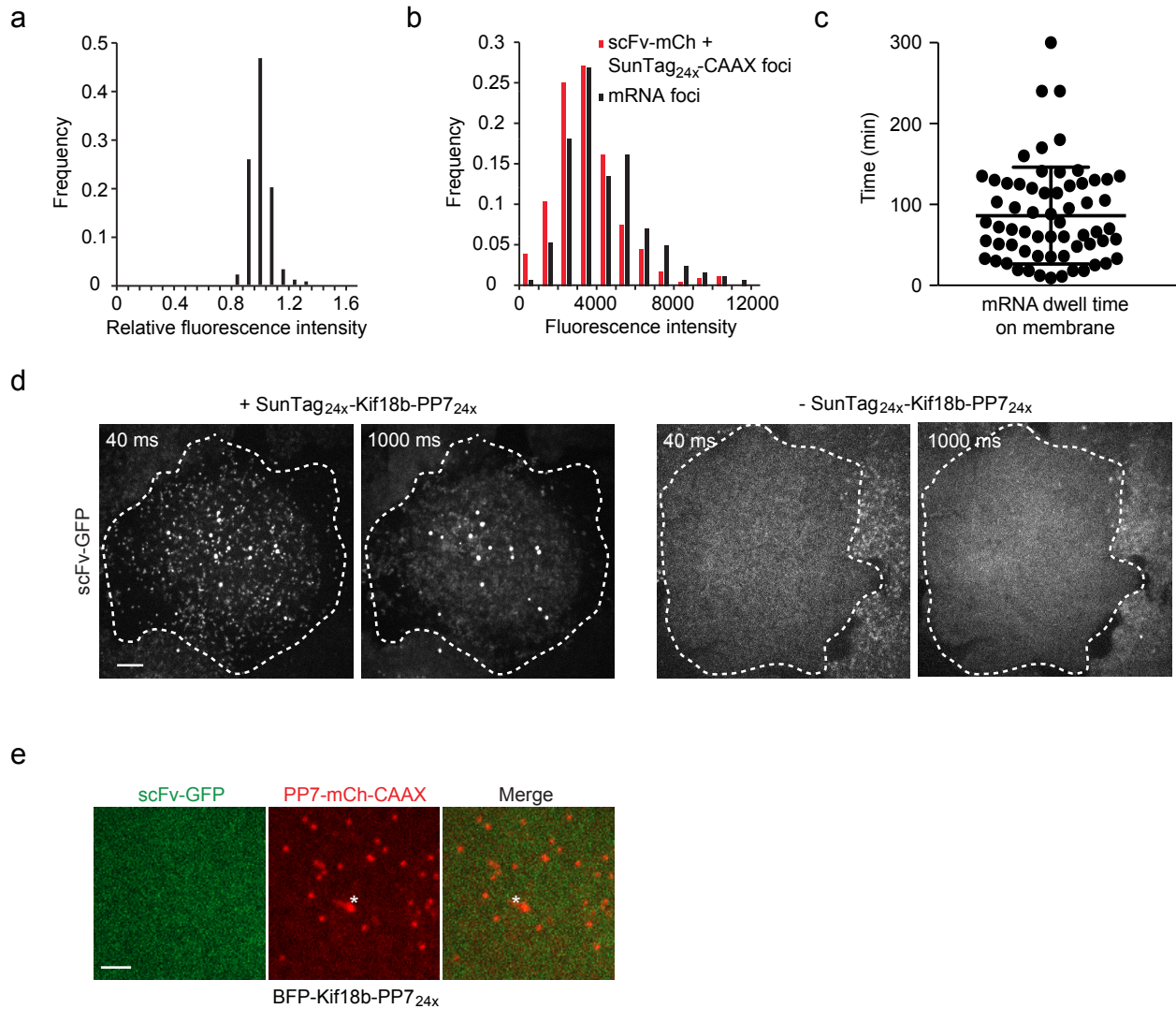


Figure S2.1 | Validation of single molecule translation visualization assay. (a-c) U2OS cells expressing scFv-GFP, PP7-2xmCherry-CAAX and the SunTag_{24x}-Kif18b-PP7_{24x} translation reporter. (a) Fluorescence intensity of mRNA foci was measured and was corrected for background fluorescence. The average corrected mRNA fluorescence intensity was set to 1 for each separate cell ($n = 3$ experiments, 14 cells, 278 mRNAs). (b) Intensity of single mRNA foci was measured and corrected for background, but intensity was not normalized as in (a) to allow comparison of absolute intensities (Black bars, $n = 3$ independent experiments, 22 cells, 377 mRNAs). In parallel, U2OS cells co-expressing SunTag_{24x}-CAAX and scFv-mCherry were imaged and the intensities of single membrane bound scFv-mCherry-SunTag_{24x} foci was measured (Red bars, $n = 4$ independent experiments, 24 cells, 162 mRNAs). (c) Dwell time of tethered mRNAs on the membrane. The time between mRNA appearance at the focal plane of the membrane and its disappearance was scored. mRNA disappearance was due to mRNA detachment or degradation, not photobleaching. Mean and SD are indicated. (d) Cells expressing scFv-GFP with (left two images) or without (right two images) the SunTag_{24x}-Kif18b-PP7_{24x} reporter were imaged with indicated exposure time. Dotted line shows outline of the cell. (e) U2OS cells expressing scFv-GFP, PP7-2xmCherry-CAAX and the BFP-Kif18b-PP7_{24x} translation

reporter. Representative image is shown. Asterisk indicates lysosome. Scale bars are 5 mm (d) and 2 mm (e).

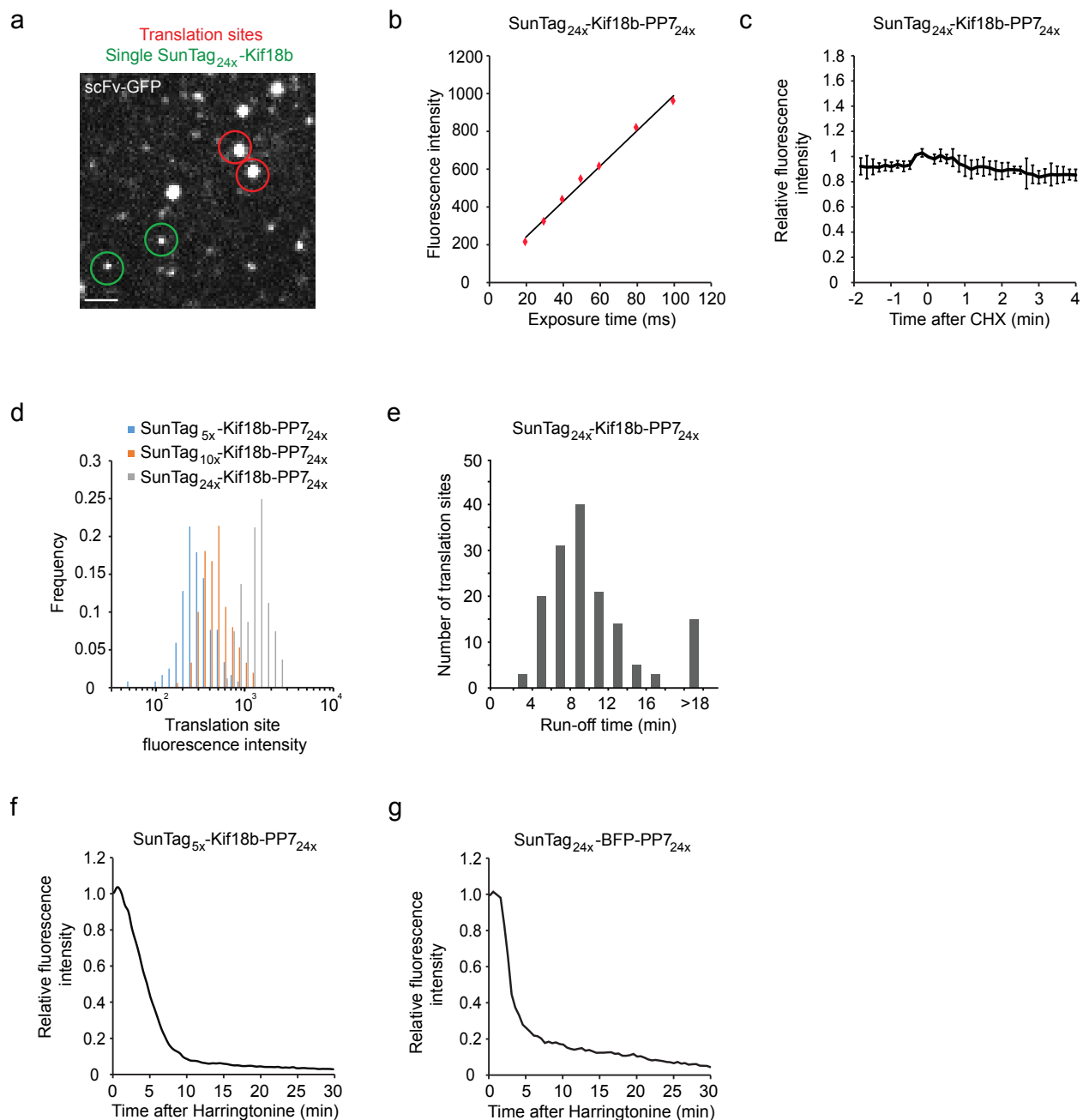
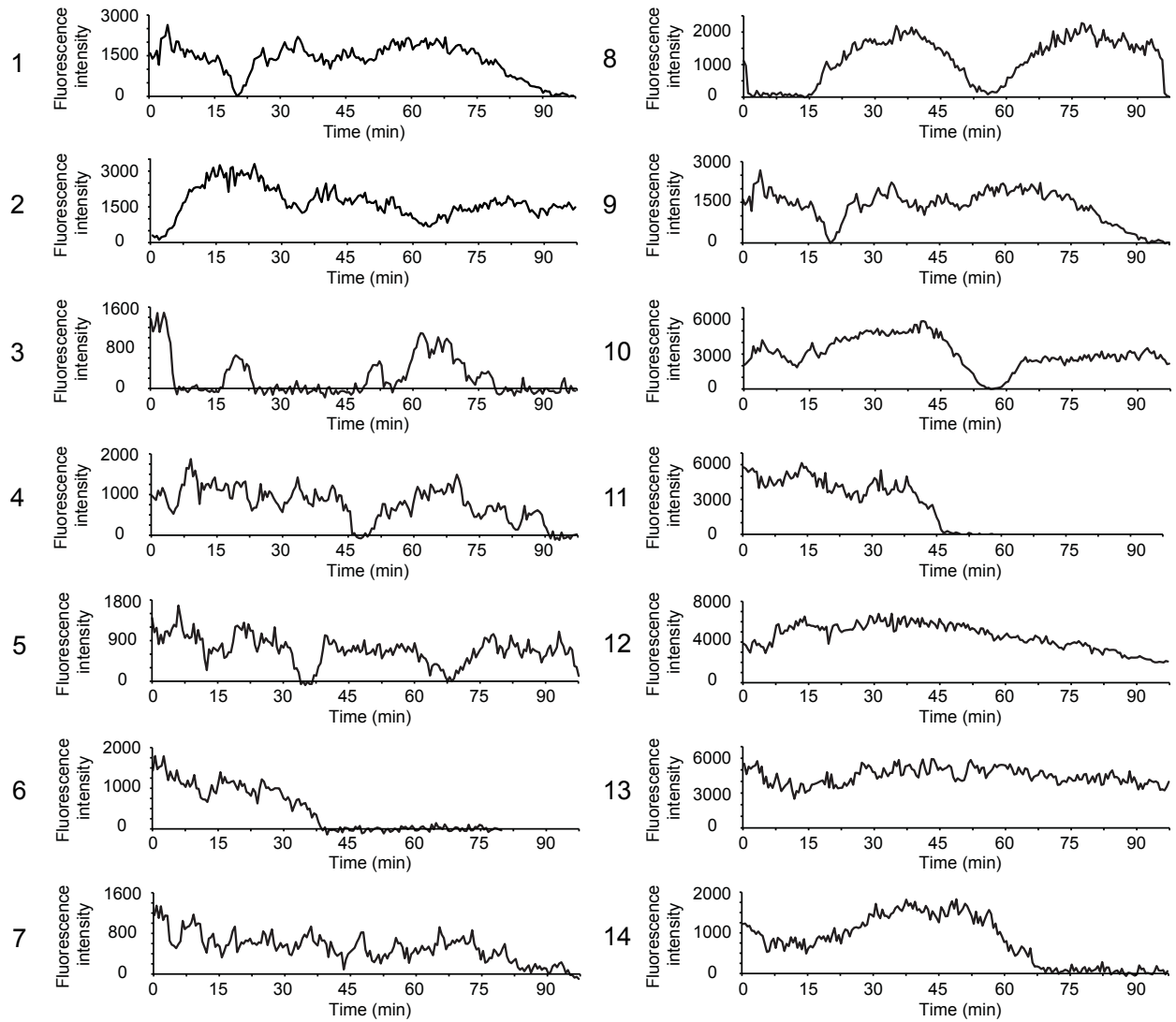


Figure S2.2 | Quantification of ribosome number and elongation speed on single mRNAs. U2OS cells expressing scFv-GFP, PP7-2xmCherry-CAAX and indicated translation reporters. (a) Images were acquired using short exposure times (40 ms), limiting motion blurring of fast moving particles, so both translation sites (red circle) and single, fully synthesized, freely diffusing SunTag proteins (green circles) could be observed as distinct foci. Fluorescence intensity of single SunTag_{24x}-Kif18b foci and single translation site was quantified in the same cell using a ROI with fixed size. (n = 45 translation sites, 15 cells, 3 experiments). (b) To determine whether the exposure time of 40 ms used in (a) was sufficiently short to prevent a reduction in fluorescence intensity of foci due to motion blurring, we measured the intensity of single fully synthesized SunTag_{24x}-Kif18b foci at different exposure times. Fluorescence intensities of the ~25 brightest foci per image were measured. Results show a linear relationship between exposure time and

fluorescence intensity at short exposure times, indicating that exposure times were short enough to prevent reduction in fluorescence intensity of foci due to motion blurring (n = 3 independent experiments, 18 cells and 400-500 spots). (c) Cells were treated with 200 mg/ml CHX at t = 0 and fluorescence intensities of translation sites were measured over time. Note that fluorescence does not increase upon CHX treatment (n = 3 independent experiments, 31 cells, 209 mRNAs). Error bars indicate SD. (d) scFv-GFP fluorescence intensity of translation sites using reporters with varying numbers of SunTag peptides (53, 103 and 243). (n = 117, 149, 80 translation sites for the 53, 103 and 243 reporters, respectively). (e-g) Cells were treated with harringtonine at t = 0 and translation-site intensity was quantified over time. (e) Histogram of the total run-off time, measured from the time of harringtonine treatment to the final disappearance of the scFv-GFP signal. 60 s was subtracted from all times to correct for the time required for harringtonine to enter the cell. (f and g) ScFv-GFP fluorescence intensity was measured over time after harringtonine addition (f, n = 3 independent experiments, 39 cells, 1883 mRNAs) (g, n = 3 independent experiments, 30 cells, 378 mRNAs). Scale bar, 2 μ m.

a



b

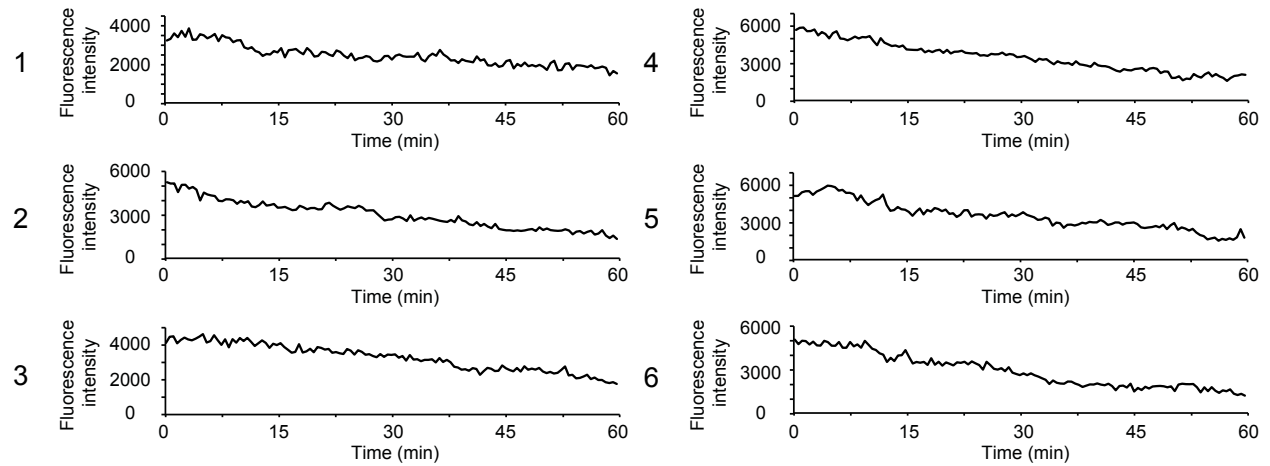


Figure S2.3 | Translation dynamics of single mRNA molecules. U2OS cells expressing scFv-GFP, PP7-2xmCherry-CAAX and the translation reporter (SunTag_{24x}-Kif18b-PP7_{24x}) were imaged by time-lapse microscopy for 2 hr. (a) or 1 hr (b) and the fluorescence intensity of single translation sites was tracked over time. 14 traces of untreated cells (a) or 6 traces of CHX treated cells (b) are shown. Note that the intensity of translation sites in CHX-treated cells slowly decreases over time, which is likely due to a decrease in the ribosome number per mRNA after prolonged CHX treatment.

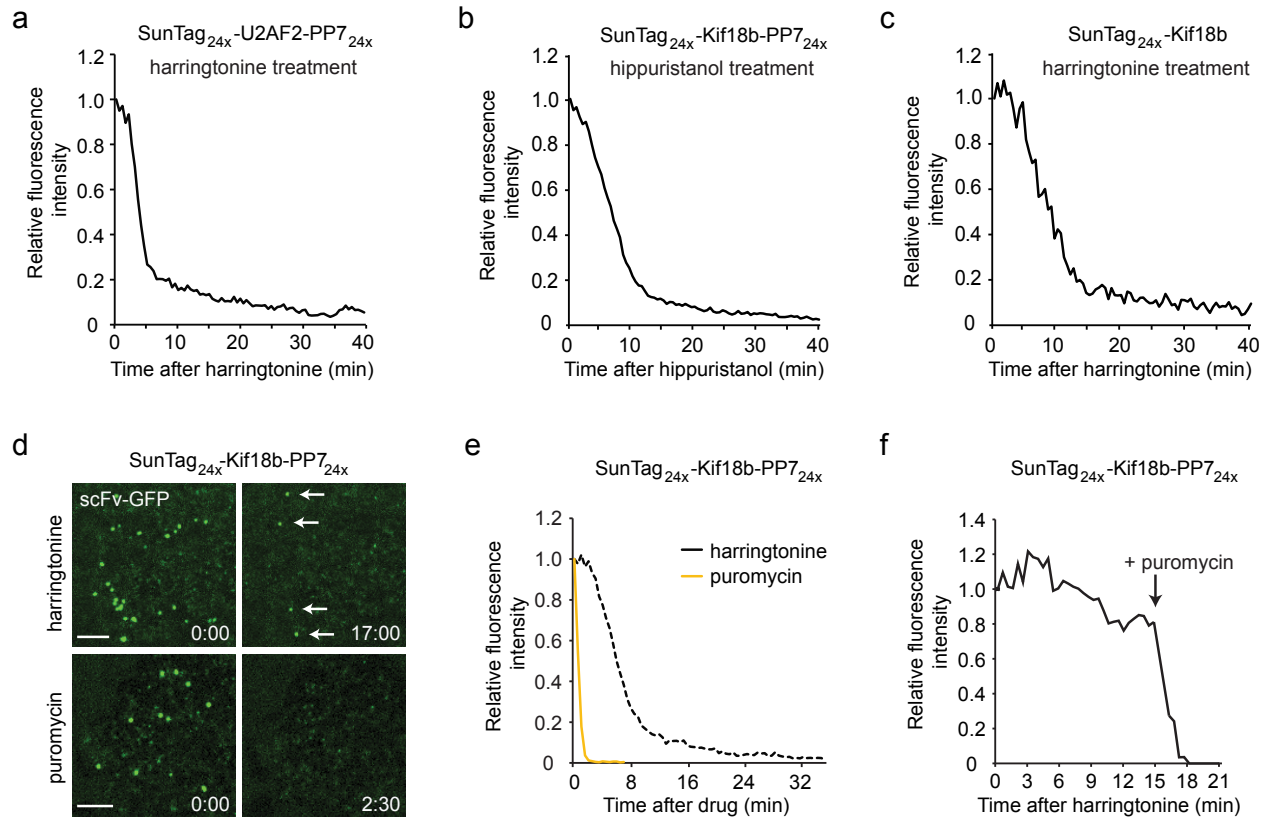


Figure S2.4 | Validation of the ribosome stalling phenotype. (a) U2OS cells expressing scFv-GFP, PP7-2xmCherry-CAAX and indicated translation reporters were treated with harringtonine at $t = 0$ and translation-site intensity was quantified over time. Reporter containing a codon optimized version of the U2AF2 coding sequence ($n = 3$ independent experiments, 29 cells, 512 mRNAs). (b) U2OS cells expressing the translation reporter (SunTag_{24x}-Kif18b-PP7_{24x}) were treated with another translation initiation inhibitor (hippuristanol) and translation-site intensity was quantified over time ($n = 2$ independent experiments, 14 cells, 515 mRNAs). (c) U2OS cells expressing scFv-GFP, PP7-2xmCherry-CAAX and indicated translation reporters were treated with harringtonine at $t = 0$ and translation-site intensity was quantified over time. Harringtonine run-off experiments were also performed on a translation reporter lacking PP7 binding sites (SunTag_{24x}-Kif18b) ($n = 2$ independent experiments, 19 cells, 248 mRNAs). (d) U2OS cells expressing scFv-GFP, PP7-2xmCherry-CAAX and indicated translation reporters were treated with harringtonine at $t = 0$ and translation-site intensity was quantified over time. Representative images in which stalled ribosome can be observed after harringtonine treatment (arrows). No ribosome stalling is observed after puromycin treatment (lower panel). (e) U2OS cells expressing scFv-GFP, PP7-2xmCherry-CAAX and indicated translation reporters were treated with harringtonine at $t = 0$ and translation-site intensity was quantified over time. At $t = 0$, either harringtonine (re-plotted from **Fig. 2.2d**) or puromycin ($n = 3$ independent experiments, 22 cells, 403 mRNAs) was added and translation-site intensity was quantified over time. (f) U2OS cells expressing scFv-GFP, PP7-2xmCherry-CAAX and indicated translation reporters were treated with harringtonine at $t = 0$ and translation-site intensity was quantified over time. Sequential addition of harringtonine and then puromycin ($n = 7$ mRNAs). Scale bar, 5 μ m.

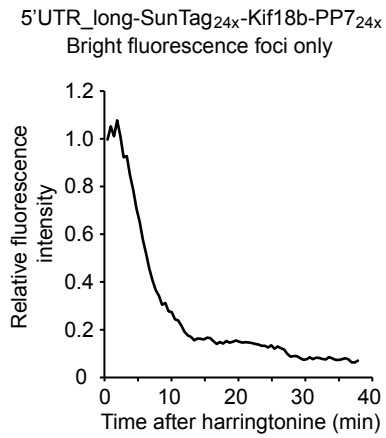


Figure S2.5 | Ribosome elongation rates on Emi1 5' UTR_long containing mRNAs. U2OS cells expressing scFv-GFP, PP7-2xmCherry-CAAX and the Emi1 5' UTR_long translation reporter were treated with harringtonine at $t = 0$. The fluorescence intensity of very bright translation sites was quantified over time ($n = 3$ independent experiments, 29 cells, 39 mRNAs).

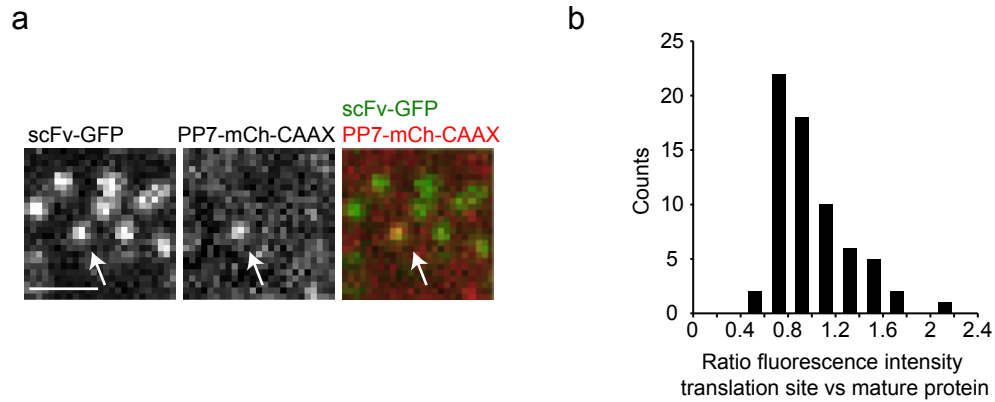


Figure S2.6 | Single ribosomes translate the Emi1 5' UTR_long mRNA. U2OS cells expressing scFv-GFP, PP7-2xmCherry-CAAX and the Emi1 5' UTR_long translation reporter were imaged using a very short (30 ms) exposure time, so fully synthesized, freely diffusing mature SunTag-Kif18b molecules can be observed together with translation sites. Translation sites could be distinguished from fully synthesized SunTag molecules, as they co-migrated with mRNAs for multiple (> 5) consecutive time points. (a) Representative image of a single translation site (arrow) surrounded by multiple mature SunTag molecules. (b) Quantification of fluorescence intensities of translation sites and mature protein. The fluorescence intensity of a single translation sites was compared to the average fluorescence intensity of 5 nearby mature SunTag molecules.

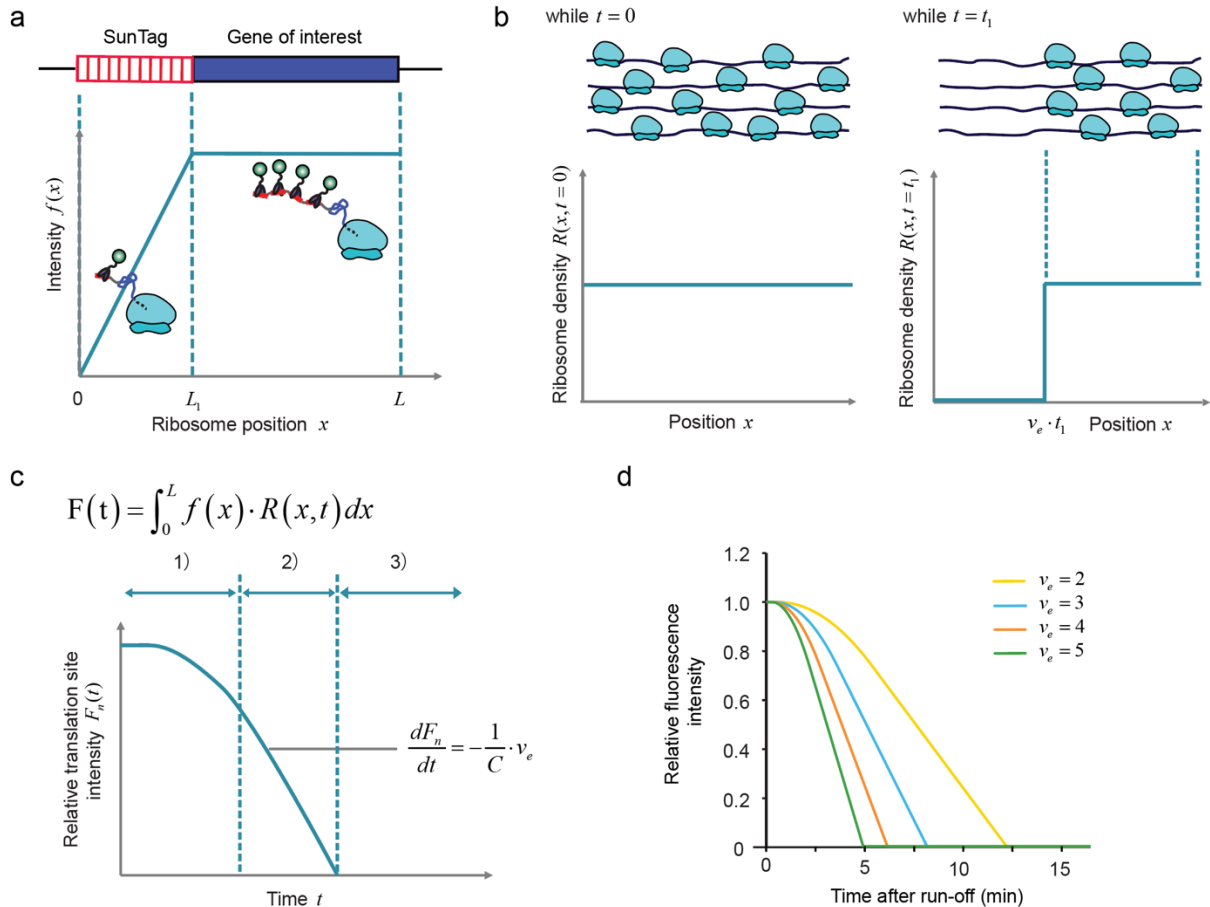


Figure S2.7 | Modeling of translation-site intensity. (a) Intensity from a single ribosome mainly depends on ribosome location on the mRNA. Due to the synthesis of SunTag peptides, ribosome intensity will increase initially as the ribosome moves toward the 3' end until SunTag peptides are fully synthesized and exposed. A typical curve for intensity function $f(x)$ is shown. For simplicity, a linear function was used to simulate the intensity increase. (b) Ribosome density changes during ribosome run-off. When there are no new initiation events, already bound ribosomes will runoff the mRNA from 5' to 3' end. Examples of the ribosome density function at $t = 0$ and $t = t_1$ are shown. (c) Translation-site intensity is dependent on both intensity from single ribosomes as well as ribosome density throughout the mRNA. A formula describing translation-site intensity is shown on top. A typical curve of intensity change during ribosome run-off process is shown at the bottom with three clear stages labeled using numbers. Intensity decreases linearly during the second stage, whose first order derivative could be used to derive elongation rate. (d) Example results of simulations of harringtonine run-off from the Kif18b reporter (SunTag_{24x}-Kif18b), which were run using different elongation rates (2, 3, 4, 5 codons/s). Run-off starts at $t = 0$.

Materials and Methods

Cell Culture and Drug Treatment

U2OS and HEK293 cells were grown in DMEM/5% with Pen/Strep. Plasmid transfections were performed with Fugene 6 (Roche), and stable transformants were selected with zeocin (Life Technologies). Unless noted otherwise, reporter transcripts were expressed from a doxycycline-inducible promoter, and expression of the reporter was induced with 1 mg/mL doxycycline (Sigma) for 1 hr before imaging. Harringtonine (Cayman Chemical) was used at 3 mg/ml. 5 mM 4NQO (Sigma) was added to cells for 1 hr before imaging. Puromycin (Life Technologies) was used at 100 mg/ml. Hippuristanol (a kind gift of Dr. J. Tanaka) was used at 5 mM. Cycloheximide (Sigma) was used at 200 mg/ml.

Microscopy

Cells were grown in 96-well glass bottom dishes (Matriplate, Brooks). Images were acquired using a Yokogawa CSU-X1 spinning disk confocal attached to an inverted Nikon TI microscope with Nikon Perfect Focus system, 1003 NA 1.49 objective, an Andor iXon Ultra 897 EM-CCD camera, and Micro-Manager software (Edelstein et al., 2010). Single z-plane images were acquired every 30 s unless noted otherwise. During image acquisition, cells were maintained at a constant temperature of 36 °C–37°C. Camera exposure times were generally set to 500 ms, unless noted otherwise. We note that stable expression of PP7-mCherry, either with or without the CAAX domain, also resulted in an accumulation of mCherry signal in lysosomes, but lysosomes could be readily distinguished from mRNA foci based on signal intensity and mobility.

FACS

GFP and scFv-GFP (**Fig. 2.1b**), mCherry, PP7-mCherry, or PP7-2xmCherry-CAAX (**Fig. 2.1f**) were expressed from a constitutive promoter, while the two reporters, SunTag_{24x}-mCherry and GFP-PP7_{24x} (**Fig. 2.1b and f**, respectively) were expressed from an inducible promoter in

U2OS cells expressing the Tet repressor protein, and their expression was induced 24 hr after transfection using doxycycline (1 mg/ml). This ensured that the reporters were translated in the presence of high levels of the scFv-GFP and PP7-2xmCherry-CAAX proteins. Cells were collected one day after doxycycline induction and analyzed by FACS. Cells were gated for GFP and mCherry double positivity, and the mCherry and GFP levels (**Fig. 2.1b and f**, respectively) were analyzed using Flowjo v10.1.

General considerations for quantifying translation based on scFv-GFP fluorescence intensity values

In this study, we provide estimates for the number of ribosomes per mRNA, reveal ribosome initiation and translocation rates and provide measurements of ribosome stalling. All these values were calculated based on fluorescence intensities of the scFv-GFP antibody bound to nascent polypeptides. It is important to note that such fluorescence intensities by themselves do not provide quantitative information, but only can be interpreted after taking several factors into consideration.

1) Ribosome position along the mRNA.

Ribosome position along the mRNA affects the number of SunTag peptides that have been synthesized, and thus the amount of scFv-GFP fluorescence associated with that ribosome/nascent polypeptide. We have generated a mathematical model to correct for this effect (see the section “Modeling of the fluorescence intensity of translation sites” below, and **Fig. S2.7**). It is important to note that this model assumes a homogeneous distribution of ribosomes along the mRNA which is likely accurate when averaging many mRNAs, but stochastic distributions can occur on an individual mRNA molecule. Also, non-homogenous distribution of ribosomes will not bias the estimate of ribosome number in ensemble statistics, as it is equally likely to underestimate or to overestimate ribosome number.

2) Variability in the number of scFv-GFPs bound to a SunTag peptide array.

The fluorescence intensity of a nascent polypeptide depends on the number of scFv-GFP molecules bound. The number of scFv-GFP antibody molecules bound to a single SunTag polypeptide can, however, vary. This variation may be limited though, as binding of the scFv-GFP to the 24 peptide SunTag array appears to saturate when sufficient cytoplasmic scFv-GFP is present, resulting in the binding of close to 24 scFv-GFPs per SunTag_{24x} array (Tanenbaum et al., 2014). Variation in the number of scFv-GFPs per SunTag_{24x} array will not affect the calculation of the number of ribosomes per mRNA, as this variation will occur on both the nascent polypeptides and the fully synthesized and released SunTag protein in the same cells which is used to normalize the fluorescence intensity of the translation sites. However, when a lot of SunTag_{24x} molecules have been synthesized and released into the cytoplasm, they can sequester the free scFv-GFP, leaving insufficient free scFv-GFP to bind all newly synthesized nascent SunTag peptides. Indeed, we have observed that at high expression levels of the reporter, translation sites become progressively dimmer after several hours, indicative of cytoplasmic depletion of the scFv-GFP. While such depletion of cytoplasmic scFv results both in dimmer translation sites and dimmer fully synthesized SunTag protein (used for normalization), and therefore does not affect our calculations of ribosome number per mRNA. Nonetheless, to circumvent this issue, we have used an inducible promoter to initiate reporter expression only ~1 hr before image acquisition. An additional source of variability in scFv-GFP binding to the SunTag peptides is introduced by the lag time that exists between synthesis of a new SunTag peptide and binding to the scFv-GFP. However, our experiments show that this lag time likely does not substantially decrease overall fluorescence intensity of the translation site, indicating that the binding reaction is relatively fast. This lag time likely depends on the level of free cytoplasmic scFv-GFP, and could therefore vary under different conditions.

3) Fluorescence decay of the scFv-GFP signal at translation sites.

In several experiments, including cases of natural translation shutdown and harringtonine ribosome run-off experiments, we observed a reduction in scFv-GFP fluorescence intensity at translation sites. We generally interpret this fluorescence decay as ribosomes terminating translation at the stop codon, which results in nascent chain release and ribosome recycling. While this conclusion is likely valid in most cases, alternative explanations are possible. For example, ribosomes stall for extended periods of time on mRNAs with chemical damage or on the Xbp1 nucleotide pausing sequence, but eventually scFv-GFP signal associated with stalled translation sites does decrease. While it is possible that ribosomes eventually normally read through such stall sequences and terminate normally at a stop codon, it is also possible that ribosomes are removed from mRNAs after a prolonged stalling event. Additional experiments are required to distinguish these events, for example by placing the stall-inducing sequence upstream of the SunTag peptide array.

4) Photobleaching

Photobleaching rates were determined by measuring the total GFP signal of the entire cell over time. In experiments where photobleaching resulted in a fluorescence decrease of >10% of the initial fluorescence intensity, fluorescence intensities of translation sites were corrected for photobleaching.

Calculation of Ribosome Elongation

Ribosome elongation rates were determined by three different methods.

1) Population measurements from harringtonine ribosome run-off (Related to **Fig. 2.2**)

The linear phase of the decrease in GFP fluorescence after new ribosome translation is prevented by addition of harringtonine provides information on the rate of ribosome movement along the mRNA. A model for the fluorescence signal is provided below (See Modeling of the

fluorescence intensity of translation sites) along with details of how the data was fit to determine the elongation rate.

2) Calculating the ribosome elongation rate based on single mRNA run-off times (Related to **Fig. S2.2e**)

The total time required for run-off of all ribosomes from individual mRNAs represents the time until the last (*i.e.* most 5') ribosome completes translation after harringtonine treatment. Individual translation site intensities were quantified using the `spot_counter` ImageJ plugin (<http://fiji.sc/SpotCounter>) developed by Nico Stuurman, which draws a box around the selected translation site and scores the number of time-points that the translation site was detected. Automatic tracking was manually curated for all spots to ensure high quality tracking data. The average run-off time for all mRNAs was then determined, and we subtracted 60 s from this time, the time required for harringtonine to enter the cell (Ingolia et al., 2011). We then divided the reporter length (1462 codons for the SunTag_{24x}-Kif18b-PP7_{24x} reporter) by the corrected run-off time to obtain the ribosome elongation rate. This rate assumes that the final ribosome loaded near the 5' end of the mRNA, which may not be true for all mRNAs. This assumption may slightly overestimate ribosome elongation rates, as the actual distance covered by the most 5' ribosome covers until completion of translation will be slightly less than 1462 codons, if it is downstream of the start codon at the time of harringtonine addition. However, considering that most mRNAs have ~20 ribosomes and the average inter-ribosomal distance is therefore ~70 codons, the most 5' ribosome will usually be within 70 codons of the start site, which represents a 5% error.

3) Determining the ribosome elongation rate from single ribosome tracking data (Related to **Fig. 2.7**)

By limiting translation initiation using the Emi1 5'UTR_{long} fused to the translation reporter, individual ribosomes could be observed decoding an mRNA molecule. For each single ribosome

translation event, we measured the duration of the event, by determining the first time-point that GFP could reliably be detected over the background until the GFP signal disappeared. Only tracks were included that lasted less than 12 min. However, determining ribosome elongation rate based on such measurements was non-trivial, as this requires knowledge of the precise position of the ribosome along the transcript when the nascent chain signal is first detected; presumably a number of SunTag peptides need to have been synthesized before sufficient signal has accumulated to allow detection. We therefore needed to identify the fluorescence intensity detection limit. To accomplish this, we first determined the maximal fluorescence intensity associated with a single ribosome; this occurs when the entire SunTag peptide array has been synthesized (**Fig. 2.7c**). Maximal intensity was calculated by aligning all of the traces ($n = 44$) at the last time point in which fluorescence was observed and averaging the fluorescence intensity of the last 3 time points before polypeptide dissociation. We then aligned all 44 traces at the first point we could visually detect fluorescence and found that we could reliably detect signal from single ribosomes when their nascent chain-associated fluorescence was $\sim 1/3$ of the maximal intensity (**Fig. 2.7b**). From this, we infer that ~ 8 of the 24 SunTag peptides likely have been synthesized at this point in time, which places these ribosomes on codon 245 (taking into account that SunTag peptides will not be labeled when inside the ribosome exit tunnel). Therefore, from the time of first fluorescence measurement to the release of the completed polypeptide on the stop codon (406 s on average for $n = 44$ individual ribosome translocations), the ribosome moves $1462 - 245 = 1217$ codons (since the whole transcript is 1462 codons long). The ribosome elongation rate can be calculated as the number of codons / time: $1217/406 = 3.0$ codons/s.

Translation Initiation Rate

1) Estimation of steady state initiation rate

To maintain a constant ribosome occupancy, the rate of removal of ribosomes from the mRNA molecule after completion of translation, must be balanced by new ribosome loading on

the mRNA through translation initiation. To estimate the translation initiation rate, we first calculated the average inter-ribosomal distance, which is 58-146 codons assuming a ribosome number per mRNA of 10-25 ribosomes on the SunTag_{24x}-Kif18b reporter. Considering the ribosome elongation rate of 3.5 codons/s (for the Kif18b reporter), a ribosome will complete translation on average every $58/3.5$ to $146/3.5 = 17$ to 42 s. Thus, if constant ribosome occupancy is maintained, the initiation rate must be between $1/17$ s⁻¹ to $1/42$ s⁻¹ or 1.4 to 3.6 min⁻¹. These values represent average initiation rates, as the initiation rate on single mRNA molecules will vary over time.

2) Comparing the translation initiation rate on newly transcribed versus polysomal mRNAs

Through the experiments in which single translating ribosomes were tracked, we determined that ribosomes could be first detected when $\sim 1/3$ of SunTag peptides had been synthesized (when the ribosome is positioned around codon 245 (see above)). Thus, when the first, pioneer ribosome was detected on a newly transcribed mRNA molecule, it was likely near this position. After initial fluorescence detection of the pioneer ribosome, translation site signal increased continuously for several min (**Fig. 2.4b**). This fluorescence increase was presumably both due to the increase in the number of synthesized SunTag peptides on the pioneer ribosome, as well as due to additional ribosomes that initiated after the pioneer ribosome. The translation site fluorescence intensity reached a steady state ~ 6 min after the initial detection of fluorescence (**Fig. 2.4b**), indicating that the ribosome density remained constant from that point on. At a translocation rate of 3-3.5 codons/s, it would take the pioneer ribosome ~ 6 min to reach the stop codon from the position of initial detection (codon 245). From these calculations, we conclude that the ribosome density on the mRNA has already reached its maximal value as soon as the pioneer ribosome completes translation. Since ribosome density is a function of ribosome initiation rate and elongation rate, we infer that the initiation rate on the newly transcribed mRNA must be

identical to that on mRNAs with established polysomes, assuming that the elongation rates are on average similar.

Estimation of Ribosome Number per mRNA

To obtain the number of ribosomes per mRNA molecule, we set up a normalization experiment to compare the average intensity of the single translation sites with the intensity coming from single, fully synthesized SunTag_{24x}-Kif18b proteins (**Fig. S2.2a**). As the fully synthesized SunTag-Kif18b proteins encompass 24 copies of the SunTag peptide, their intensity should be comparable with the fluorescence intensity associated with a ribosome at the 3' end on the mRNA. However, the ribosomes at the 5' end of the mRNA will be much dimmer because they have not fully synthesized the SunTag epitopes. Thus, dividing the translation site intensity by the intensity from a single fully synthesized SunTag-Kif18b protein will underestimate the actual number of ribosomes on an mRNA. A correction factor derived from our model presented below was used to obtain a more accurate estimate of ribosome number.

Modeling of the fluorescence intensity of translation sites

Every active translation site is composed of one single mRNA with multiple ribosomes undergoing translation. The SunTag peptides produced from ribosomes are bound by scFv-GFP antibodies floating in the cytoplasm, giving rise to a fluorescence signal from the translation complex. Ribosomes on the 5' end of the mRNA are still translating part of the SunTag peptide, thus resulting in a partial fluorescence signal from the emerging polypeptide chain, while ribosomes that are translocating on the 3' end of the gene have already synthesized the entire SunTag peptide array and are fully covered by the antibodies. Therefore, to interpret intensity changes from a transcript loaded with multiple ribosomes, we have generated a model that takes into account ribosome location as well as density on the mRNA as a function of time.

The intensity from a single ribosome will increase gradually as the ribosome moves towards the 3' end until it reaches a plateau where all SunTag peptides are synthesized (**Fig. S2.7a**). This relationship is described as following:

$$f(x) = \begin{cases} g(x) & x \in [0, L_1) \\ I_{sun} & x \in [L_1, L] \end{cases} \dots\dots\dots (1)$$

The parenthesis and square bracket are notations for intervals. The two numbers are the endpoints of the interval. Parenthesis indicates exclusion of corresponding endpoint while square bracket indicates inclusion of it.

$f(x)$: intensity from a single ribosome at position x

$g(x)$: intensity from a single ribosome at position x when x is in between 0 and L_1 , simplified as a linear function in the diagram in **Fig. S2.7a**

x : ribosome position on the mRNA at time t

L : the length of positions that can be covered by ribosomes, which includes the open reading frame decoding both the SunTag peptides and the gene of interest (GOI), as shown in **Fig. S2.7a**

L_1 : position on the transcript where intensity from a ribosome at that position reaches plateau intensity, as shown in **Fig. S2.7a**

I_{sun} : the intensity of a single SunTag array that is fully covered by scFv-GFP, with the intensity depending on the number of SunTag peptides fused to the gene of interest, multiplied by the intensity of a single scFv-GFP antibody

$$I_{sun} = n_{sun} \times i_{GFP} \dots\dots\dots (2)$$

n_{sun} : the number of SunTag peptides which are fused to the gene of interest (mostly 24 in this study)

i_{GFP} : the intensity of a single scFv-GFP

Thus the average intensity from a single translation site could therefore be described using the following formula:

$$F(t) = \int_0^L f(x) \cdot R(x, t) \cdot dx \dots\dots\dots (3)$$

$F(t)$: average fluorescence intensity from a single translation site at time t

$R(x, t)$: ribosome probability density at position x at time t .

With the assumption that ribosomes are randomly positioned in a population of transcripts (while this distribution may not be homogeneous on a single mRNA, it will approximate a random distribution when averaging hundreds of mRNAs, as is done in these experiments; thus this model is only accurate when a large number of mRNAs is analyzed), the probability density will be the same at different positions on the transcript at steady state (**Fig. S2.7b**, left), which leads to:

$$R(x, t) = R_s$$

$$R_s = \frac{n_r}{L_r} \dots\dots\dots (4)$$

R_s : ribosome density during steady state

n_r : the number of ribosomes on a single transcript

L_r : the length of transcript that is covered by the ribosomes

Modeling of shutdown process and elongation rate

Based on the above, we are able to model fluorescence intensity as a function of time. We use the translation shutdown process as an example:

When new ribosomes are no longer added at the 5' end (e.g. when initiation is blocked by harringtonine), previously bound ribosomes will run off the transcript from the 5' to the 3' end. This will change the ribosome distribution as a function of time t , as described by the following equations and the illustration depicted in **Fig. S2.7b**:

$$R(x, t) = \begin{cases} 0 & x \in [0, v_e \cdot t) \\ R_s & x \in [v_e \cdot t, L] \end{cases} \dots\dots\dots (5)$$

v_e : elongation rate; assumed to be a constant in the model.

If we incorporate this function into equation (3), we derive the following formula, which describes the change in intensity of a single translation site as a function of time ($t = 0$ marking the beginning of harringtonine treatment) (**Fig. S2.7c**).

$$F(t) = \begin{cases} R_s \cdot \left[\int_{v_e \cdot t}^{L_1} g(x) \cdot dx + I_{sun}(L - L_1) \right] & t \in \left[0, \frac{L_1}{v_e} \right) \quad 1) \\ R_s \cdot I_{sun}(L - v_e \cdot t) & t \in \left[\frac{L_1}{v_e}, \frac{L}{v_e} \right) \quad 2) \dots\dots\dots (6) \\ 0 & t \in \left[\frac{L}{v_e}, \infty \right) \quad 3) \end{cases}$$

This formula describes absolute intensity, which is dependent on many variables including the laser power used to excite the fluorophore. Therefore, we normalize this equation to the initial steady state fluorescence using the constant C_1 :

$$F_s = R_s \cdot \left[\int_0^{L_1} g(x) \cdot dx + I_{sun}(L - L_1) \right] = R_s \cdot C_1 \dots\dots\dots (7)$$

$$F_n(t) = \frac{F(t)}{F_s} = \begin{cases} \left[\int_{v_e \cdot t}^{L_1} g(x) \cdot dx + I_{sun}(L - L_1) \right] / C_1 & t \in \left[0, \frac{L_1}{v_e} \right) \quad 1) \\ I_{sun}(L - v_e \cdot t) / C_1 & t \in \left[\frac{L_1}{v_e}, \frac{L}{v_e} \right) \quad 2) \dots\dots\dots (8) \\ 0 & t \in \left[\frac{L}{v_e}, \infty \right) \quad 3) \end{cases}$$

F_s : translation site intensity during steady state

C_1 : constant number, $C_1 = \int_0^{L_1} g(x) \cdot dx + I_{sun}(L - L_1)$

To further simplify the model, a linear function was used to describe $g(x)$, and the results are shown in **Fig. S2.7d**. Data acquired in real experiments have stalling events, which complicates fitting the data to the simulated results. To simplify the determination of the elongation rate from the harringtonine run-off experiments, we only fit the second stage of the curve (**Fig. S2.7c**), which is given by:

$$F_n(t) = F(t)/F_s = I_{sun}(L - v_e \cdot t)/C_1 = -\frac{1}{C} \cdot v_e \cdot t + \frac{L}{C} \dots\dots\dots (9)$$

C : constant number, $C = C_1/I_{sun}$

The first order derivative of this stage will give us the elongation rate v_e :

$$\frac{dF_n(t)}{dt} = -\frac{1}{C} \cdot v_e \dots\dots\dots (10)$$

Intensity from stalled or slowly elongating translation sites contributes about 10-20% of the initial fluorescence intensity (**Fig. 2.2d**), which will lead to underestimation of the slope without correction. To overcome this, a linear regression was fit to the second half of the curve after initial run-off, and the intercept value was subtracted to exclude the influence of stalled translation sites. The curve was normalized after correction, and relative fluorescence intensity in between 40-80% was fit with a linear function to extract the slope. Elongation rate was calculated using the formula mentioned above. However, several more corrections were applied to the previous equation to account for the influence of 1) the codons translated before the SunTag array (23 aa) as well as the codons in between the SunTag array and the GOI (6 aa) 2) the nascent peptide buried in the ribosome exit tunnel (36 aa was used).

Modeling of the number of ribosomes on a single mRNA

Fully synthesized SunTag_{24x}-Kif18b proteins encompass 24 copies of the SunTag peptide, so their intensity should be comparable to the fluorescence intensity of a ribosome at the 3' end of the mRNA. However, ribosomes at the 5' end of the mRNA will be dimmer, since they have not fully synthesized the SunTag epitopes. Thus, dividing the translation site intensity by the intensity of a single fully synthesized SunTag_{24x}-Kif18b protein will underestimate the actual number of ribosomes on an mRNA.

To overcome this, a correction factor γ was applied. As described earlier in equation (4), ribosome density R_s is:

$$R_s = \frac{n_r}{L_r}$$

If we assume the transcript is fully covered by the ribosomes during steady state, the ribosome number n_r can be derived as:

$$n_r = R_s \cdot L_r = R_s \cdot L = \frac{(F_s \cdot L)}{C_1} = \gamma \cdot F_s / I_{sun} \dots\dots\dots (11)$$

The constant $\gamma = 2L / (2L - L_1)$ when a linear function was used to describe $g(x)$. As discussed in the previous section, more corrections were applied to the equation to account for influence of things such as ribosome exit tunnel length. Thus, the actual correction factor used for the Kif18b reporter is 1.32 (1/0.76).

Acknowledgements

We would like to thank Nico Stuurman for developing the spot_counter plugin for imageJ and for general help with microscopy, Dr. J. Tanaka for providing hippuristanol, and S. Ruijtenberg and L. Krenning for helpful discussions. This work was supported by a European Research Council (ERC) starting grant (ERC-STG 677936-RNAREG). M.E.T. was supported by a fellowship of the Dutch Cancer Society (KWF), and R.D.V. was supported by the Howard Hughes Medical Institute. A patent has been filed for the SunTag technology (M.E.T. and R.D.V.).

CHAPTER 3

High-Content Imaging-Based Pooled CRISPR Screens in Mammalian Cells

Abstract

CRISPR (clustered regularly interspaced short palindromic repeats)-based gene inactivation provides a powerful means for linking genes to particular cellular phenotypes. CRISPR-based screening typically uses large genomic pools of single guide RNAs (sgRNAs). However, this approach is limited to phenotypes that can be enriched by chemical selection or FACS sorting. Here, we developed a microscopy-based approach, which we name optical enrichment, to select cells displaying a particular CRISPR-induced phenotype by automated imaging-based computation, mark them by photo-activation of an expressed photo-activatable fluorescent protein, and then isolate the fluorescent cells using fluorescence-activated cell sorting (FACS). A plugin was developed for the open source software μ Manager to automate the phenotypic identification and photo-activation of cells, allowing ~1.5 million individual cells to be screened in 8 hr. We used this approach to screen 6092 sgRNAs targeting 544 genes for their effects on nuclear size regulation and identified 14 bona fide hits. These results present a scalable approach to facilitate imaging-based pooled CRISPR screens.

Introduction

High throughput sequencing in combination with CRISPR technology has greatly accelerated discoveries in biology through unbiased identification of many new molecular players in key biological processes (Hsu et al., 2014; Barrangou and Doudna, 2016; Kweon and Kim, 2018; Schuster et al., 2018). Using a high diversity sgRNA library, large numbers of genes can be manipulated simultaneously in a pooled manner and sgRNA abundance differences can be determined with high throughput sequencing quickly with low labor and financial cost. Thus far, pooled CRISPR screens have been limited to phenotypes that can be transformed into sgRNA abundance differences, such as growth effects (Gilbert et al., 2014; Shalem et al., 2014; Wang et al., 2014) or phenotypes that can be directly examined by flow cytometry (Parnas et al., 2015) or single cell molecular profiling (Dixit et al., 2016; Jaitin et al., 2016; Datlinger et al., 2017; Adamson et al., 2018; Wroblewska et al., 2018; Rubin et al., 2018). However, many important cellular phenotypes can only be detected under a microscope, which requires a robust method for transforming optically identified phenotypes into differences in sgRNA abundance. Arrayed sgRNA libraries greatly facilitate such microscope-based screens, but are not widely available. Several *in situ* sequencing (Feldman et al., 2019; Wang et al., 2019) and cell isolation methods (Chien et al., 2015; Piatkevich et al., 2018; Wheeler et al., 2020) were developed that allow microscopes to be used for screening. However, these methods contain non-high throughput steps that limit their scalability.

Recently, an imaging based method named “visual cell sorting” was described that uses the photo-convertible fluorescent protein Dendra2 to enrich phenotypes optically, enabling pooled genetic screens and transcription profiling (Hasle et al., 2020). Here, we developed an analogous approach to execute an imaging-based pooled CRISPR screen using optical enrichment by automated photo-activation of the photo-activatable fluorescent protein, PA-mCherry. Similar to traditional enrichment based pooled CRISPR screens, cells are infected with a sgRNA library and high throughput sequencing is used to examine sgRNA abundance. Instead of traditional

enrichment strategies, we use optical enrichment — cells exhibiting the desired phenotype are identified and photo-activated automatically under a microscope. Photo-activated cells are then isolated using flow cytometry and analyzed by high throughput sequencing. We first evaluated this approach using a synthetic fluorescent reporter to estimate screening accuracy and capacity. We then applied this approach to identify genes that regulate nuclear size. This methodology is modular, allows millions of cells to be screened within a few hours, and can be scaled to a genome wide level.

Results

An optical approach for cell enrichment by patterned illumination followed by FACS sorting

We developed an approach, which we term optical enrichment, to select cells of interest using a microscope and mark them by photo-activation, enabling cell isolation using FACS (**Fig. 3.1**). To achieve this, we engineered hTERT-RPE1 cells expressing the photo-activatable fluorescent protein PA-mCherry and observed them under a microscope. A photo-activatable fluorescent protein was chosen over a photo-convertible fluorescent protein to increase the number of channels available for imaging. PA-mCherry was chosen to leave the better performing green channel open for labeling of other cellular features. Moreover, non-activated PA-mCherry has low background fluorescence in the mCherry channel (**Fig. S3.1b**), and it can be activated to different intensities when photo-activated for various amounts of time. Cells of interest were selected by automated image analysis and then photo-activated with patterned illumination using a digital micromirror device (DMD) (**Fig. S3.1a**). To avoid undesired photo-activation of neighboring cells, we limited the activation pattern to nuclei as identified by the H2B-mGFP signal (**Fig. S3.1b**). We developed a plugin for the open source microscope control software μ Manager (Edelstein et al., 2014) called Auto-PhotoConverter that automates these steps and has a pluggable interface for image analysis code so that it can be used for any desired phenotype (<https://github.com/nicost/mnfinder>) (**Fig. S3.1c**). After harvesting the cells, the photo-activated cells were isolated by FACS. By varying the activation time of the PA-mCherry, we were also able to create multiple populations of cells with different intensities that were distinguishable by FACS (**Fig. 3.2a and b**), enabling analysis of multiple phenotypes simultaneously, as discussed below.

We next tested the precision of our automated photo-activation platform in a “mock screen” consisting of a mixture of cells expressing the fluorescent marker mIFP and cells not expressing mIFP (outlined in **Fig. 3.2c**). In this mock screen, mIFP fluorescence was used as a “phenotype” to indicate cells of interest (mIFP positive cells). The Auto-PhotoConverter plugin was used to

identify and generate an activation mask based on mIFP fluorescence to photo-activate the mIFP expressing cells into mIFP-mCherry double positive cells (true positive cells). Cells without mIFP fluorescence might also be accidentally photo-activated, leading to false positive cells (mCherry single positive cells). To evaluate the precision (the fraction of called positives that are true positives) of this assay, all cells were collected and analyzed by FACS after image analysis and photo-activation (**Fig. 3.2d and e**). We calculated precision as the fraction of photo-activated cells (mCherry positive cells) that are true positives (mIFP-mCherry double positive cells) (**Fig. 3.2f**). When the initial subset of mIFP positive cells was 30%, the precision was 98.3% (**Fig. 3.2e**). The precision varied with the initial percentage of mIFP positive cells and ranged from 80% to ~100% (initial percentage of mIFP positive cells ranging between 2.3% and 43.7%) (**Fig. 3.2f**). Precision is expected to fall below 80% with initial percentage of mIFP positive cells less than 2.3%. However, these results indicate that optical enrichment can be used to identify hits with high precision even at relatively low hit rates.

Optical enrichment enables accurate sgRNA identification

Having established that we can recover photo-activated cells with high precision, we next tested if we can successfully identify specific sgRNA sequences present in these cells. mIFP negative cells and mIFP positive cells were separately infected with two different CRISPRa sgRNA libraries (6100 sgRNAs for mIFP negative cells; 860 sgRNAs for mIFP positive cells) at a low multiplicity of infection (MOI) to guarantee a single sgRNA per cell. Note that in these experiments, the sgRNAs only function as barcodes to be read out by sequencing, but do not cause phenotypic changes as the cells do not express corresponding CRISPR reagents. These two populations were then mixed at a ratio of 9:1 mIFP negative cells: mIFP positive cells. We again used mIFP expression as our phenotype of interest (outlined in **Fig. 3.3a**). Two biological replicates were performed and at least 200-fold coverage of each sgRNA library was guaranteed throughout the screen, including library infection, puromycin selection, imaging/photo-activation

and FACS. For each replicate, we screened a single imaging plate. A total of 1,825,740 and 1,490,188 RPE-1 cells in the two replicates were imaged separately using a 20x objective. Automated imaging and photo-activation of the plate took ~8 hr. The mCherry positive cells were isolated by FACS, and cells passing through the sgRNA gate without further analysis were also collected as a control (unanalyzed sample). These cells were separately prepared for high throughput sequencing for sgRNA information extraction.

For simplicity, we use the terms “mIFP sgRNAs” for the sgRNAs used to infect mIFP positive cells and “control sgRNAs” (simplified as “Ctrl sgRNAs”) for the sgRNAs used to infect mIFP negative cells. Typically, sgRNA libraries contain multiple sgRNAs that target a single gene, which minimizes confounding effects that arise from differences in sgRNA efficacy. Since the mIFP positive phenotype is not induced by our sgRNA library, we emulated genes in normal sgRNA libraries in our analysis by grouping different numbers of randomly selected sgRNAs.

Our results show that the sgRNA groups from mIFP positive cells (mIFP groups) could be well distinguished from the sgRNA groups in mIFP negative cells (control groups, simplified as Ctrl groups) (**Fig. 3.3b**). To further investigate how library composition and number of replicates influence screening results, we also analyzed the data by grouping the sgRNAs differently (either 1 sgRNA or 2 sgRNAs were assigned to each group) and 2 different numbers of replicates (phenotypic scores calculated from 1 replicate versus phenotypic scores averaged between 2 replicates). As shown in Fig. S2, mIFP sgRNAs could be distinguished from control sgRNAs in a single experimental replicate (**Fig. S3.2**, top left). Combining data from both replicates significantly improved segregation of the mIFP and control groups (**Fig. S3.2**, top right). Not surprisingly, the greater the number of sgRNAs assigned to a group, the better the detection of mIFP groups (**Fig. S3.2**, bottom). Two sgRNAs per group is sufficient for a reliable screening result, even using a single replicate (**Fig. S3.2** bottom left). Thus, we demonstrate that pooled CRISPR libraries can be screened for phenotypes under a microscope by optical enrichment.

Improved phenotype identification through multi-intensity labeling

In most pooled CRISPR screens, only cells showing the phenotype of interest are selected and the relative enrichment of a given sgRNA is calculated based on comparison with the whole cell population. However, this whole cell population is usually collected separately and includes both positive and negative cells, which reduces the perceived enrichment in the positive population. We therefore investigated calculating the relative enrichment of a given sgRNA by comparing with true negative cells. Not all mCherry negative cells are true negative cells since there are unanalyzed regions outside of the microscope field of view (grey region in **Fig. 3.3c** top panel) and cells that fail to pass the filters for phenotype identification. Thus, true negative cells also need to be labeled before harvesting. This task requires selecting for multiple phenotypes simultaneously. We achieved this within the same experiment using different photo-activation times for true positives (2 sec) and true negatives (100 ms) and separating them by FACS (**Fig. 3.3c**). For comparison, we also collected cells going through the same experimental procedure that were not analyzed during image analysis (unanalyzed cells, mCherry negative cells) to determine the sgRNA composition in the total cell population. As shown in **Fig. 3.3d**, the peaks indicating groups of mIFP sgRNAs and control sgRNAs were separated to a much greater extent when comparing with true negative cells rather than with the whole cell population (**Fig. 3.3b**), demonstrating that this approach can indeed improve screening of pooled sgRNA libraries. Additionally, this approach can be used to screen for multiple different phenotypes simultaneously.

Pooled CRISPR screen for factors involved in nuclear size regulation

To further test our screening method, we performed a screen for regulators of nuclear size. We generated a CRISPRi library of 6092 sgRNAs targeting 544 genes (10 sgRNAs/gene with 22 non-targeting sgRNAs) whose translation efficiency is upregulated during the G2 phase of the cell cycle. This library includes sgRNAs targeting FBXO5, which is known to cause larger nuclei after knock down (Machida and Dutta, 2007; Verschuren et al., 2007), and served as the positive

control. For this experiment, hTERT-RPE1 cells were further engineered with a CRISPRi modality (dCas9-KRAB-BFP) to inhibit transcription of genes targeted by the sgRNA library. We defined nuclear size as the 2D area in square microns measured by H2B-mGFP using an epifluorescence microscope, as determined by automated image analysis (**Fig. 3.4a**). We selected 2 control sgRNAs that have no targeting sites in the human genome, and as expected had no discernible effect on nuclear size (**Fig. S3.3a**). Nuclear sizes were measured for control cells and the value of the top 0.5% was used as the screening threshold ($1,000 \mu\text{m}^2$).

Positive cells were photo-activated and sorted together with unanalyzed cells as a comparison. Two biological replicates were performed, each consisting of 4 imaging plates, containing 5,521,518 and 5,795,313 RPE-1 cells in total respectively. Both replicates were completed within 2 days (each plate taking 7-10 hr of imaging/photo-activation). The 4 imaging plates per replicate were carried out as separate screening experiments, termed runs, and data was only combined after sgRNA abundance determination (**Fig. 3.4b**). Simulated negative controls were generated computationally by randomly regrouping all the sgRNAs (10 sgRNAs/group) and a phenotypic score was calculated for each gene and simulated negative control as described in the Methods. A score η summarizing effects from both severity of the phenotype (phenotypic score) as well as trustworthiness of the phenotype ($-\ln(\text{p value})$) were calculated, and empirical false discovery rate (eFDR) = 0.1% was used to call hits for further analysis (**Fig. 3.4c, S3.3b and S3.3c**). The two replicates correlate well (**Fig. 3.4d**) and combined yielded 30 hits of which 15 genes were found in both replicates, including the positive control FBXO5 (**Fig. 3.4e**).

To estimate the minimum requirements for performing an optical enrichment pooled CRISPR screen, we computationally analyzed the effect of both library composition and number of runs on the screening results. Utilizing data from replicate 2, we re-ran the analysis, comparing results when fewer sgRNAs per gene and/or fewer runs were included. We binned the sgRNAs based upon three commercially available CRISPRi libraries: 10 sgRNAs/gene and the “Top5” or

“Supp5” sub-pools of the sgRNA library (Horlbeck et al., 2016). “Top5” and “Supp5” libraries were first designed in Jonathan Weissman’s lab by splitting their original 10 sgRNAs/gene library based on predicted sgRNA knockdown activity (Horlbeck et al., 2016). As expected, the “Top5” sgRNAs yielded more hits than “Supp5” sgRNAs (**Fig. S3.3d**). In addition, the “Top5” sgRNA library behaves similarly to the original 10 sgRNAs/gene library, suggesting 5 efficient sgRNAs are sufficient for hit identification using our imaging based screening approach. Even in the scenario of Top5 sgRNAs for 2 runs, around 20 hits were successfully identified (**Fig. S3.3d**). Thus, based upon factors such as the time to run a screen and available sgRNAs, fewer sgRNAs/gene and/or fewer runs can be used in a screen.

Since nuclear size often positively correlates with DNA content and cell size, we also sorted cells based upon H2B-mGFP intensity (as a proxy for DNA content) or forward scattering (FSC) signal (cell size) (**Fig. 3.4f and S3.4**). To compare results directly, these two screens were performed at the same time as our imaging-based nuclear size screen (**Fig. 3.4f**). The top 10 percentile of cells based on either GFP fluorescence or FSC signal were separately sorted and prepared for high throughput sequencing. In the H2B-mGFP intensity screen, two replicates identified 11 and 16 hits respectively, with 7 in common, while 7 and 0 were identified in the FSC screen (**Fig. S3.4**). Together, a total of 21 genes were captured in the H2B-mGFP and FSC screens (**Fig. 3.4g**); 15 out of these 21 genes were also identified through the imaging-based nuclear size screen. These data suggest that a direct measurement utilizing a microscope can provide different information and reveal hits that are inaccessible using other screening approaches.

Genes involved in nuclear size regulation

We applied optical enrichment to a screen for genes involved in nuclear size determination and identified 15 hits. To validate the 15 genes that emerged in both replicates of the microscope-based screen for enlarged nuclei, each gene was individually targeted using the sgRNAs from

our sgRNA library. 11 out of 15 genes showed >75% knockdown, as revealed by RT-qPCR with most genes demonstrating almost complete knock down (**Fig. S3.5**). Furthermore, 14 out of 15 hits were confirmed to be real hits (Kolmogorov-Smirnov test two tailed p-value < 0.01 for at least 2 out of 3 replicates; the exception was TACC3 which could be explained by incomplete knock down; **Fig. 3.5a, S3.5 and S3.6**). Among the 14 verified genes, all have known roles during cell cycle regulation except KRI1 which is involved in cell death regulation in *C. elegans* (Ito et al., 2010) (**Table S3.1**). Six genes are involved in spindle function and chromosome segregation, which includes KIF11 (Rapley et al., 2008), NUP62 (Hashizume et al., 2013), SPDL1 (Gassmann et al., 2008) and three core chromosomal passenger complex (CPC) components INCENP, AURKB and CDCA8 (Terada, 2001; Carmena et al., 2012). Three genes function in DNA damage and repair, namely TICRR (Sansam et al., 2010; Yu et al., 2019), TOP2A (Bower et al., 2010; Yoshida and Azuma, 2016) and RAD51 (Yoon et al., 2014; Sullivan and Bernstein, 2018), while the remaining four play roles in histone synthesis (CASP8AP2(Sokolova et al., 2017)), DNA maintenance (DNA2 (Duxin et al., 2009; Pawłowska et al., 2017)) and cell cycle regulation (SKA1 (Sivakumar et al., 2014, 2016) and FBXO5 (Verschuren et al., 2007; Machida and Dutta, 2007)) (**Table S3.1**). Some of these functions might directly explain the larger nuclei phenotype after knock down. For example, the loss of FBXO5 was suggested to lead to cellular senescence, resulting in larger nuclei (Verschuren et al., 2007). Knockdown of CPC components including AURKB, INCENP and CDCA8 leads to asymmetrical distribution of nuclear material and cytokinesis failure, thereby generating abnormally large nuclei (Terada, 2001; Carmena et al., 2012).

To begin to understand the mechanism underlying nuclear size regulation of our 14 hits, we investigated changes in DNA content, measured by DRAQ5 staining, and cell size, assessed using forward scattering on FACS, after knock down. Almost all hit genes show increases in the forward scattering signal (**Fig. 3.5b and S3.6**). This matches with the karyoplasmic ratio theory, which suggests that nuclear size is always related with cellular size (Webster et al., 2009; Edens

et al., 2013; Mukherjee et al., 2016; Cantwell and Nurse, 2019). On the other hand, DRAQ5 signal was unchanged or somewhat lower after knock down (**Fig. 3.5b and S3.6**), suggesting that these gene knockdowns do not change cellular DNA content.

Discussion

High throughput sequencing has transformed our ability to perform pooled genetic screens on a broad scale. However, applying enrichment-based pooled CRISPR screens to optical based phenotypes has been challenging. In this study, we developed an imaging-based pooled CRISPR screening method. Using the photo-activatable fluorescent protein PA-mCherry, cells of interest can be labeled through photo-activation and isolated with FACS sorting, which enables sgRNA identification by high throughput sequencing. We have combined this optical enrichment strategy with pooled CRISPR-Cas9 libraries to perform imaging based CRISPR screens. Independently, Kanfer *et al.* described a similar method to ours for imaging-based pooled CRISPR screening (Kanfer et al., 2020).

Advantages and limitations of phenotypic screening by optical enrichment

Image processing and microscope operations are the time limiting steps of most imaging-based genetic screens. Our optical enrichment based pooled screening method is relatively fast and scalable. For example, the image analysis code developed for this study can be run on millisecond time scale per field of view, and cells can be separated from the control population on a FACS machine with as little as 100 ms photo-activation time (**Fig. 3.2b**), enabling screening of large numbers of cells. In our system, 1.5 million RPE-1 cells can be imaged and photo-activated in 8 hr with a 20x objective. This is significantly faster than *in situ* methods which process millions of cells over a period of a few days (Feldman et al., 2019). For phenotypes as penetrant as mIFP expression, a library of 6092 sgRNAs with 2 sgRNA/group can be successfully screened with a single replicate. A genome scale screen of such a phenotype can be executed within 3 days (time of image analysis and photo activation). Even for more complex phenotypes, such as the nuclear size screen described here, a genomic screen can be finished within 2 weeks using the “Top5” sgRNA library and 2 runs. This time can be shortened with further optimization such as

the use of a microscope with a larger field of view, a lower magnification objective, optimization of imaging analysis algorithms, *etc.*

Optical enrichment screening also is possible for phenotypic screens with relatively low hit rates (defined as the fraction of all genes screened that are true hits). The ability to detect hits at low hit rates in our method depends on multiple factors, including: 1) the penetrance of the phenotype; 2) cellular fitness effect of the phenotype; 3) detection and photo-activation accuracy of the phenotype; 4) limitations imposed by FACS recovery and sequencing sample preparations of low cell numbers. The first three factors vary with the phenotype of interest. We optimized the genomic DNA preparation protocol (Methods), and are now able to process sequencing samples from a few thousand cells, enabling screens of low hit rate phenotypes. In our nuclear size screen, more than 1.5 millions cells were analyzed during each run with 2000-4000 cells recovered after FACS sorting. The hit rate of this screen was 2.76%, similar to optical CRISPR screens performed in an arrayed format (Groot et al., 2018), demonstrating the possibility to apply our approach to investigate phenotypes with low hit rates.

Our optical-enrichment screening approach can screen for multiple phenotypes simultaneously by using different photo-activation times. With PA-mCherry, we show that 4 distinct phenotypes could be potentially sorted (**Fig. 3.2b**). We demonstrate this in practice by differential photo-activation of true positive and negative cells to improve screening sensitivity. However, differential time of photo-activation could also be applied to analyze different phenotypes. This approach can be further developed by combining multiple photo-activatable fluorescent proteins in each cell.

In our study, optical enrichment was utilized for pooled CRISPR screens on phenotypes identifiable through microscopy. However, optical enrichment can be used for other purposes, as demonstrated previously (Hasle et al., 2020). In a recent study by Hasle *et al.*, the process of separating cells by FACS after optical enrichment was termed “visual cell sorting”. This method was used to evaluate hundreds of nuclear localization sequence variants in a pooled format and

to identify transcriptional regulatory pathways associated with paclitaxel resistance using single cell sequencing, demonstrating the broad applicability and power of this approach beyond CRISPR screening.

Our approach has limitations. Phenotypes of interest should be identifiable under the microscope and generally require fluorescent labeling. Commonly used fluorescence microscopes use four channels for fluorescent imaging with little spectral overlap: blue, green, red and far red. In our study, the red channel was occupied by cell labeling with PA-mCherry and the blue channel was used to estimate sgRNA transduction efficiency. Since sgRNA transduction efficiency can be measured by other approaches, the blue channel could be used together with the remaining two channels to label cellular structures. Combining bright field imaging with deep learning can be used to reconstruct the localization of fluorescent labels (Ounkomol et al., 2018), making it possible to use bright field imaging to further expand the phenotypes that can be studied with our technique.

Another limitation is the computational cost. Phenotypes were identified directly after imaging, thus the analysis code has to be fast and robust. In our study, the code identified phenotypes within a few hundred milliseconds. Each phenotype required writing specific image analysis code. This requirement can be overcome by implementing other image analysis strategies, including trainable machine learning, combining with existing image analysis software like CellProfiler, *etc.*, which will benefit laboratories that do not have the expertise to develop custom image processing code. Additionally, our approach is currently not compatible with fixed cells, thus transient phenotypes might be difficult to capture. However, we expect this to be solvable by further optimizing our screening pipeline to make it possible to prepare sequencing samples after fixation.

Optical enrichment compared to other methods for phenotypic screening

Two other methods have recently been developed to use imaging both for phenotypic screening and decoding to permit sgRNA identification in individual cells *in situ* (Feldman et al., 2019; Wang et al., 2019). In both methods, CRISPR sgRNA expression constructs were modified to express both a sgRNA and a barcode. The barcode can be read out either by *in situ* sequencing (Feldman et al., 2019) or sequential fluorescence *in situ* hybridization (FISH) (Wang et al., 2019). Both methods require sgRNAs to be re-barcoded necessitating *de-novo* design and library re-synthesis preventing reuse of most existing sgRNA libraries. In addition, cells need to be fixed, preventing further cell-based assays of the identified cells. Most importantly, both of these methods cannot easily scale to the whole genome because of barcoding limitations and the long imaging time required.

Another newly published method, similar to ours, also uses high throughput sequencing as an end point assay. Instead of using FACS to enrich cells of interest, this method cultures cells on microcraft arrays (magnetic polystyrene particles designed to capture single clones) to enable cell isolation as separate clones (CRaft-ID) (Wheeler et al., 2020). This method also can use most available sgRNA libraries and is compatible with further live cell studies. However, it is difficult to perform a genome wide screen with CRaft-ID, since it requires single cell isolation during cell culture and thus limits the number of cells that can be screened (6000 colonies/array). In addition, CRaft-ID can not be used to screen for phenotypes that cause defects in monoclonal growth, including essential genes. Our assay, on the other hand, provides an option for genome-wide screens and allows study of genes essential to growth.

Conclusion

In summary, our data demonstrate the power of our optical enrichment based pooled CRISPR screening method to study previously inaccessible phenotypes with high efficiency and accuracy. This method is simple, fast, uses open source software, and can be applied to

commercial or institutional genome-scale CRISPR sgRNA libraries. A digital micromirror device is required, but this can be introduced into the light path of common commercial microscopes. This screening approach could be broadly applied across many biological phenotypes including morphological changes, sub-cellular organization and cellular dynamics. Pluggable image analysis code enables selection of any desired morphological phenotypes as long as fast and robust detection code can be created, which is an area suited for deep learning approaches. We anticipate that this screening approach can be integrated with other profiling technologies such as single cell sequencing, further expanding its application to other research fields.

Figures and figure legends

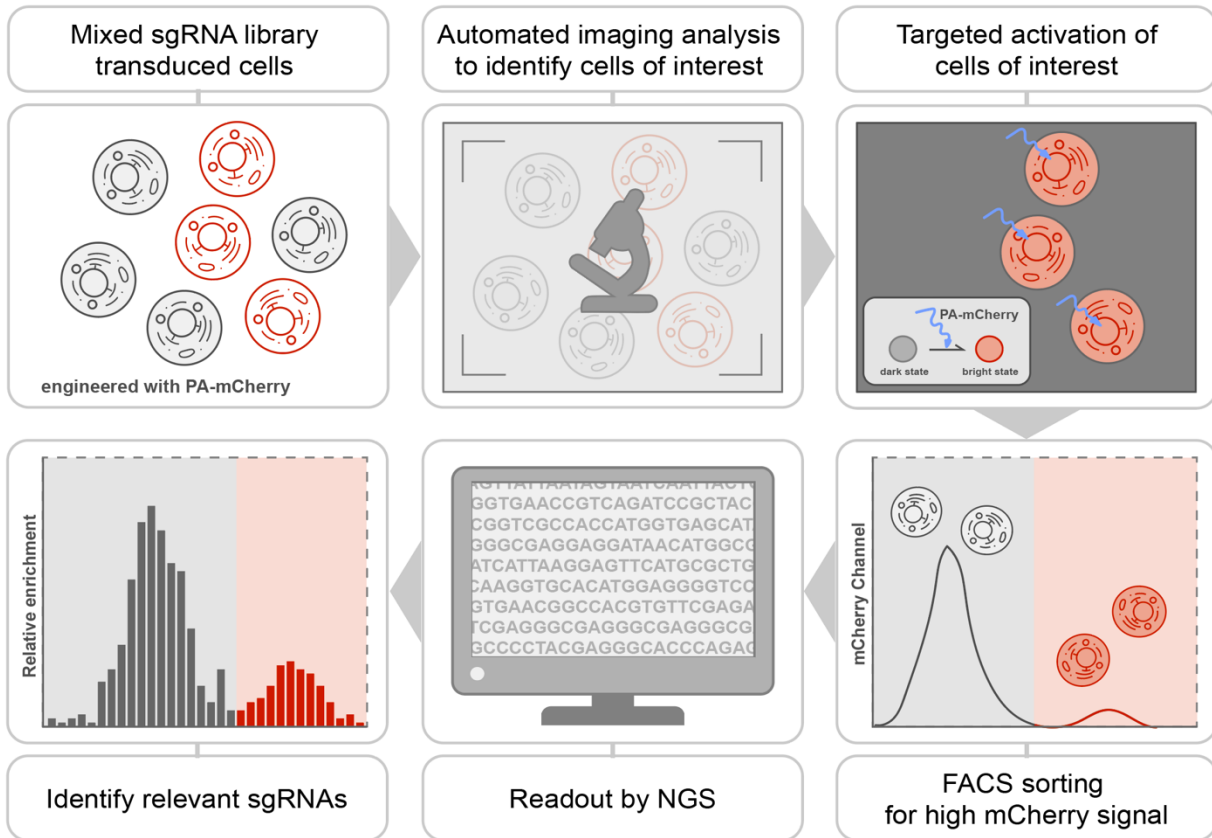


Fig. 3.1| Imaging-based pooled CRISPR screen. Schematic of imaging-based pooled CRISPR screen. Cells expressing PA-mCherry are infected with a pooled sgRNA library and imaged using a microscope. Images are collected and analyzed automatically to generate an activation mask targeting cells of interest. Exposure with blue light photo-activates cells of interest into mCherry positive cells that are subsequently isolated by FACS. Samples are analyzed by high-throughput sequencing for sgRNA identification.

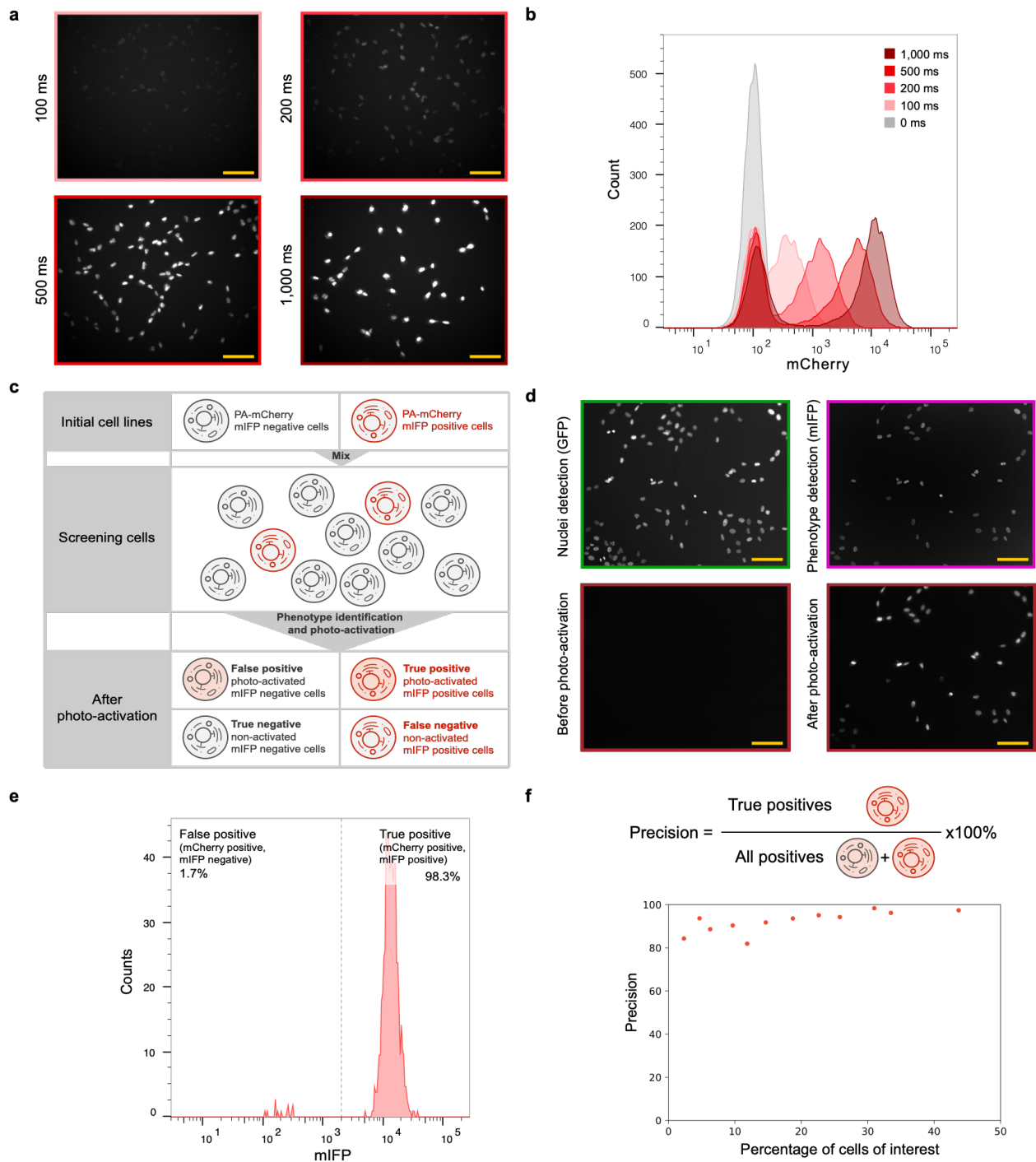


Fig. 3.2 | Optical enrichment enables accurate cell identification and isolation. (a-b) Cells can be activated into different fluorescent intensity levels that are clearly distinguished by FACS. Example images of cells (hTERT-RPE1 PA-mCherry) that have undergone various photo-activation times (a) and their corresponding FACS result (b) (number of cells analyzed for each condition is around 10,000, replicate number = 2) are shown. Scale bar: 100 μ m. (c) Schematic of the experiment to measure precision. mIFP positive (hTERT-RPE1 PA-mCherry H2B-mGFP mIFP-NLS) and mIFP negative cells (hTERT-RPE1 PA-mCherry H2B-mGFP) were mixed and

analyzed under GFP and mIFP channels separately. mIFP expression was used as a phenotype to indicate cells of interest (mIFP positive cells). An activation mask was generated for each mIFP positive cell. Cells identified by FACS to be mIFP and mCherry double positive are true positives while mCherry positive cells without mIFP fluorescence result from accidental activation (false positive cells). (d-e) Cells of interest can be identified by automated image analysis followed by photo-activation and distinguished through FACS with high accuracy. (d) Example images of image analysis (GFP channel, —; mIFP channel, —), before and after photo-activation (mCherry channel, —) are shown. Scale bar: 100 μm . (e) Example FACS data is shown (number of cells analyzed = 662). (f) Different ratios of mIFP positive cells and mIFP negative cells were mixed to measure precision at different percentages of hits.

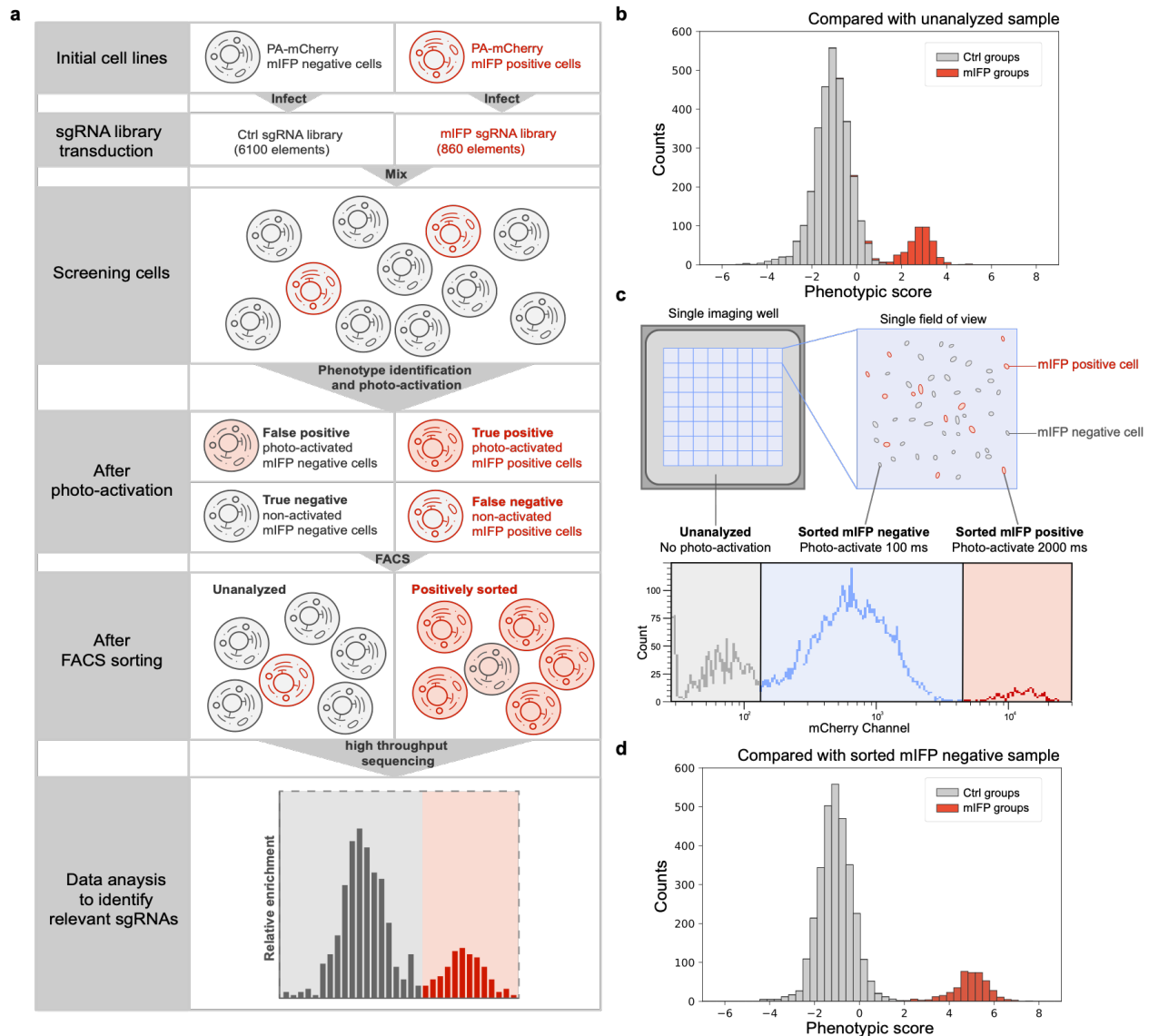


Fig. 3.3 | Optical enrichment enables accurate sgRNA identification in a pooled CRISPR screen. (a) Schematic of mIFP proof-of-principle screen. A mixed population of mIFP positive and negative cells was imaged and photo-activated as described in **Fig. 3.2c**. mCherry positive and unanalyzed cells were then isolated by FACS and prepared for high throughput sequencing to extract sgRNA information. (b) Screening result presented by distribution of phenotypic scores of all the sgRNA groups. Red and Grey: mIFP groups and Ctrl groups calculated by comparing with unanalyzed sample. (c) Schematic of dual-activation experiment. Experiment as described in **Fig. 3.3a** but mIFP negative cells were also photo-activated (100 ms). mIFP positive cells were activated using a longer activation time (2000 ms) to guarantee a clear distinction by FACS. Image acquisition generally does not cover the complete imaging well which leaves cells not imaged and unanalyzed. Lower panel shows an example of FACS data. Cells sorted for mIFP expression (sorted mIFP positive), cells sorted for lack of mIFP (sorted mIFP negative) and unanalyzed cells were separately collected and prepared for high throughput sequencing. (d) Distribution of phenotypic scores of all the sgRNA groups compared with the sorted mIFP negative sample. Phenotype identification is improved by comparing with true negative cells

rather than unanalyzed cells as shown in **Fig. 3.2b**. Red and Grey: mIFP groups and Ctrl groups calculated by comparing with sorted mIFP negative sample.

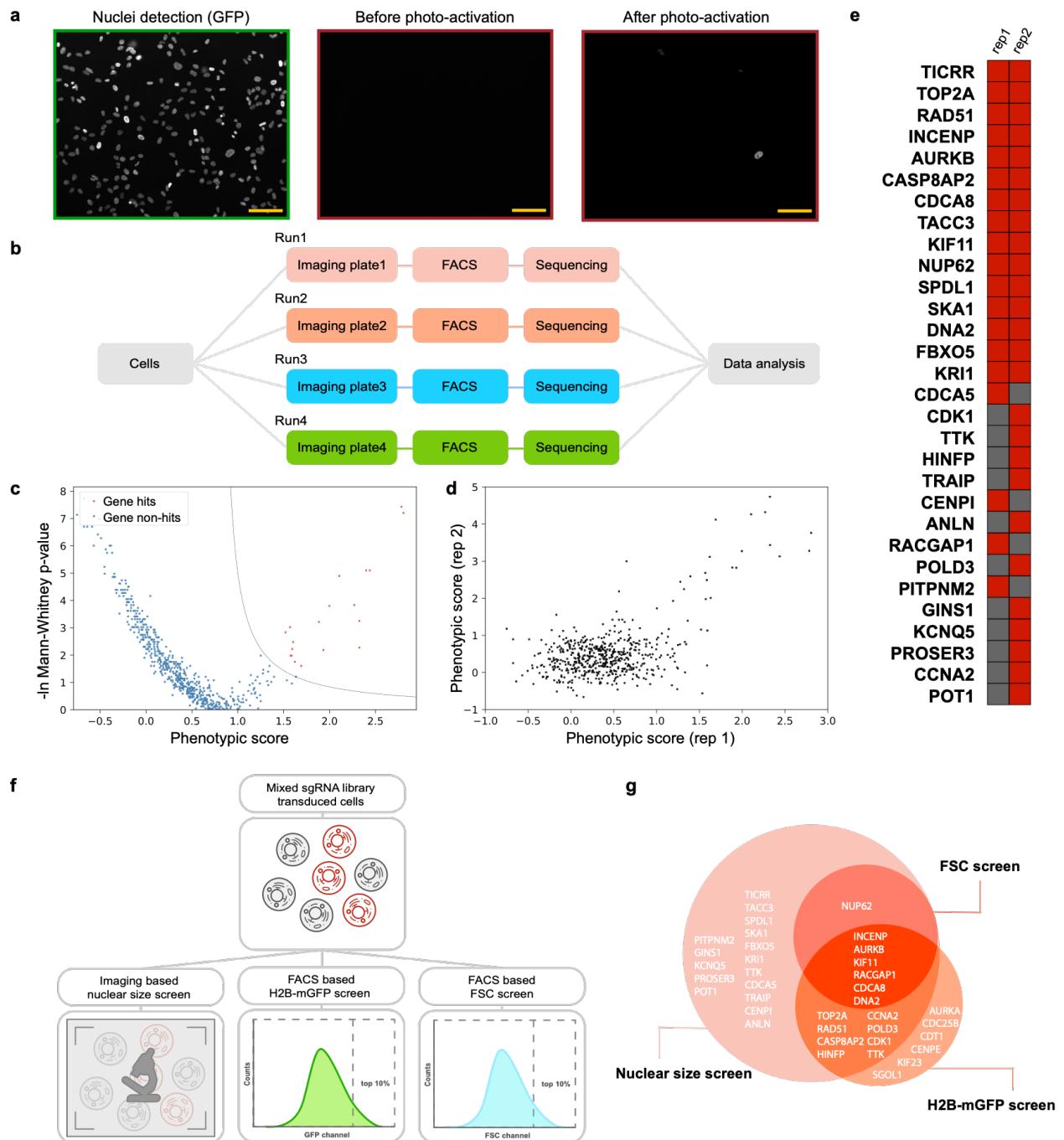


Fig. 3.4 | Screens for nuclear size regulators. (a) Example images of the nuclear size screen. Cells (hTERT-RPE1 dCas9-KRAB-BFP PA-mCherry H2B-mGFP) were transduced with a CRISPRi sgRNA library and imaged under the GFP channel. Cells with nuclei larger than $1000 \mu\text{m}^2$ were photo-activated, sorted and analyzed by deep sequencing. Example images of nuclei, (GFP channel, —), before and after photo-activation (mCherry channel, —) are shown. Note that the example images were from experiments done with dual-activation setup as described in Fig. 3.3c. Background cells with low fluorescence intensity in mCherry channel after photo-activation were true negative cells that were photo-activated with a shorter exposure time (200ms). Scale bar: $100 \mu\text{m}$. (b) Work flow of one replicate of the nuclear size screen. For each replicate,

transduced cells were seeded into 4 glass-bottom imaging plates. Cells in each single imaging plate were imaged, analyzed, photo-activated, sorted and sequenced separately, termed as separate runs. (c) Screening result of one replicate shown in volcano plot. The corresponding eFDR- η curve is shown in **Fig. S3.3b** and the other replicate is shown in **Fig. S3.3c**. (d) Comparison between two replicates. (e) List of genes identified from two replicates. Red: hit. Grey: non-hit. (f) Work flow of three screens, namely nuclear size screen, H2B-mGFP screen and FSC screen. After transducing the sgRNA library, cells were split and prepared for either image analysis (nuclear size screen) or FACS sorting (H2B-mGFP screen and FSC screen). (g) Comparison of hits identified in FSC screen, H2B-mGFP screen and nuclear size screen.

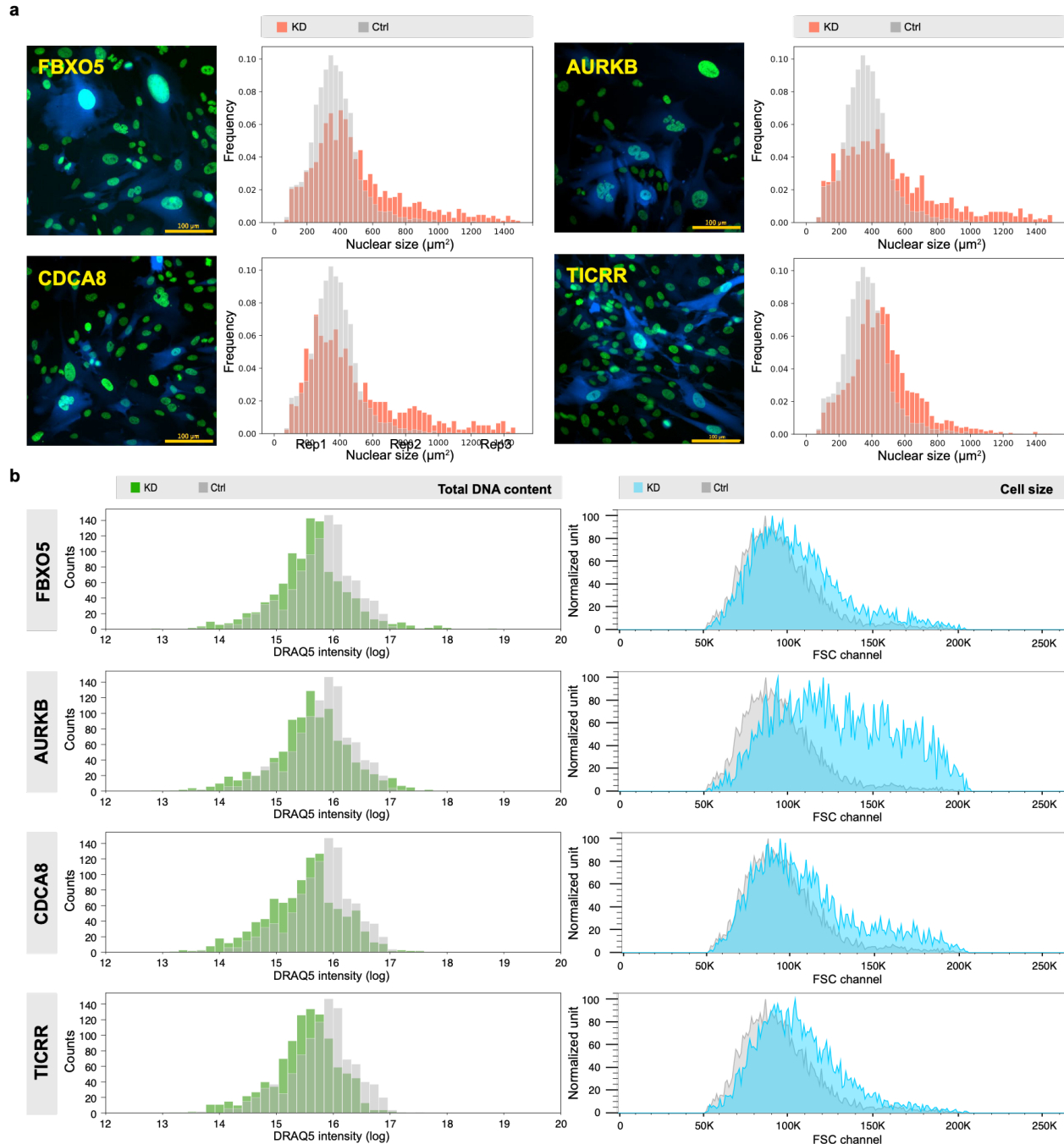


Fig. 3.5 | Characterization of hits identified in nuclear size screen. (a) Each hit identified in both replicates was verified under the microscope after infecting with a mixture of 3-4 sgRNA constructs targeting the gene ($n=3$). Cells were puromycin selected for 2 days before imaging. Example images of 4 hits and their distribution of nuclear sizes from one replicate are shown in (a), all the others are listed in **Fig. S3.6**. For each gene in each replicate, at least 1,000 cells were analyzed using the Auto-PhotoConverter μ Manager plugin. The cell population is heterogeneous due to inefficient knockdown, incomplete puromycin selection, and penetrance of the phenotype. A BFP was expressed from the same sgRNA construct. Only cells with high BFP intensity, indicating successful sgRNA transduction, were included for data analysis as described in Methods. Red, nuclear size distribution of corresponding gene after knockdown; Grey, nuclear size distribution of cells infected with non-targeting control sgRNAs. (b) Some cells

developed a larger cellular size while maintained a similar DNA content level after knock down. For DNA content measurement, cells were infected with the same 3-4 sgRNAs/gene, puromycin selected for 2 days and stained with 5 μ M DRAQ5 for 1 hr before imaging (1,000 cells were analyzed for each gene). For cellular size measurement, cells were infected with the same 3-4 sgRNAs/gene and puromycin selected for 3 days before FACS analysis (at least 2787 cells were analyzed for each gene). Example imaging analysis data and FACS data of the same 4 hits are shown in (b) and all the others are shown in **Fig. S3.6**. Green, distribution analysis of DRAQ5 staining intensity after knockdown of corresponding gene; Blue, FACS of FSC signal after knockdown of corresponding gene; Grey, distribution analysis of DRAQ5 staining intensity or FSC signal of cells infected with non-targeting control sgRNAs.

Supplemental figures and table legends

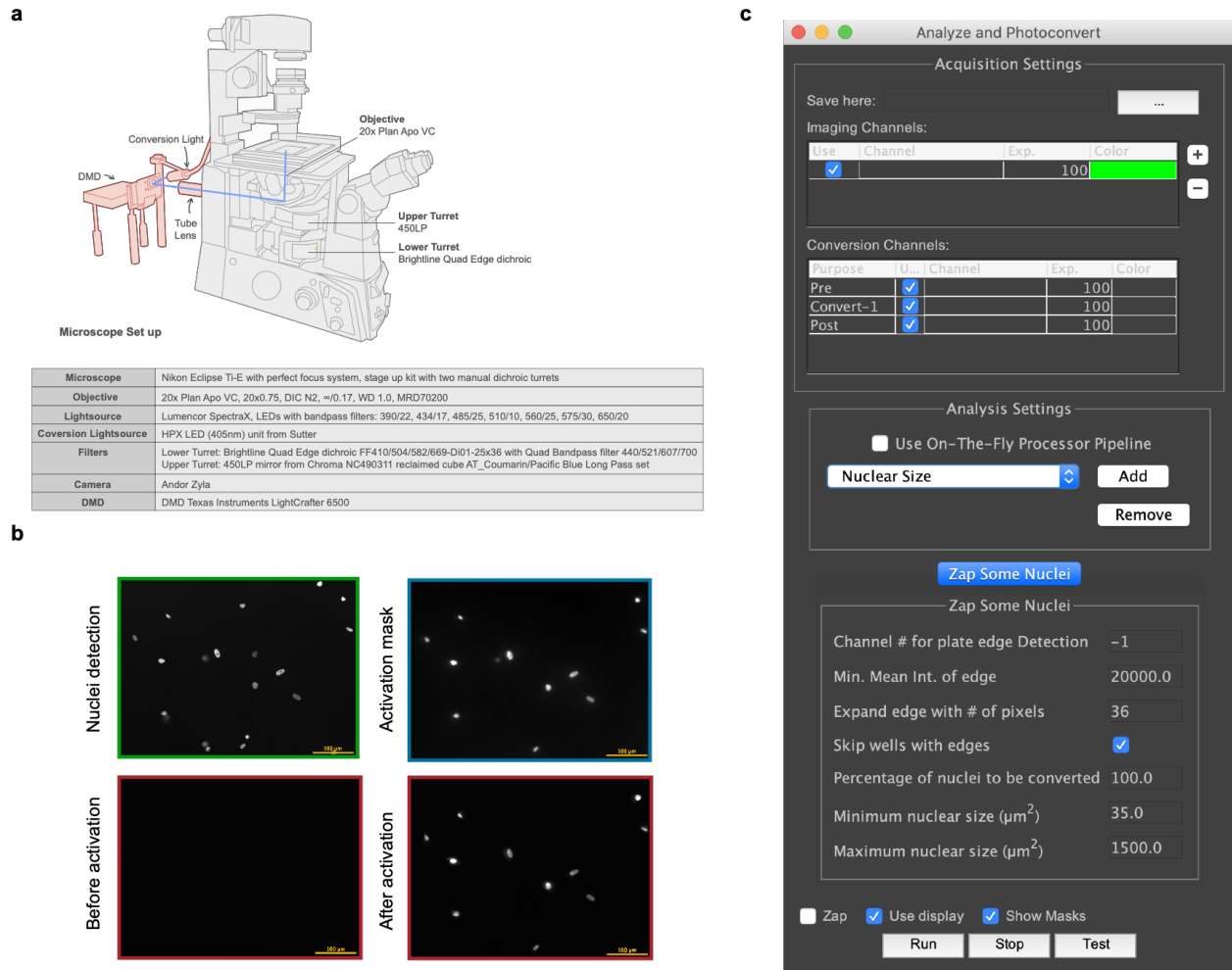


Fig. S3.1 | Microscope and μ Manager plugin for photo-activation experiments. (a) A digital micro-mirror device (DMD) and a blue LED (centered around 405nm) light source were engineered on a Nikon Eclipse Ti-E microscope as shown in the figure. A computer was used to control the DMD which reflects light into the microscope only when pixels are in the “on” position, so displaying a mask matching the cell photo-activates that cell. (b) Example images of a photo-activation experiment. Cells (hTERT-RPE1 PA-mCherry H2B-mGFP) are shown imaged in the GFP channel (—), during photo-activation (Blue light channel, —), and before and after photo-activation (mCherry channel, —). Scale bar: 100 μm . (c) A μ Manager plugin was developed to enable automatic image acquisition, analysis and photo-activation. An analysis plugin defines its own set of parameters that can be manipulated by the user. Two analysis plugins were used in this study: one for cell identification and another for nuclear size measurement.

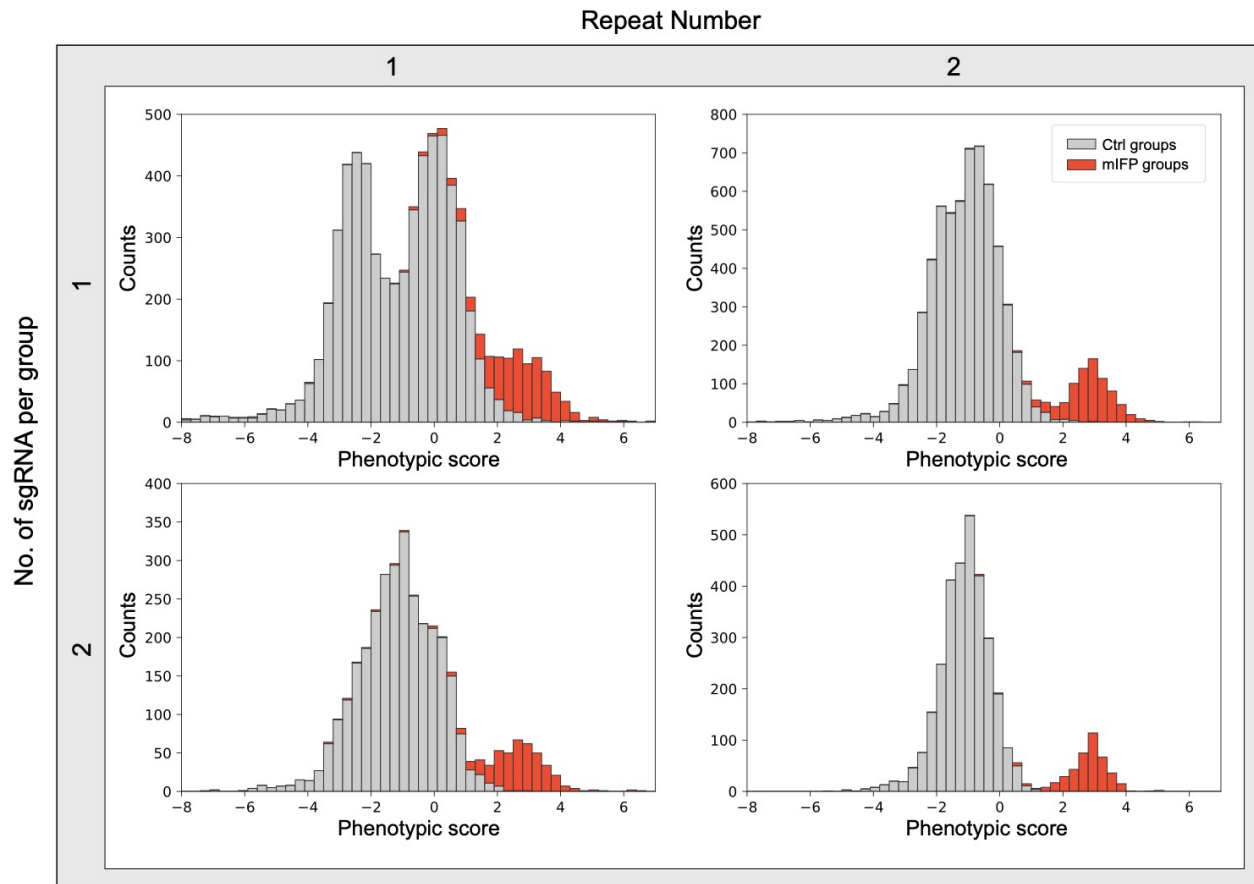


Fig. S3.2 | Library composition and number of replicates affect screening results. Distribution of phenotypic scores of all sgRNA groups in 4 different analysis modes. Phenotypic score of a sgRNA group was calculated based on the average phenotypic scores from: Top left: a single sgRNA from a single replicate; Top right: a single sgRNA averaged between 2 replicates; Bottom left: 2 sgRNAs from a single replicate; Bottom right: 2 sgRNAs averaged between 2 replicates.

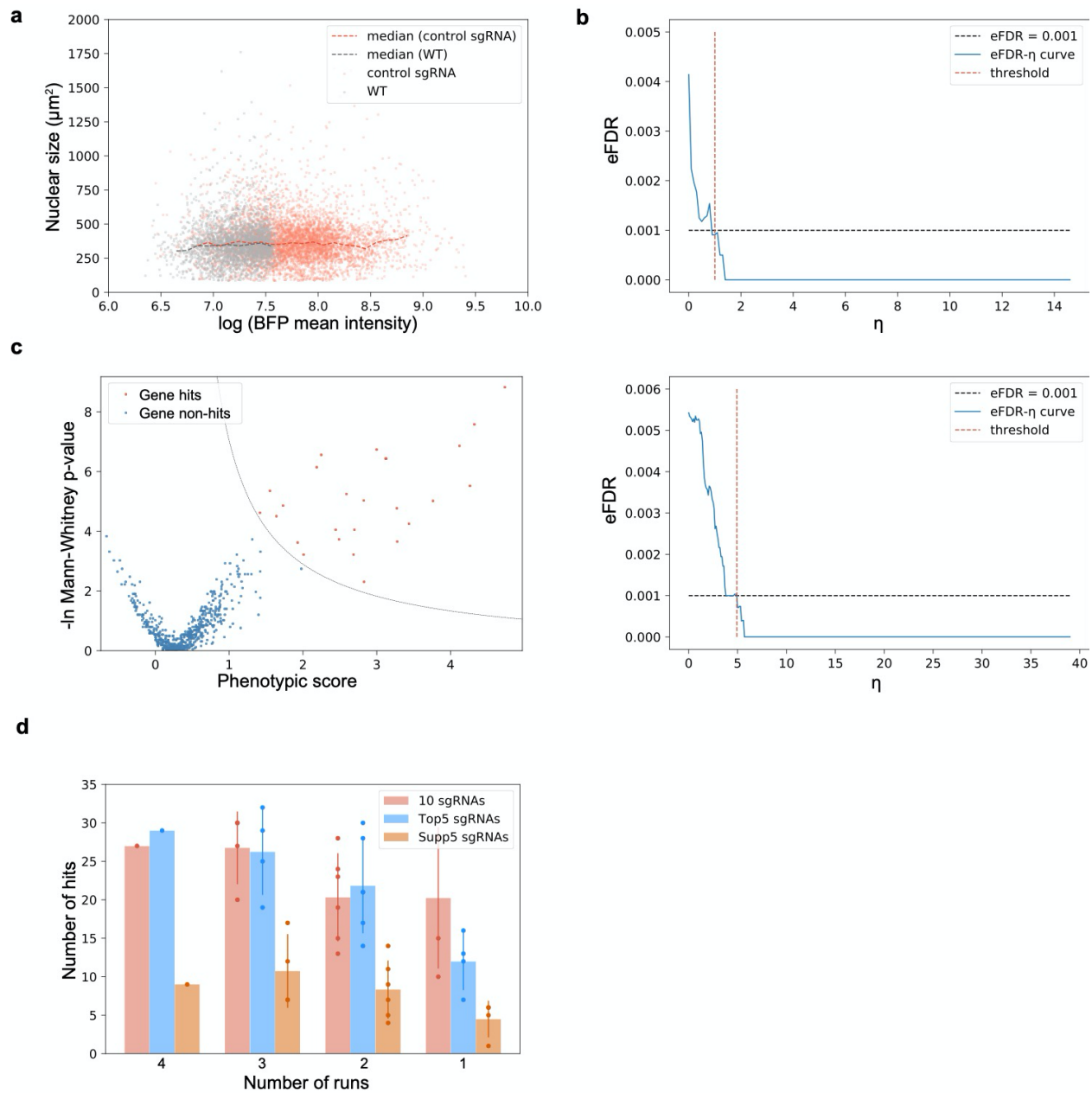


Fig. S3.3 | Screens for nuclear size regulators. (a) sgRNA transduction results in cells with higher BFP intensity and negative control sgRNAs do not affect nuclear size after viral infection. Two negative control sgRNAs were designed to have no target sites in the human genome. Cells (hTERT-RPE1 dCas9-KRAB-BFP PA-mCherry H2B-mGFP) underwent viral transduction and puromycin selection for 3 days before imaging. Both wild type cells without viral transduction (grey dots) and cells infected with negative control sgRNAs (red dots) were seeded into a 96-well glass bottom imaging dish. Images were collected for cells with/without sgRNA viral transduction and both nuclear size and mean BFP intensity of each nucleus were analyzed using the Auto-PhotoConverter μ Manager plugin (number of wild type cells analyzed = 2756; number of negative control sgRNA infected cells analyzed = 5653). Besides the BFP expressed from the dCas9 construct, another BFP was encoded on the sgRNA construct and higher BFP intensity was used as a marker for successful infection. The boundary measured from comparison between sgRNA

infected cells and wild type cells: $\ln(\text{mean BFP intensity}) = 7.6$ was also used as a threshold to determine which cells were successfully transduced with sgRNA. Analysis from imaging data show no correlation between nuclear size and BFP intensity. (b) eFDR- η curve for screening result shown in **Fig. 3.4c**. A score η summarizing effects from both severity of the phenotype (phenotypic score) as well as trustworthiness of the phenotype ($-\ln(p \text{ value})$) was calculated for each gene and simulated negative control. The optimal cut-off for score η (red dotted line) was determined by calculating an empirical false discovery rate (eFDR) at multiple values of η as the number of simulated negative controls with score η higher than the cut-off (false positives) divided by the sum of genes and simulated negative controls with score η higher than the cut-off (all positives). The cut-off score η resulting in an eFDR of 0.1% (black dotted line) was used to call hits for further analysis. (c) Screening result of the other replicate shown in volcano plot and its corresponding eFDR- η curve as described in **Fig. S3.3b**. (d) Number of hits identified using data averaging using different numbers of runs and/or different library compositions. Error bar: standard deviation.

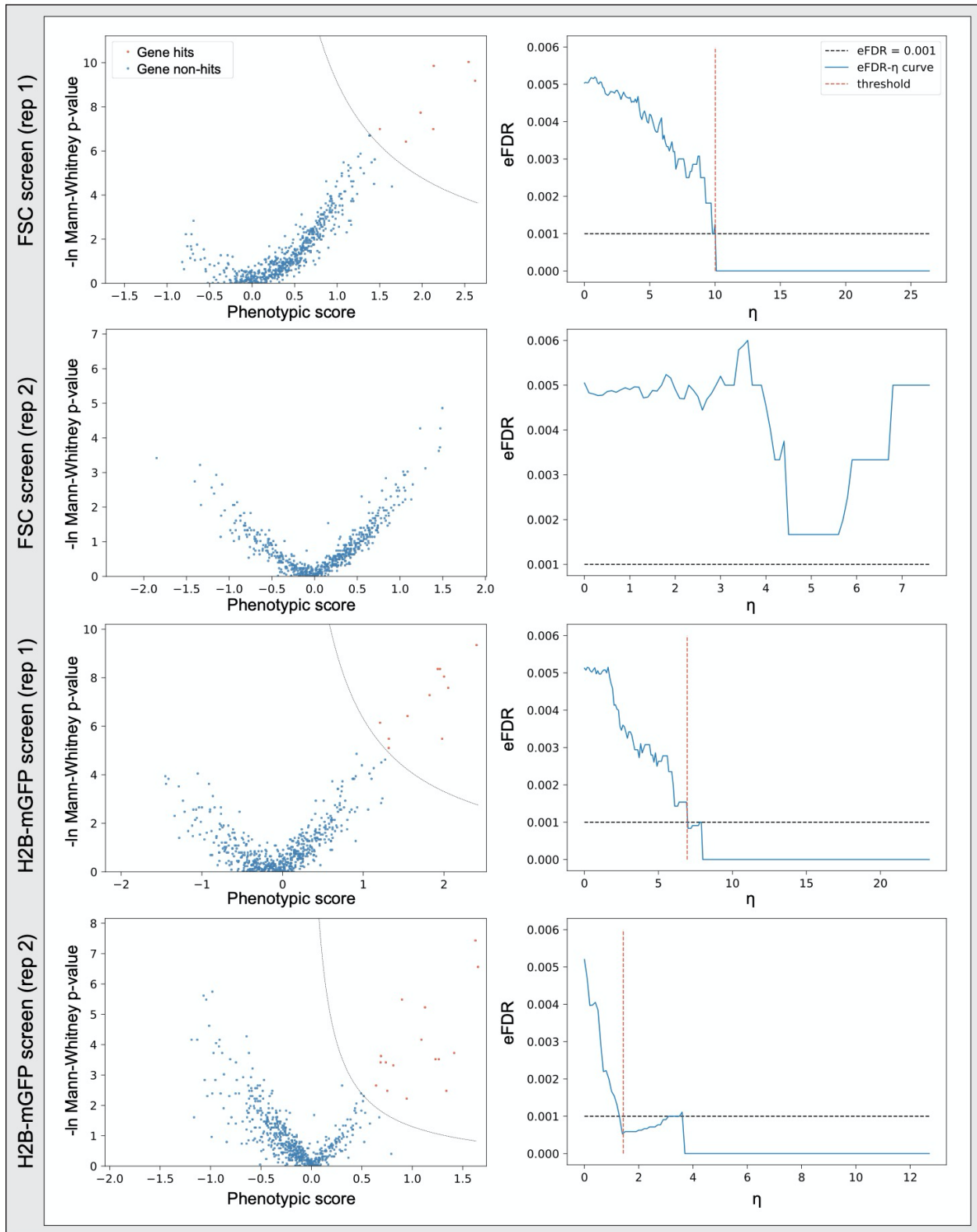
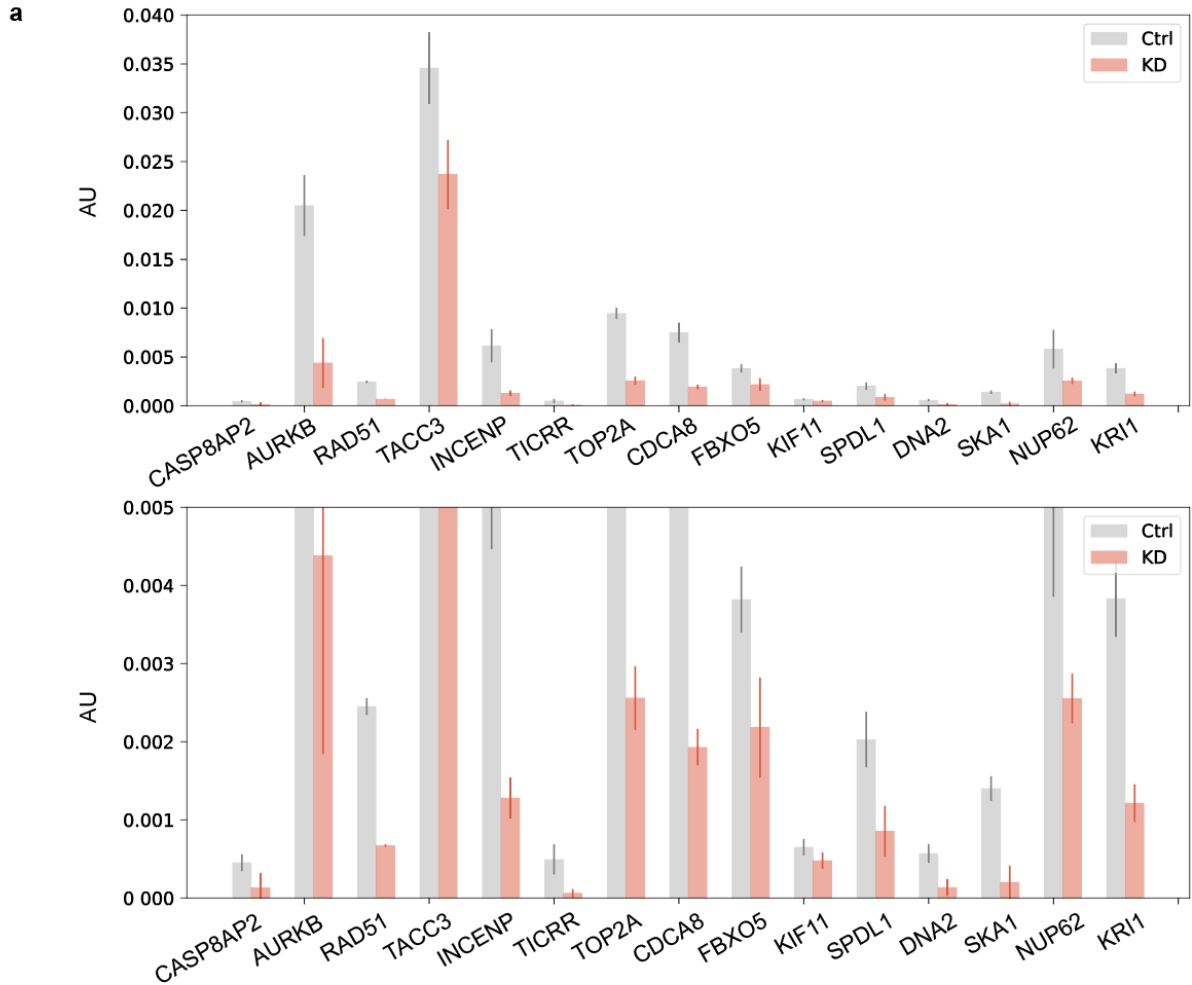


Fig. S3.4 | Screen results of FSC and H2B-mGFP screens. Cells (hTERT-RPE1 dCas9-KRAB-BFP PA-mCherry H2B-mGFP) were infected and puromycin selected for 3 days. The top 10 percentile of cells based on either GFP fluorescence or FSC signal were separately sorted and

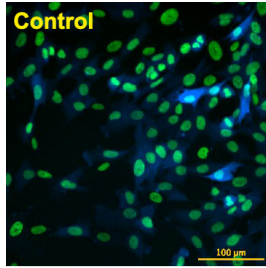
prepared for high throughput sequencing. Screen results shown in volcano plot and their corresponding eFDR- η curve of two replicates are shown as described in **Fig. S3.3b**.



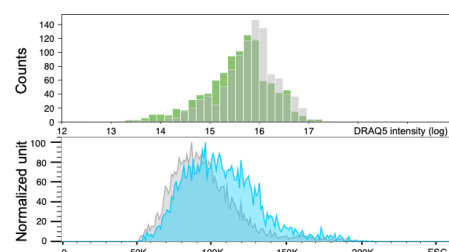
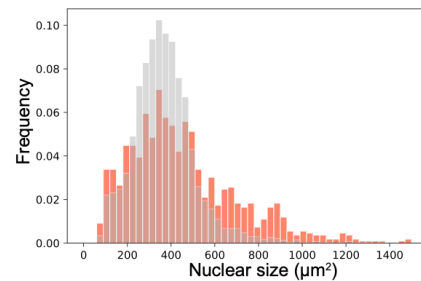
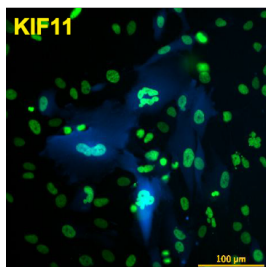
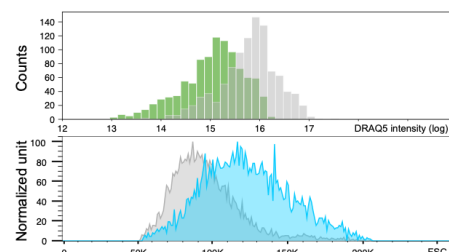
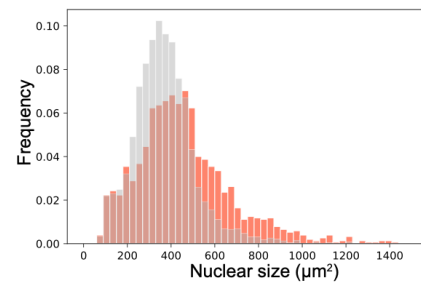
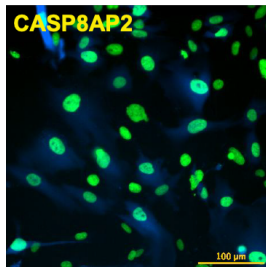
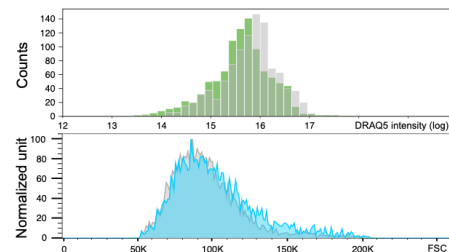
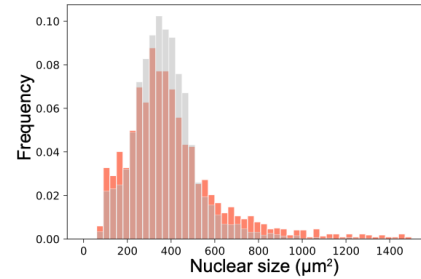
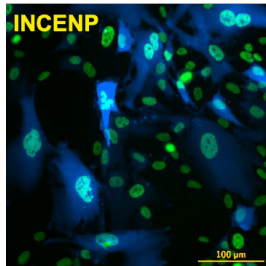
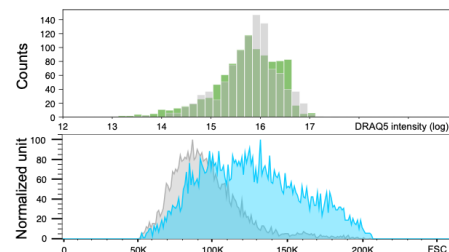
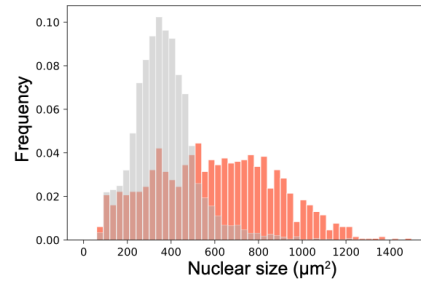
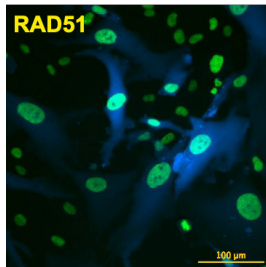
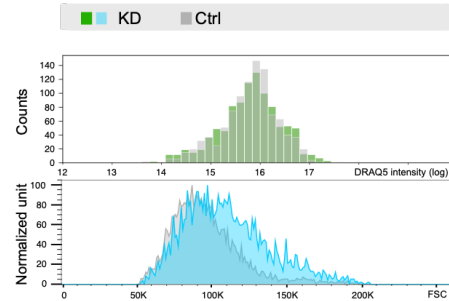
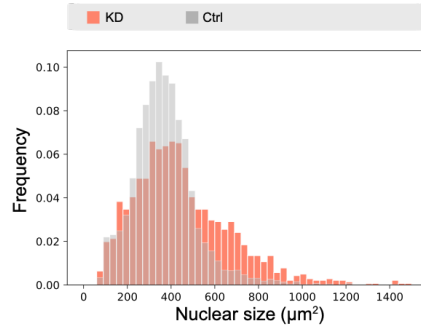
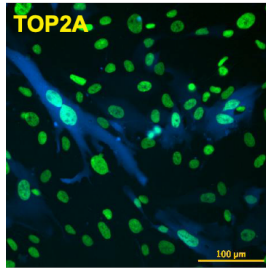
b

Gene	KD percentage	Infection percentage	KD efficiency
CASP8AP2	70.60	67.3	105
AURKB	78.63	51.1	154
RAD51	72.69	67.9	107
TACC3	31.50	85.7	36.8
INCENP	79.14	73.0	108
TICRR	87.27	80.9	108
TOP2A	72.93	74.0	98.5
CDCA8	74.29	66.9	111
FBXO5	42.86	76.6	55.95
KIF11	26.37	69.7	37.8
SPDL1	57.77	77.3	74.7
DNA2	76.25	79.5	95.9
SKA1	85.56	80.2	107
NUP62	55.97	64.5	86.8
KRI1	68.29	82.4	82.9

Fig. S3.5 | Measurement of knock down efficiency for hit verification. (a) RT-qPCR results of all the hits identified in both replicates after knock down. An enlarged graph of the bottom part of the original graph was also included. Cells (hTERT-RPE1 dCas9-KRAB-BFP PA-mCherry H2B-mGFP) were infected with corresponding sgRNAs and puromycin selected for 3 days before harvesting. Harvested cells were split and one was used for RNA extraction and RT-qPCR analysis to measure the percentage of knock down (KD percentage), while the other half was used for FACS analysis to measure the percentage of sgRNA infection (Infection percentage) (**Fig. S3.5b**). ACTB was used as an internal control to normalize the variability on expression levels. Error bar: standard deviation between triplicates. (b) Knock down efficiency (KD efficiency) of all the hits identified in both replicates. Knock down percentage (KD percentage) was measured based on RT-qPCR results. BFP was co-expressed on the sgRNA construct and only cells with BFP intensity above a threshold value that determined by control cells were considered successfully infected cells. Percentage of successful infection (Infection percentage) was measured by FACS and for each gene, KD efficiency was calculated using KD percentage divided by its corresponding infection percentage.



- Nuclear size distribution, DRAQ5 staining intensity or FSC signal of cells infected with non-targeting control sgRNAs
- Nuclear size distribution of corresponding gene after knockdown
- DRAQ5 staining intensity after knockdown of corresponding gene
- FACS of FSC signal after knockdown of corresponding gene



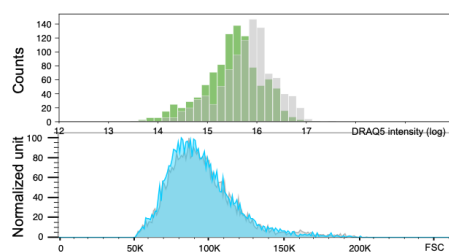
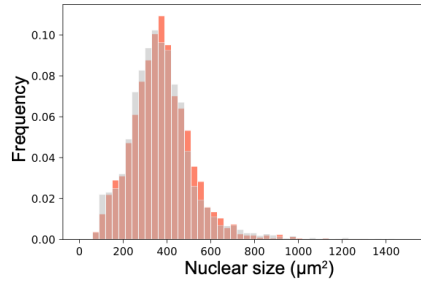
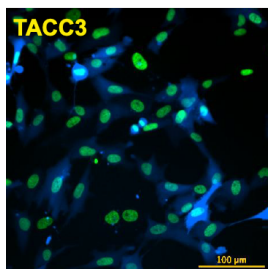
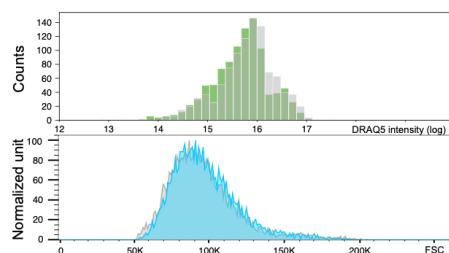
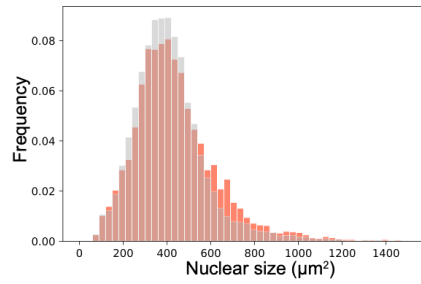
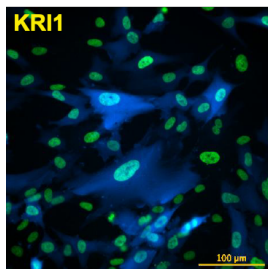
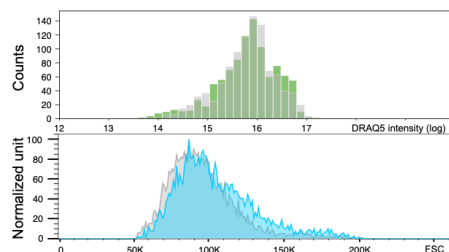
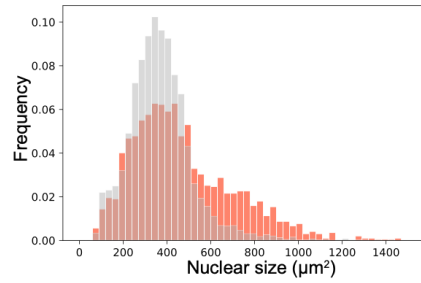
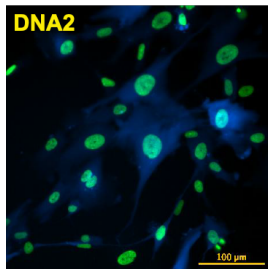
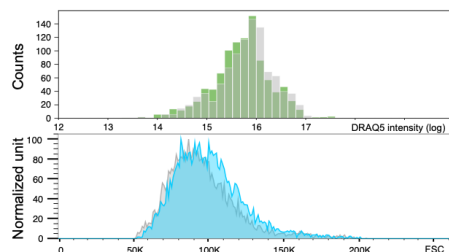
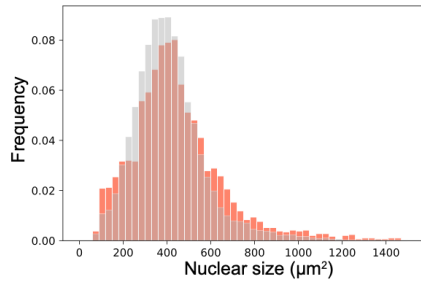
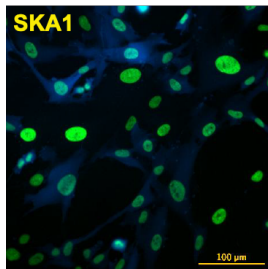
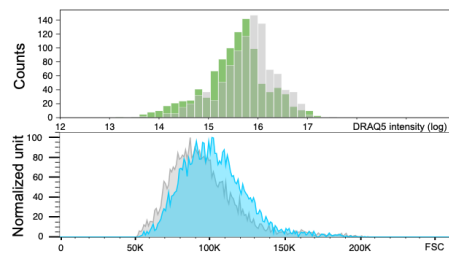
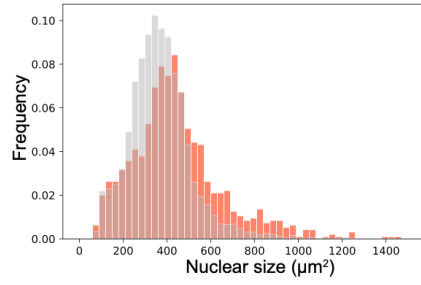
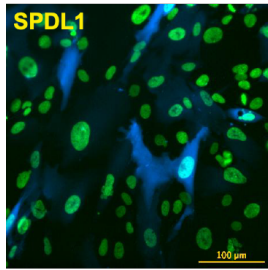
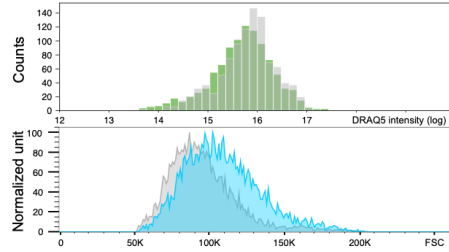
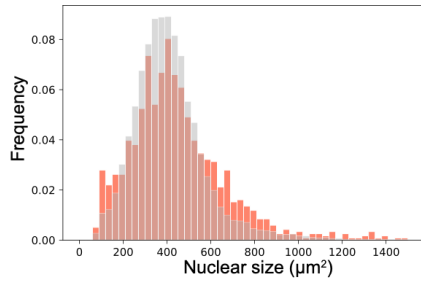
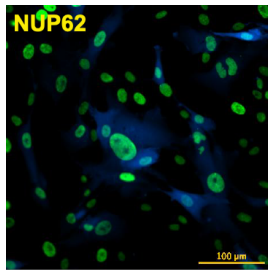


Fig. S3.6 I Characterization of hits identified in both replicates of the nuclear size screens. Example images, distribution of nuclear size (at least 1,000 cells were analyzed for each gene, replicate number = 3), distribution analysis data of DRAQ5 staining fluorescence (1,000 cells were analyzed for each gene) and FACS of FSC distribution of each hit (at least 2787 cells were analyzed for each gene) (other than the 4 shown in **Fig. 3.5**) after knock down are shown.

Table S3.1 | Description of shared hits (summary from genecards.org, SGD and cited literature)

	Aliases	Summaries
TICRR	TOPBP1 Interacting Checkpoint And Replication Regulator; Treslin	Regulator of DNA replication and S/M and G2/M checkpoints. Regulates the triggering of DNA replication initiation via its interaction with TOPBP1 by participating in CDK2-mediated loading of CDC45L onto replication origins. Required for the transition from pre-replication complex (pre-RC) to pre-initiation complex (pre-IC). Required to prevent mitotic entry after treatment with ionizing radiation.
TOP2A	DNA Topoisomerase II Alpha	Control of topological states of DNA by transient breakage and subsequent rejoining of DNA strands. Topoisomerase II makes double-strand breaks. Essential during mitosis and meiosis for proper segregation of daughter chromosomes.
RAD51	RAD51 Recombinase	Plays an important role in homologous strand exchange, a key step in DNA repair through homologous recombination (HR) (PubMed:28575658). Binds to single and double-stranded DNA and exhibits DNA-dependent ATPase activity. Catalyzes the recognition of homology and strand exchange between homologous DNA partners to form a joint molecule between a processed DNA break and the repair template. Binds to single-stranded DNA in an ATP-dependent manner to form nucleoprotein filaments which are essential for the homology search and strand exchange (PubMed:26681308). Part of a PALB2-scaffolded HR complex containing BRCA2 and RAD51C and which is thought to play a role in DNA repair by HR. Plays a role in regulating mitochondrial DNA copy number under conditions of oxidative stress in the presence of RAD51C and XRCC3. Also involved in interstrand cross-link repair (PubMed:26253028).
INCENP	Inner Centromere Protein	Component of the chromosomal passenger complex (CPC), a complex that acts as a key regulator of mitosis. The CPC complex has essential functions at the centromere in ensuring correct chromosome alignment and segregation and is required for chromatin-induced microtubule stabilization and spindle assembly. Acts as a scaffold regulating CPC localization and activity. The C-terminus associates with AURKB or AURKC, the N-terminus associated with BIRC5/survivin and CDCA8/borealin tethers the CPC to the inner centromere, and the microtubule binding activity within the central SAH domain directs AURKB/C toward substrates near microtubules (PubMed:15316025, PubMed:12925766, PubMed:27332895). The flexibility of the SAH domain is proposed to allow AURKB/C to follow substrates on dynamic microtubules while ensuring CPC docking to static chromatin (By similarity). Activates AURKB and AURKC (PubMed:27332895). Required for localization of CBX5 to mitotic centromeres (PubMed:21346195). Controls the kinetochore localization of BUB1 (PubMed:16760428).
AURKB	Aurora Kinase B	Serine/threonine-protein kinase component of the chromosomal passenger complex (CPC), a complex that acts as a key regulator of mitosis. The CPC complex has essential functions at the centromere in ensuring correct chromosome alignment and segregation and is required for chromatin-induced microtubule stabilization and spindle assembly. Involved in the bipolar attachment of spindle microtubules to kinetochores and is a key regulator for the onset of cytokinesis during mitosis. Required for central/midzone spindle assembly and cleavage furrow formation. Key component of the cytokinesis checkpoint, a process required to delay abscission to prevent both premature resolution of intercellular chromosome bridges and accumulation of DNA damage: phosphorylates CHMP4C, leading to retain abscission-competent VPS4 (VPS4A and/or VPS4B) at the midbody ring until abscission checkpoint signaling is terminated at late cytokinesis (PubMed:22422861, PubMed:24814515). AURKB phosphorylates the CPC complex subunits BIRC5/survivin, CDCA8/borealin and INCENP. Phosphorylation of INCENP leads to increased AURKB activity. Other known AURKB substrates involved in centromeric functions and mitosis are CENPA, DES/desmin, GPAF, KIF2C, NSUN2, RACGAP1, SEPTIN1, VIM/vimentin, HASPIN, and histone H3. A positive feedback loop involving HASPIN and AURKB contributes to localization of CPC to centromeres. Phosphorylation of VIM controls vimentin filament segregation in cytokinetic process, whereas histone H3 is phosphorylated at 'Ser-10' and 'Ser-28' during mitosis (H3S10ph and H3S28ph, respectively). A positive feedback between HASPIN and AURKB contributes to CPC localization. AURKB is also required for kinetochore localization of BUB1 and SGO1. Phosphorylation of p53/TP53 negatively regulates its transcriptional activity. Key regulator of active promoters in resting B- and T-lymphocytes: acts by mediating phosphorylation of H3S28ph at active promoters in resting B-cells, inhibiting RNF2/RING1B-mediated ubiquitination of histone H2A and enhancing binding and activity of the USP16 deubiquitinase at transcribed genes.
CASP8AP2	Caspase 8 Associated Protein 2	Participates in TNF-alpha-induced blockade of glucocorticoid receptor (GR) transactivation at the nuclear receptor coactivator level, upstream and independently of NF-kappa-B. Suppresses both NCOA2- and NCOA3-induced enhancement of GR transactivation. Involved in TNF-alpha-induced activation of NF-kappa-B via a TRAF2-dependent pathway. Acts as a downstream mediator for CASP8-induced activation of NF-kappa-B. Required for the activation of CASP8 in FAS-mediated apoptosis. Required for histone gene transcription and progression through S phase.

	Aliases	Summaries
CDCA8	Cell Division Cycle Associated 8; Borealin	Component of the chromosomal passenger complex (CPC), a complex that acts as a key regulator of mitosis. The CPC complex has essential functions at the centromere in ensuring correct chromosome alignment and segregation and is required for chromatin-induced microtubule stabilization and spindle assembly. Major effector of the TTK kinase in the control of attachment-error-correction and chromosome alignment.
KIF11	Kinesin Family Member 11; Eg5	Motor protein required for establishing a bipolar spindle during mitosis (PubMed:19001501). Required in non-mitotic cells for transport of secretory proteins from the Golgi complex to the cell surface (PubMed:23857769).
NUP62	Nucleoporin 62	Essential component of the nuclear pore complex (PubMed:1915414). The N-terminal is probably involved in nucleocytoplasmic transport (PubMed:1915414). The C-terminal is involved in protein-protein interaction probably via coiled-coil formation, promotes its association with centrosomes and may function in anchorage of p62 to the pore complex (PubMed:1915414, PubMed:24107630). Plays a role in mitotic cell cycle progression by regulating centrosome segregation, centriole maturation and spindle orientation (PubMed:24107630). It might be involved in protein recruitment to the centrosome after nuclear breakdown (PubMed:24107630).
SPDL1	Spindle Apparatus Coiled-Coil Protein 1; Spindly	Required for the localization of dynein and dynactin to the mitotic kintochore. Dynein is believed to control the initial lateral interaction between the kinetochore and spindle microtubules and to facilitate the subsequent formation of end-on kinetochore-microtubule attachments mediated by the NDC80 complex. Also required for correct spindle orientation. Does not appear to be required for the removal of spindle assembly checkpoint (SAC) proteins from the kinetochore upon bipolar spindle attachment (PubMed:17576797, PubMed:19468067). Acts as an adapter protein linking the dynein motor complex to various cargos and converts dynein from a non-processive to a highly processive motor in the presence of dynactin. Facilitates the interaction between dynein and dynactin and activates dynein processivity (the ability to move along a microtubule for a long distance without falling off the track) (PubMed:25035494).
SKA1	Spindle And Kinetochore Associated Complex Subunit 1	Component of the SKA1 complex, a microtubule-binding subcomplex of the outer kinetochore that is essential for proper chromosome segregation (PubMed:17093495, PubMed:19289083, PubMed:23085020). Required for timely anaphase onset during mitosis, when chromosomes undergo bipolar attachment on spindle microtubules leading to silencing of the spindle checkpoint (PubMed:17093495). The SKA1 complex is a direct component of the kinetochore-microtubule interface and directly associates with microtubules as oligomeric assemblies (PubMed:19289083). The complex facilitates the processive movement of microspheres along a microtubule in a depolymerization-coupled manner (PubMed:19289083). Affinity for microtubules is synergistically enhanced in the presence of the ndc-80 complex and may allow the ndc-80 complex to track depolymerizing microtubules (PubMed:23085020). In the complex, it mediates the interaction with microtubules (PubMed:19289083, PubMed:23085020).
DNA2	DNA Replication Helicase/Nuclease 2	Key enzyme involved in DNA replication and DNA repair in nucleus and mitochondrion. Involved in Okazaki fragments processing by cleaving long flaps that escape FEN1: flaps that are longer than 27 nucleotides are coated by replication protein A complex (RPA), leading to recruit DNA2 which cleaves the flap until it is too short to bind RPA and becomes a substrate for FEN1. Also involved in 5'-end resection of DNA during double-strand break (DSB) repair: recruited by BLM and mediates the cleavage of 5'-ssDNA, while the 3'-ssDNA cleavage is prevented by the presence of RPA. Also involved in DNA replication checkpoint independently of Okazaki fragments processing. Possesses different enzymatic activities, such as single-stranded DNA (ssDNA)-dependent ATPase, 5'-3' helicase and endonuclease activities. While the ATPase and endonuclease activities are well-defined and play a key role in Okazaki fragments processing and DSB repair, the 5'-3' DNA helicase activity is subject to debate. According to various reports, the helicase activity is weak and its function remains largely unclear. Helicase activity may promote the motion of DNA2 on the flap, helping the nuclease function.

	Aliases	Summaries
FBXO5	F-Box Protein 5; Emi1	Regulator of APC activity during mitotic and meiotic cell cycle (PubMed:17485488, PubMed:17234884, PubMed:17875940, PubMed:23708001, PubMed:23708605, PubMed:16921029). During mitotic cell cycle plays a role as both substrate and inhibitor of APC-FZR1 complex (PubMed:29875408, PubMed:17485488, PubMed:17234884, PubMed:17875940, PubMed:23708001, PubMed:23708605, PubMed:16921029). During G1 phase, plays a role as substrate of APC-FZR1 complex E3 ligase (PubMed:29875408). Then switches as an inhibitor of APC-FZR1 complex during S and G2 leading to cell-cycle commitment (PubMed:29875408). As APC inhibitor, prevents the degradation of APC substrates at multiple levels: by interacting with APC and blocking access of APC substrates to the D-box coreceptor, formed by FZR1 and ANAPC10; by suppressing ubiquitin ligation and chain elongation by APC by preventing the UBE2C and UBE2S activities (PubMed:23708605, PubMed:23708001, PubMed:16921029). Plays a role in genome integrity preservation by coordinating DNA replication with mitosis through APC inhibition in interphase to stabilize CCNA2 and GMNN in order to promote mitosis and prevent rereplication and DNA damage-induced cellular senescence (PubMed:17234884, PubMed:17485488, PubMed:17875940). During oocyte maturation, plays a role in meiosis through inactivation of APC-FZR1 complex. Inhibits APC through RPS6KA2 interaction that increases FBXO5 affinity for CDC20 leading to the metaphase arrest of the second meiotic division before fertilization (By similarity). Controls entry into the first meiotic division through inactivation of APC-FZR1 complex (By similarity). Promotes migration and osteogenic differentiation of mesenchymal stem cells (PubMed:29850565).
KRI1	KRI1 Homolog	Studied mostly in <i>S. cerevisiae</i> and <i>C. elegans</i> . In <i>S. cerevisiae</i> , Kri1p is an essential nucleolar protein required for 40S ribosome biogenesis; associate with snR30; physically and functionally interacts with Krr1p. KRI1 is also involved in cell death regulation in <i>C. elegans</i> (Ito, S., Greiss, S., Gartner, A. & Derry, W. B. Cell-nonautonomous regulation of <i>C. elegans</i> germ cell death by kri-1. <i>Curr Biol</i> 20 , 333-338, doi:10.1016/j.cub.2009.12.032 (2010)).

Table S3.2 | sgRNAs used for hit verification

sgID	gene	protospacer sequence	transcript
CASP8AP2_+_90539869.23-P1P2	CASP8AP2	GAGGACCACAGAGGGACGCA	P1P2
CASP8AP2_-_90539614.23-P1P2	CASP8AP2	GATGCCAGGGAGACCTCGGT	P1P2
CASP8AP2_-_90539669.23-P1P2	CASP8AP2	GGGCGGGTAAGTTGTCGTAG	P1P2
CASP8AP2_-_90539639.23-P1P2	CASP8AP2	GAAAGGAACCGGTTGTCTT	P1P2
AURKB_-_8113837.23-P1P2	AURKB	GGCGCAAGGCCTGCGACAGG	P1P2
AURKB_+_8113846.23-P1P2	AURKB	GTAGCAGTGCCTTGGACCCC	P1P2
AURKB_+_8113717.23-P1P2	AURKB	GGTAGGGACGATAGCAGGGC	P1P2
RAD51_+_40987440.23-P1P2	RAD51	GGTCCCAGCTGCACGCCTC	P1P2
RAD51_+_40987382.23-P1P2	RAD51	GGCTGCGCGCGGTCCGCCAG	P1P2
RAD51_+_40987418.23-P1P2	RAD51	GAAGCGCCGCACTCTCCTTA	P1P2
RAD51_-_40987434.23-P1P2	RAD51	GCCGAGCCCTAAGGAGAGTG	P1P2
TACC3_+_1725435.23-P2	TACC3	GCACCTGTTGCGGAGTGAGG	P2
TACC3_+_1725634.23-P2	TACC3	GCAAGAACCATCGGACAGAG	P2
TACC3_+_1725283.23-P2	TACC3	GTAGCTTTGGCCAGGTTCTT	P2
TACC3_+_1725320.23-P2	TACC3	GGGCCAGCCACACACTGTAC	P2
INCENP_-_61891556.23-P1P2	INCENP	GCGAAGGCCCGGAGCCAAGT	P1P2
INCENP_+_61891577.23-P1P2	INCENP	GCAAGTCCGGTGGTATCCCT	P1P2
INCENP_-_61891562.23-P1P2	INCENP	GCCCGGAGCCAAGTGGGTCT	P1P2
INCENP_+_61891697.23-P1P2	INCENP	GGGTCCGGAGTCGACAGGGA	P1P2
TICRR_-_90118774.23-P1P2	TICRR	GGGAAGGGACTAAGGGACGG	P1P2
TICRR_-_90119155.23-P1P2	TICRR	GGACGCTGCTAGACTACCAG	P1P2
TICRR_+_90118911.23-P1P2	TICRR	GCCGAATCGGCAACTCAGAT	P1P2
TICRR_-_90118820.23-P1P2	TICRR	GggcggcggcACGGCCGATA	P1P2
TOP2A_+_38573771.23-P1P2	TOP2A	GTCTAGGGGCAGGAGTCCGT	P1P2
TOP2A_-_38573777.23-P1P2	TOP2A	GGGATCCACTACGGGACCAA	P1P2
TOP2A_+_38574121.23-P1P2	TOP2A	GACGCGCTGTGGAGAAG	P1P2
TOP2A_-_38573979.23-P1P2	TOP2A	GGGCGGCCAGAGAGATGCCC	P1P2
CDCA8_+_38158616.23-P1P2	CDCA8	GGTATCCACCTCCTTGAGG	P1P2
CDCA8_+_38158405.23-P1P2	CDCA8	GGCGAGCTTCGCCTCCGTA	P1P2
CDCA8_+_38158347.23-P1P2	CDCA8	GGAGCCATGGCGCTCGGCGG	P1P2
CDCA8_+_38158230.23-P1P2	CDCA8	GACCTGCGGTCCCGAGACAA	P1P2
FBXO5_+_153304114.23-P1P2	FBXO5	GGGAGCGGTTCTCCACCTG	P1P2
FBXO5_+_153303697.23-P1P2	FBXO5	GTGCGGGTCGTGGCATCCCA	P1P2
FBXO5_-_153304123.23-P1P2	FBXO5	GACGTGGAGTCTGCCTCAGG	P1P2
FBXO5_-_153304068.23-P1P2	FBXO5	GGAGCAGCGGGGTGGCCGTA	P1P2

sgID	gene	protospacer sequence	transcript
KIF11_-_94352926.23-P1P2	KIF11	GGGAGACGAGATTAGTGATT	P1P2
KIF11_+_94353079.23-P1P2	KIF11	GCGCGGACCTGGGCGCTGTG	P1P2
KIF11_-_94353272.23-P1P2	KIF11	GTGCTGGGCCCCCTACTGCG	P1P2
SPDL1_+_169010802.23-P1P2	SPDL1	GACGCTAACTCCACAGTCGC	P1P2
SPDL1_-_169011175.23-P1P2	SPDL1	GGGATCTGCGCTGTGCGGTA	P1P2
SPDL1_+_169010824.23-P1P2	SPDL1	GGCTCAGGGCGTCCACATTG	P1P2
DNA2_-_70231407.23-P1P2	DNA2	GAGGCAGGCCCGACTCCGG	P1P2
DNA2_+_70231419.23-P1P2	DNA2	GGGACCCGGTGCAGCTCCCT	P1P2
DNA2_-_70231468.23-P1P2	DNA2	GCCCGGGCCCGGTCAGTCTG	P1P2
DNA2_+_70231531.23-P1P2	DNA2	GGGAGCTGTGAGCGGAGAGG	P1P2
KRI1_-_10676181.23-P1P2	KRI1	GCACATCAGAGTGGGCAGGT	P1P2
KRI1_+_10675454.23-P1P2	KRI1	Gacggtagcatctcccctca	P1P2
KRI1_+_10676632.23-P1P2	KRI1	GTCGCAGCTGCGGGTGAACG	P1P2
KRI1_-_10676291.23-P1P2	KRI1	GAGGATGGGACGCTCCCGCA	P1P2
SKA1_+_47901696.23-P1P2	SKA1	GCTCCGCCGGTGAAGAGAGAA	P1P2
SKA1_+_47901591.23-P1P2	SKA1	GCTACCTCGGGCGCTTCCAA	P1P2
SKA1_+_47901470.23-P1P2	SKA1	GAAGCCCAGGATGGTGAGAC	P1P2
SKA1_-_47901488.23-P1P2	SKA1	GAGTCCCGTCTCACCATCCT	P1P2
NUP62_+_50432687.23-P1P2	NUP62	GAGCTGCGAAAGGGCGGGAA	P1P2
NUP62_-_50432350.23-P1P2	NUP62	GCCCTGATCCCCAGGAGGAG	P1P2
NUP62_-_50432793.23-P1P2	NUP62	GGGCGCAGAAGTAGTACGGT	P1P2
NUP62_+_50432696.23-P1P2	NUP62	GAGGGCCACGAGCTGCGAAA	P1P2

Table S3.3 | FACS laser and filters used in the study

Channel	Laser	Filter	Measured Fluorophore
Pacific Blue	Violet 407 nm	450/50	BFP
FITC	Blue 488 nm	530/30 - 502LP	GFP
PE	Green 561 nm	582/15	mCherry
APC-Cy5.5	Red 633 nm	710/50 - 685LP	mIFP

Table S3.4 | qPCR primers used in the study

Targeted gene	Forward primer	Reverse primer
CASP8AP2	CACTTGCCACTTCTACAAGTC	TGGCGGCTAAATATGCAAATG
AURKB	TGGAACGTGTACTTGGCTC	AGGATGTTGGGATGGTGC
RAD51	GTGGTAGCTCAAGTGGATGG	GGGAGAGTCGTAGATTTTGCAG
TACC3	CCTCTTCAAGCGTTTTGAGAAAC	GCCCTCCTGGGTGATCCTT
INCENP	GAGAGGCTCCTGAATGTTGAG	AATCTCCGTGTCATTGTGGG
TICRR	AGCAGGTGATGGAGAAGTTG	AAACAGTCCAGTATCCAAGGTG
TOP2A	CCTTTGCCAATGCTTCCAAGTTAC	GTGTCTTCTCGGTGCCATTCAAC
CDCA8	TTGACTACTTCGCCCTTG	CTTCTTCTCCTCTTCCACTA
FBX05	GTGTCTAAAGTGAGCACAACCTTG	TTCTCTGGTTGAAGCATGAGG
KIF11	TATTGAATGGGCGCTAGCTT	TCGTCTGCGAAGAAGAAAGA
SPDL1	GTTTGAGCTGCGAATGTGAAG	GCCTGGCTTCATCTAATTCCAC
SKA1	AGTTGAAGAACCTGAACCCG	GGTTAAGCGGGATTTCATGTAC
DNA2	GCCGCTGGAACCTAAAACCTG	GTACATCTGACCAGTCTTGAGG
NUP62	CTGAATCAAGGTCCAGAGAAGG	GTGCCTCCAAAATTAACCCG
KRI1	AAGCCTTGGAGAAGCAGAAG	CTGGAGTCTGTGATTGGAAGTC
ACTB	GCTACGAGCTGCCTGACG	GGCTGGAAGAGTGCCTCA

Materials and Methods

Plasmid sequences

CRISPRi construct (Addgene 85969) and sgRNA parental construct (Addgene 84832) were a kind gift from Jonathan Weissman's lab.

Cell line generation

hTERT-RPE1 dCas9-KRAB-BFP

All the hTERT-RPE1 cells were grown in DMEM/F-12 media supplemented with 10% FBS and Pen/Strep (complete DMEM/F-12 medium). CRISPRi modality dCas9-KRAB-BFP construct was stably expressed in hTERT-RPE1 cells via lentiviral infection, as described below. BFP positive cells were sorted after 2 days.

hTERT-RPE1 dCas9-KRAB-BFP PA-mCherry

The photo-activatable cell line was generated starting with hTERT-RPE1 dCas9-KRAB-BFP cell line. The PA-mCherry construct was stably expressed in hTERT-RPE1 dCas9-KRAB-BFP cells via lentiviral infection as described below. Monoclonal cell lines were grown and screened under the microscope to select clones with successfully integrated PA-mCherry construct. A cell line that showed high and homogeneous fluorescence after photo-activation was chosen to use in this study.

hTERT-RPE1 dCas9-KRAB-BFP PA-mCherry H2B-mGFP and hTERT-RPE1 dCas9-KRAB-BFP PA-mCherry H2B-mGFP mIFP-NLS

H2B-mGFP and mIFP-NLS constructs were sequentially integrated into hTERT-RPE1 dCas9-KRAB-BFP PA-mCherry cells via lentiviral infection. GFP positive cells or GFP/mIFP double positive cells were selected by FACS at 2 days post-infection.

sgRNA sequences

Two negative control sgRNAs were used in this study and their protospacer sequences are GCTGCATGGGGCGCGAATCA and GTGCACCCGGCTAGGACCGG. sgRNA libraries used in this study were gifts from Jonathan Weissman's lab. Since the cell line used (hTERT-RPE1 dCas9-KRAB-BFP PA-mCherry) for the mIFP proof-of-principle screen has CRISPRi modality, we used two CRISPRa sgRNA libraries for this screen. sgRNAs used for hit verification are listed in **Table S3.2**.

Lentivirus preparation and transduction

For CRISPRi modality construct and sgRNA libraries, lentiviral particles were packaged by transfecting HEK293T in a 15 cm cell culture dishes at 70% confluency with 8 µg plasmid, 1 µg PMD2.G, 8 µg dR8.91, 48 µl TransIT-LT1 transfection reagent (Mirus Bio) and 1300ul serum-free Opti-MEM medium. Medium containing lentivirus was collected 72 hr post-transfection and concentrated 10 fold using an Amicon Ultra Centrifugal Unit (MilliporeSigma). For other constructs including PA-mCherry, H2B-mGFP, mIFP-NLS and small scale sgRNA virus preparations, lentiviral particles were packaged by transfecting HEK293T in a 6-well plate at 70% confluency with 1 µg PA-mCherry plasmid, 0.1 µg PMD2.G, 0.9 µg psPAX2, 10 µl TransIT-LT1 transfection reagent (Mirus Bio) and 250 µl serum free Opti-MEM medium. Medium containing lentivirus was collected 72 hr post-transfection and concentration was not needed. 250 µl supernatant was used to transduce a 6-well plate of corresponding cells by spinning infection at 2000rpm for 1 hr. Polybrene infection reagent (Sigma) was used to increase infection efficiency. Medium was replaced with complete DMEM/F-12 medium immediately after spinning infection. Cells were puromycin selected at 5 µg/ml to select for cells successfully receiving the sgRNA (sgRNA construct harbors puromycin resistance cassette). For screening, cells were puromycin selected for 3 days.

Microscopy

Cells were grown in 96-well glass bottom dishes (Matriplate, Brooks) after puromycin selection. Images were acquired by fluorescence imaging using a Nikon Eclipse Ti-E microscope with a Nikon 20x 0.75na (Plan APO VC) objective. A digital micromirror device (DMD, DLP LightCrafter 6500 Evaluation Model, Texas Instruments) was positioned behind the back port of the microscope and illuminated using a Sutter HPX-L5UVLambda LED light source (8W output centered around 405 nm) coupled through a 5 mm diameter liquid light guide. The DMD image was projected into the sample plane using a 100 mm focal length achromatic doublet lens, and a 1x beam “expander” consisting of a pair of 80 mm focal length achromatic lenses, followed by a 450 nm long pass dichroic mirror positioned on top of the dichroic mirror used for epi-illumination (Fig. S1a). With all pixels of the DMD in the “on” position, we measured ~40 mW in the back focal plane of the objective. During image acquisition, cells were maintained in DMEM/F-12 complete medium at a constant temperature of 36°C-37°C using a stage top incubator (Tokai Hit). Fluorescence illumination was with a liquid light guide coupled LED illuminator (SpectraX from Lumencor) using a multi-band pass dichroic mirror (Semrock FF410/504/582/669-Di01-25x36) in a cube with the Semrock FF01-440/521/607/700-25 as emission filter. Camera (Andor Zyla) exposure times were usually set to 500ms for GFP channel; 100ms for mCherry channel and 1000 ms for mIFP channel.

Imaging processing

All image processing code was written in Java and used ImageJ functions to execute image processing steps. Image processing routines were plugins to the Micro-Manager Auto-PhotoConverter plugin (unless indicated), and source code can be found at <https://github.com/nicost/mnfinder> branch: mm2-gamma (<https://doi.org/10.5281/zenodo.4274219>).

1). Flatfielding and background subtraction

Background images were collected using a camera exposure of 0 ms, no illumination, and by blocking light from reaching the camera. One hundred background images were averaged. For each channel, flat field images were generated by acquiring 1000 images of cells, each slightly displaced from the other, median projection, and smoothing by Gaussian filtering. Background was subtracted, and the image was normalized so that the average of all pixels in the flat field image was 1.0.

All images acquired by the microscope were then flat field corrected by pixel wise subtraction of the background image, and division by the flat field image. This was carried out using the Micro-Manager Flat-Field Correction plugin.

2). Well edge detection

The plastic of the 96-well imaging plates is fluorescent and clearly visible in images taken at the edge of an imaging well, interfering with automated thresholding of images. To illustrate the problem, example images of automated and manual thresholding of the same image (mIFP channel) are shown below. To detect wells, the image of the user-specified channel was thresholded (threshold value was determined using the “Huang dark” ImageJ method), filled (using “Close” and “Fill Holes” commands, and expanded by a user-specified number of pixels. The detected objects were considered to be a well wall if their mean intensity was higher than a user defined value. Well edges can either be excluded from further analysis, or images containing well edges can be skipped entirely.

3). Image pre-processing

Even though images were background subtracted and flat field corrected, results were improved preprocessing using the ImageJ rolling ball background removal code (size 5, sliding) and the Smooth command.

4). Initial nuclei mask generation

Threshold values for each image were calculated using the “Otsu” method and used to segment the image. Sequential application of the “Fill Holes” and “Watershed” methods resulted in binary nuclei masks.

5). Nuclei filters to generate final nuclei mask

Nuclei particles within initial nuclei mask has to be filtered in order to exclude clustered cells or cells out of focus *etc.* Normal range of the following parameters were determined with control cells and nuclei particles out of range were discarded.

FACS

For all the screens required cell sorting, cells were trypsinized and sorted using a BD FACSAria III. For hit analysis, cells were analyzed with a BD FACSAria II after 3 days' puromycin selection. Cells were gated for single cell population and the FSC levels were analyzed using Flowjo v10.6.2.

FACS gating strategy

For all FACS experiments, forward versus side scatter (FSC-A vs. SSC-A) gating was used to identify cells of interest based on size and granularity; forward scatter width (FSC-W) vs. forward scatter area (FSC-A) density plot was used to exclude doublets. BFP (Pacific Blue-A) vs. GFP (FITC-A) density plot was used to gate H2B-mGFP cells successfully receiving sgRNAs (high BFP high GFP group). For single activation experiment (single activation mIFP proof-of-principle screen and nuclear size screen), unanalyzed samples were collected using sgRNA gate. mCherry (PE-A) vs. GFP (FITC-A) density plot was used to gate for photo-activated cells. Cells activated with different photo-activation times were used to determine corresponding gates.

Cells in dual-activation experiments were collected with mCherry gates of corresponding photo-activation time. Laser and filters used in this study were listed in **Table S3.3**.

Sequencing sample preparation

Sequencing sample was prepared using a protocol from Jonathan Weissman's lab (<https://weissmanlab.ucsf.edu/CRISPR/IlluminaSequencingSamplePrep.pdf>) except that genomic DNA of samples less than 10,000 cells was extracted with the Arcturus PicoPure DNA Extraction Kit.

mIFP proof-of-principle screen, Nuclear size screen, FSC screen and H2B-mGFP screen

For the mIFP proof-of-principle screen, mIFP positive cells (hTERT-RPE1 dCas9-KRAB-BFP PA-mCherry H2B-mGFP mIFP-NLS) and mIFP negative cells (hTERT-RPE1 dCas9-KRAB-BFP PA-mCherry H2B-mGFP) were stably transduced with the "mIFP sgRNA library" (CRISPRa library with 860 elements) and the "control sgRNA library" (CRISPRa library with 6100 elements) separately. For the nuclear size screen, FSC screen and H2B-mGFP screen, cells (hTERT-RPE1 dCas9-KRAB-BFP PA-mCherry H2B-mGFP) were stably transduced with the "nuclear size library" (CRISPRi library with 6190 elements). To guarantee that cells receive no more than one sgRNA per cell, BFP was expressed on the same sgRNA construct and cells were analyzed by FACS the day after transduction. The experiment only continued when 10-15% of the cells were BFP positive. These cells were further enriched by puromycin selection (a puromycin resistance gene was expressed from the sgRNA construct) for 3 days to prepare for imaging. For FSC and H2B-mGFP screens, cells were then subjected to FACS sorting. Cells before FACS (unsorted sample for FSC and H2B-mGFP screens) and top 10% cells based on either FSC signal (high FSC sample) or GFP fluorescence signal (high GFP sample) were separately collected and prepared for high throughput sequencing. For mIFP proof-of-principle screen and nuclear size screen, cells were then seeded into 96-well glass bottom imaging dishes (Matriplate, Brooks) and imaged

starting from the morning of the next day (around 15 hr after plating). A series of densities ranging from 0.5E4 cells/well to 2.5E4 cells/well with 0.5E4 cells/well interval were selected and seeded. The imaging dish with cells around 70% confluency was selected to be screened on the imaging day. For mIFP proof-of-principle screen, a single imaging plate was performed for each replicate while 4 imaging plates per replicate were imaged for the nuclear size screen. When executing multiple imaging runs, 2 consecutive runs could be imaged on the same day (day run and night run). 64 (8x8, day run) or 81 (9x9, night run) fields of view were selected for each imaging well and each field of view was subjected to an individual round of imaging directly followed by photo-activation. Around 200-250 cells were present in each given field of view and 60% to 80% surface area of each well was covered. Either mIFP positive cells or cells passing the nuclear size filter were identified and photo-activated automatically using the Auto-PhotoConverter μ Manager plugin. The total time to perform imaging and photo-activation of a single 96-well imaging dish with around 1.5 million cells was around 8 hr. The night run generally took longer, since more fields of view were included than in the day run. Cells were then harvested by trypsinization and pooled into a single tube for isolation by FACS. Sorting gates were pre-defined using samples with different photo-activation times (e.g. 0s, 200ms, 2s) and detailed gating strategies are described in Supplementary file 1. Sorted samples were used to prepare sequencing samples.

Bioinformatic analysis of the screen

Analysis was based on the ScreenProcessing pipeline developed in the Weissman lab (<https://github.com/mhorlbeck/ScreenProcessing>) (Horlbeck et al., 2016). The phenotypic score (ϵ) of each sgRNA was quantified as previously defined (Kampmann et al., 2013). For the mIFP proof-of-principle screen, phenotypic score of each group was the average score of two sgRNAs assigned to the group and averaged between two replicates except otherwise described. For the nuclear size screen, FSC screen and H2B-mGFP screen, genes were scored based on the

average phenotypic scores of the sgRNAs targeting them. For the nuclear size screen, phenotypic scores were further averaged between 4 runs for each replicate. For the nuclear size screen, FSC screen and H2B-mGFP screen, sgRNAs were first clustered by transcription start site (TSS) and scored by the Mann-Whitney U test against 22 non-targeting control sgRNAs included in the library. Since only 22 control sgRNAs were included, significance of hits was assessed by comparison with simulated negative controls that were generated by random assignment of all sgRNAs in the library and phenotypic scores of these simulated negative controls were scored in the same way as phenotypic scores for genes. A score η that includes the phenotypic score and its significance was calculated for each gene and simulated negative control. The optimal cut-off for score η was determined by calculating an empirical false discovery rate (eFDR) at multiple values of η as the number of simulated negative controls with score η higher than the cut-off (false positives) divided by the sum of genes and simulated negative controls with score η higher than the cut-off (all positives). The cut-off score η resulting in an eFDR of 0.1% was used to call hits for further analysis.

Verification of hits from nuclear size screen

For each hit in the nuclear size screen, the two sgRNAs with the highest phenotypic score in the screen and the two sgRNAs with the highest score predicted by the CRISPRi-v2 algorithm (Horlbeck et al., 2016) were selected and pooled to generate a mixed sgRNA pool of 3-4 sgRNAs (**Table S3.2**). Cells (hTERT-RPE1 dCas9-KRAB-BFP PA-mCherry H2B-mGFP) were transduced with pooled sgRNAs targeting each gene and puromycin selected for 2 days to prepare for imaging. Cells were then seeded into 96-well glass bottom imaging dishes. For DRAQ5 staining experiment, cells were further stained with 5 μ M DRAQ5 (Cell Signaling) for 1hr before imaging. Images were collected the next day and nuclear size and DRAQ5 staining intensity was measured using the Auto-PhotoConverter μ Manager plugin. To focus on cells with successful transduction, BFP was co-expressed on the sgRNA construct and only cells with BFP intensity above a

threshold value were included in nuclear size measurements. This BFP threshold was established by comparing the average BFP intensity of cells with and without sgRNA transduction (**Fig.S3.3a**).

RNA extraction and RT-qPCR

Total RNA was extracted using Trizol reagent (Invitrogen) according to the manufacturer's instructions. 2 µg of total RNA was treated with Turbo DNase I (Invitrogen) and 1 µg of treated RNA was used for cDNA synthesis using SuperScript™ III First-Strand Synthesis SuperMix for qRT-PCR (Invitrogen). For RT-qPCR amplification of corresponding hit genes, an initial amplification using corresponding primers (**Table S3.4**) was done with a denaturation step at 95°C for 15 min, followed by 40 cycles of denaturation at 95°C for 30 s, primer annealing at 60°C for 30 s, and primer extension at 72°C for 30 s. qPCR was carried out using SYBR Green PCR Master Mix (Applied Biosystems) with a Biorad CFX 96 Real Time System. Reactions were run in triplicate and the housekeeping gene ACTB was used as an internal control.

Data and software availability

The raw and processed data for the high throughput sequencing results have been deposited in NCBI GEO database with the accession number (GSE156623). The plugin Auto-PhotoConverter developed for open source microscope control software µManager (Edelstein et al., 2014) has been deposited on github (<https://github.com/nicost/mnfinder>).

Acknowledgements

We thank Luke A. Gilbert, Martin Kampmann, Jess Sheu-Gruttadauria, Taylor Skokan, Kara McKinley, and all the other Vale lab members for helpful discussions.

This work was supported by the Howard Hughes Medical Institute (NIH R35GM118106).

The authors declare no competing financial interests.

Author contributions: M.E. Tanenbaum conceived of the project with input from R.D. Vale and J.S. Weissman; N. Stuurman developed the Auto-PhotoConverter plugin; N. Stuurman developed image analysis code with input from X. Yan; M.A. Horlbeck designed sgRNA libraries with input from M.E. Tanenbaum; X. Yan, C.R. Liem, and M. Jost cloned the sgRNA libraries; X. Yan and S.A. Ribeiro performed the experiments; X. Yan analyzed the data; X. Yan drafted the manuscript; X. Yan, N. Stuurman, and R.D. Vale edited the manuscript with input from M.E. Tanenbaum and S.A. Ribeiro; and all authors read and approved the final article.

REFERENCES

- Aakalu, G., W.B. Smith, N. Nguyen, C. Jiang, and E.M. Schuman. 2001. Dynamic Visualization of Local Protein Synthesis in Hippocampal Neurons. *Neuron*. 30:489–502. doi:10.1016/s0896-6273(01)00295-1.
- Adamson, B., T.M. Norman, M. Jost, and J.S. Weissman. 2018. Approaches to maximize sgRNA-barcode coupling in Perturb-seq screens. *Biorxiv*. 298349. doi:10.1101/298349.
- Babendure, J.R., J.L. Babendure, J.-H. Ding, and R.Y. Tsien. 2006. Control of mammalian translation by mRNA structure near caps. *Rna*. 12:851–861. doi:10.1261/rna.2309906.
- Barrangou, R., and J.A. Doudna. 2016. Applications of CRISPR technologies in research and beyond. *Nat Biotechnol*. 34:933–941. doi:10.1038/nbt.3659.
- Blanchard, S.C. 2009. Single-molecule observations of ribosome function. *Curr Opin Struc Biol*. 19:103–109. doi:10.1016/j.sbi.2009.01.002.
- Bordeleau, M.-E., A. Mori, M. Oberer, L. Lindqvist, L.S. Chard, T. Higa, G.J. Belsham, G. Wagner, J. Tanaka, and J. Pelletier. 2006. Functional characterization of IRESes by an inhibitor of the RNA helicase eIF4A. *Nat Chem Biol*. 2:213–220. doi:10.1038/nchembio776.
- Bower, J.J., G.F. Karaca, Y. Zhou, D.A. Simpson, M. Cordeiro-Stone, and W.K. Kaufmann. 2010. Topoisomerase II α maintains genomic stability through decatenation G2 checkpoint signaling. *Oncogene*. 29:4787–4799. doi:10.1038/onc.2010.232.

- Brar, G.A., M. Yassour, N. Friedman, A. Regev, N.T. Ingolia, and J.S. Weissman. 2012. High-Resolution View of the Yeast Meiotic Program Revealed by Ribosome Profiling. *Science*. 335:552–557. doi:10.1126/science.1215110.
- Brittis, P.A., Q. Lu, and J.G. Flanagan. 2002. Axonal Protein Synthesis Provides a Mechanism for Localized Regulation at an Intermediate Target. *Cell*. 110:223–235. doi:10.1016/s0092-8674(02)00813-9.
- Cantwell, H., and P. Nurse. 2019. Unravelling nuclear size control. *Curr Genet*. 65:1281–1285. doi:10.1007/s00294-019-00999-3.
- Carmena, M., M. Wheelock, H. Funabiki, and W.C. Earnshaw. 2012. The chromosomal passenger complex (CPC): from easy rider to the godfather of mitosis. *Nat Rev Mol Cell Bio*. 13:789–803. doi:10.1038/nrm3474.
- Chao, J.A., Y. Patskovsky, S.C. Almo, and R.H. Singer. 2008. Structural basis for the coevolution of a viral RNA–protein complex. *Nat Struct Mol Biol*. 15:103–105. doi:10.1038/nsmb1327.
- Chen, J., A. Tsai, S.E. O’Leary, A. Petrov, and J.D. Puglisi. 2012. Unraveling the dynamics of ribosome translocation. *Curr Opin Struc Biol*. 22:804–814. doi:10.1016/j.sbi.2012.09.004.
- Chien, M.-P., C.A. Werley, S.L. Farhi, and A.E. Cohen. 2015. Photostick: a method for selective isolation of target cells from culture. *Chem Sci*. 6:1701–1705. doi:10.1039/c4sc03676j.
- Choi, J., K.-W. Jeong, H. Demirci, J. Chen, A. Petrov, A. Prabhakar, S.E. O’Leary, D. Dominissini, G. Rechavi, S.M. Soltis, M. Ehrenberg, and J.D. Puglisi. 2016. N6-methyladenosine in mRNA disrupts tRNA selection and translation-elongation dynamics. *Nat Struct Mol Biol*. 23:110–115. doi:10.1038/nsmb.3148.

- Cornish, P.V., D.N. Ermolenko, H.F. Noller, and T. Ha. 2008. Spontaneous Intersubunit Rotation in Single Ribosomes. *Mol Cell*. 30:578–588. doi:10.1016/j.molcel.2008.05.004.
- Datlinger, P., A.F. Rendeiro, C. Schmidl, T. Krausgruber, P. Traxler, J. Klughammer, L.C. Schuster, A. Kuchler, D. Alpar, and C. Bock. 2017. Pooled CRISPR screening with single-cell transcriptome readout. *Nat Methods*. 14:297–301. doi:10.1038/nmeth.4177.
- Dixit, A., O. Parnas, B. Li, J. Chen, C.P. Fulco, L. Jerby-Arnon, N.D. Marjanovic, D. Dionne, T. Burks, R. Raychowdhury, B. Adamson, T.M. Norman, E.S. Lander, J.S. Weissman, N. Friedman, and A. Regev. 2016. Perturb-Seq: Dissecting Molecular Circuits with Scalable Single-Cell RNA Profiling of Pooled Genetic Screens. *Cell*. 167:1853-1866.e17. doi:10.1016/j.cell.2016.11.038.
- Duxin, J.P., B. Dao, P. Martinsson, N. Rajala, L. Guittat, J.L. Campbell, J.N. Spelbrink, and S.A. Stewart. 2009. Human Dna2 Is a Nuclear and Mitochondrial DNA Maintenance Protein ∇ †. *Mol Cell Biol*. 29:4274–4282. doi:10.1128/mcb.01834-08.
- Edelstein, A., N. Amodaj, K. Hoover, R. Vale, and N. Stuurman. 2010. Computer Control of Microscopes Using μ Manager. *Curr Protoc Mol Biology*. 92:14.20.1-14.20.17. doi:10.1002/0471142727.mb1420s92.
- Edelstein, A.D., M.A. Tsuchida, N. Amodaj, H. Pinkard, R.D. Vale, and N. Stuurman. 2014. Advanced methods of microscope control using μ Manager software. *J Biological Methods*. 1:10. doi:10.14440/jbm.2014.36.
- Edens, L.J., K.H. White, P. Jevtic, X. Li, and D.L. Levy. 2013. Nuclear size regulation: from single cells to development and disease. *Trends Cell Biol*. 23:151–159. doi:10.1016/j.tcb.2012.11.004.

- Elkon, R., A.P. Ugalde, and R. Agami. 2013. Alternative cleavage and polyadenylation: extent, regulation and function. *Nat Rev Genet.* 14:496–506. doi:10.1038/nrg3482.
- Fei, J., P. Kosuri, D.D. MacDougall, and R.L. Gonzalez. 2008. Coupling of Ribosomal L1 Stalk and tRNA Dynamics during Translation Elongation. *Mol Cell.* 30:348–359. doi:10.1016/j.molcel.2008.03.012.
- Feldman, D., A. Singh, J.L. Schmid-Burgk, R.J. Carlson, A. Mezger, A.J. Garrity, F. Zhang, and P.C. Blainey. 2019. Optical Pooled Screens in Human Cells. *Cell.* 179:787-799.e17. doi:10.1016/j.cell.2019.09.016.
- Gassmann, R., A. Essex, J.-S. Hu, P.S. Maddox, F. Motegi, A. Sugimoto, S.M. O'Rourke, B. Bowerman, I. McLeod, J.R. Yates, K. Oegema, I.M. Cheeseman, and A. Desai. 2008. A new mechanism controlling kinetochore–microtubule interactions revealed by comparison of two dynein-targeting components: SPDL-1 and the Rod/Zw1ch/Zw10 complex. *Gene Dev.* 22:2385–2399. doi:10.1101/gad.1687508.
- Gilbert, L.A., M.A. Horlbeck, B. Adamson, J.E. Villalta, Y. Chen, E.H. Whitehead, C. Guimaraes, B. Panning, H.L. Ploegh, M.C. Bassik, L.S. Qi, M. Kampmann, and J.S. Weissman. 2014. Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. *Cell.* 159:647–661. doi:10.1016/j.cell.2014.09.029.
- Groot, R., J. Lüthi, H. Lindsay, R. Holtackers, and L. Pelkmans. 2018. Large-scale image-based profiling of single-cell phenotypes in arrayed CRISPR-Cas9 gene perturbation screens. *Mol Syst Biol.* 14:e8064. doi:10.15252/msb.20178064.
- Gupta, I., S. Clauder-Münster, B. Klaus, A.I. Järvelin, R.S. Aiyar, V. Benes, S. Wilkening, W. Huber, V. Pelechano, and L.M. Steinmetz. 2014. Alternative polyadenylation diversifies post-

transcriptional regulation by selective RNA–protein interactions. *Mol Syst Biol.* 10:719.
doi:10.1002/msb.135068.

Halstead, J.M., T. Lionnet, J.H. Wilbertz, F. Wippich, A. Ephrussi, R.H. Singer, and J.A. Chao. 2015. An RNA biosensor for imaging the first round of translation from single cells to living animals. *Science.* 347:1367–1671. doi:10.1126/science.aaa3380.

Han, K., A. Jaimovich, G. Dey, D. Ruggero, O. Meyuhas, N. Sonenberg, and T. Meyer. 2014. Parallel measurement of dynamic changes in translation rates in single cells. *Nat Methods.* 11:86–93. doi:10.1038/nmeth.2729.

Hashizume, C., A. Moyori, A. Kobayashi, N. Yamakoshi, A. Endo, and R.W. Wong. 2013. Nucleoporin Nup62 maintains centrosome homeostasis. *Cell Cycle.* 12:3804–3816.
doi:10.4161/cc.26671.

Hasle, N., A. Cooke, S. Srivatsan, H. Huang, J.J. Stephany, Z. Krieger, D. Jackson, W. Tang, S. Pendyala, R.J. Monnat, C. Trapnell, E.M. Hatch, and D.M. Fowler. 2020. High-throughput, microscope-based sorting to dissect cellular heterogeneity. *Mol Syst Biol.* 16:e9442.
doi:10.15252/msb.20209442.

Hir, H.L., J. Saulière, and Z. Wang. 2016. The exon junction complex as a node of post-transcriptional networks. *Nat Rev Mol Cell Bio.* 17:41–54. doi:10.1038/nrm.2015.7.

Holt, C.E., and E.M. Schuman. 2013. The Central Dogma Decentralized: New Perspectives on RNA Function and Local Translation in Neurons. *Neuron.* 80:648–657.
doi:10.1016/j.neuron.2013.10.036.

Horlbeck, M.A., L.A. Gilbert, J.E. Villalta, B. Adamson, R.A. Pak, Y. Chen, A.P. Fields, C.Y. Park, J.E. Corn, M. Kampmann, and J.S. Weissman. 2016. Compact and highly active next-

generation libraries for CRISPR-mediated gene repression and activation. *Elife*. 5:e19760. doi:10.7554/elifesciences.19760.

Hsu, P.D., E.S. Lander, and F. Zhang. 2014. Development and Applications of CRISPR-Cas9 for Genome Engineering. *Cell*. 157:1262–1278. doi:10.1016/j.cell.2014.05.010.

Hüttelmaier, S., D. Zenklusen, M. Lederer, J. Dichtenberg, M. Lorenz, X. Meng, G.J. Bassell, J. Condeelis, and R.H. Singer. 2005. Spatial regulation of β -actin translation by Src-dependent phosphorylation of ZBP1. *Nature*. 438:512–515. doi:10.1038/nature04115.

Ingolia, N.T., S. Ghaemmaghami, J.R.S. Newman, and J.S. Weissman. 2009. Genome-Wide Analysis in Vivo of Translation with Nucleotide Resolution Using Ribosome Profiling. *Science*. 324:218–223. doi:10.1126/science.1168978.

Ingolia, N.T., L.F. Lareau, and J.S. Weissman. 2011. Ribosome Profiling of Mouse Embryonic Stem Cells Reveals the Complexity and Dynamics of Mammalian Proteomes. *Cell*. 147:789–802. doi:10.1016/j.cell.2011.10.002.

Ishigaki, Y., X. Li, G. Serin, and L.E. Maquat. 2001. Evidence for a Pioneer Round of mRNA Translation mRNAs Subject to Nonsense-Mediated Decay in Mammalian Cells Are Bound by CBP80 and CBP20. *Cell*. 106:607–617. doi:10.1016/s0092-8674(01)00475-5.

Ito, S., S. Greiss, A. Gartner, and W.B. Derry. 2010. Cell-Nonautonomous Regulation of *C. elegans* Germ Cell Death by *kri-1*. *Curr Biol*. 20:333–338. doi:10.1016/j.cub.2009.12.032.

Jaitin, D.A., A. Weiner, I. Yofe, D. Lara-Astiaso, H. Keren-Shaul, E. David, T.M. Salame, A. Tanay, A. van Oudenaarden, and I. Amit. 2016. Dissecting Immune Circuits by Linking CRISPR-Pooled Screens with Single-Cell RNA-Seq. *Cell*. 167:1883-1896.e15. doi:10.1016/j.cell.2016.11.039.

Jovanovic, M., M.S. Rooney, P. Mertins, D. Przybylski, N. Chevrier, R. Satija, E.H. Rodriguez, A.P. Fields, S. Schwartz, R. Raychowdhury, M.R. Mumbach, T. Eisenhaure, M. Rabani, D. Gennert, D. Lu, T. Delorey, J.S. Weissman, S.A. Carr, N. Hacohen, and A. Regev. 2015. Dynamic profiling of the protein life cycle in response to pathogens. *Science*. 347:1259038. doi:10.1126/science.1259038.

Kampmann, M., M.C. Bassik, and J.S. Weissman. 2013. Integrated platform for genome-wide screening and construction of high-density genetic interaction maps in mammalian cells. *Proc National Acad Sci*. 110:E2317–E2326. doi:10.1073/pnas.1307002110.

Kanfer, G., S.A. Sarraf, Y. Maman, H. Baldwin, K.R. Johnson, M.E. Ward, M. Kampmann, J. Lippincott-Schwartz, and R.J. Youle. 2020. Image-based pooled whole genome CRISPR screening for Parkin and TFEB subcellular localization. *Biorxiv*. 2020.07.02.184390. doi:10.1101/2020.07.02.184390.

Katz, Z.B., B.P. English, T. Lionnet, Y.J. Yoon, N. Monnier, B. Ovryn, M. Bathe, and R.H. Singer. 2016. Mapping translation “hot-spots” in live cells by tracking single molecules of mRNA and ribosomes. *Elife*. 5:e10415. doi:10.7554/elife.10415.

Kertesz, M., Y. Wan, E. Mazor, J.L. Rinn, R.C. Nutter, H.Y. Chang, and E. Segal. 2010. Genome-wide measurement of RNA secondary structure in yeast. *Nature*. 467:103–107. doi:10.1038/nature09322.

Kweon, J., and Y. Kim. 2018. High-throughput genetic screens using CRISPR–Cas9 system. *Arch Pharm Res*. 41:875–884. doi:10.1007/s12272-018-1029-z.

- Leung, K.-M., F.P. van Horck, A.C. Lin, R. Allison, N. Standart, and C.E. Holt. 2006. Asymmetrical β -actin mRNA translation in growth cones mediates attractive turning to netrin-1. *Nat Neurosci.* 9:1247–1256. doi:10.1038/nn1775.
- Machida, Y.J., and A. Dutta. 2007. The APC/C inhibitor, Emi1, is essential for prevention of rereplication. *Gene Dev.* 21:184–194. doi:10.1101/gad.1495007.
- Mukherjee, C., D.P. Patil, B.A. Kennedy, B. Bakthavachalu, R. Bundschuh, and D.R. Schoenberg. 2012. Identification of Cytoplasmic Capping Targets Reveals a Role for Cap Homeostasis in Translation and mRNA Stability. *Cell Reports.* 2:674–684. doi:10.1016/j.celrep.2012.07.011.
- Mukherjee, R.N., P. Chen, and D.L. Levy. 2016. Recent advances in understanding nuclear size and shape. *Nucleus.* 7:167–186. doi:10.1080/19491034.2016.1162933.
- Ounkomol, C., S. Seshamani, M.M. Maleckar, F. Collman, and G.R. Johnson. 2018. Label-free prediction of three-dimensional fluorescence images from transmitted-light microscopy. *Nat Methods.* 15:917–920. doi:10.1038/s41592-018-0111-2.
- Park, H.Y., H. Lim, Y.J. Yoon, A. Follenzi, C. Nwokafor, M. Lopez-Jones, X. Meng, and R.H. Singer. 2014. Visualization of Dynamics of Single Endogenous mRNA Labeled in Live Mouse. *Science.* 343:422–424. doi:10.1126/science.1239200.
- Parnas, O., M. Jovanovic, T.M. Eisenhaure, R.H. Herbst, A. Dixit, C.J. Ye, D. Przybylski, R.J. Platt, I. Tirosh, N.E. Sanjana, O. Shalem, R. Satija, R. Raychowdhury, P. Mertins, S.A. Carr, F. Zhang, N. Hacohen, and A. Regev. 2015. A Genome-wide CRISPR Screen in Primary Immune Cells to Dissect Regulatory Networks. *Cell.* 162:675–686. doi:10.1016/j.cell.2015.06.059.

- Pawłowska, E., J. Szczepanska, and J. Blasiak. 2017. DNA2—An Important Player in DNA Damage Response or Just Another DNA Maintenance Protein? *Int J Mol Sci.* 18:1562. doi:10.3390/ijms18071562.
- Peshkin, L., M. Wühr, E. Pearl, W. Haas, R.M. Freeman, J.C. Gerhart, A.M. Klein, M. Horb, S.P. Gygi, and M.W. Kirschner. 2015. On the Relationship of Protein and mRNA Dynamics in Vertebrate Embryonic Development. *Dev Cell.* 35:383–394. doi:10.1016/j.devcel.2015.10.010.
- Piatkevich, K.D., E.E. Jung, C. Straub, C. Linghu, D. Park, H.-J. Suk, D.R. Hochbaum, D. Goodwin, E. Pnevmatikakis, N. Pak, T. Kawashima, C.-T. Yang, J.L. Rhoades, O. Shemesh, S. Asano, Y.-G. Yoon, L. Freifeld, J.L. Saulnier, C. Riegler, F. Engert, T. Hughes, M. Drobizhev, B. Szabo, M.B. Ahrens, S.W. Flavell, B.L. Sabatini, and E.S. Boyden. 2018. A robotic multidimensional directed evolution approach applied to fluorescent voltage reporters. *Nat Chem Biol.* 14:352–360. doi:10.1038/s41589-018-0004-9.
- Raab-Graham, K.F., P.C.G. Haddick, Y.N. Jan, and L.Y. Jan. 2006. Activity- and mTOR-Dependent Suppression of Kv1.1 Channel mRNA Translation in Dendrites. *Science.* 314:144–148. doi:10.1126/science.1131693.
- Rapley, J., M. Nicolàs, A. Groen, L. Regué, M.T. Bertran, C. Caelles, J. Avruch, and J. Roig. 2008. The NIMA-family kinase Nek6 phosphorylates the kinesin Eg5 at a novel site necessary for mitotic spindle formation. *J Cell Sci.* 121:3912–3921. doi:10.1242/jcs.035360.
- Rodriguez, A.J., S.M. Shenoy, R.H. Singer, and J. Condeelis. 2006. Visualization of mRNA translation in living cells. *J Cell Biology.* 175:67–76. doi:10.1083/jcb.200512137.

- Rojas-Duran, M.F., and W.V. Gilbert. 2012. Alternative transcription start site selection leads to large differences in translation activity in yeast. *Rna*. 18:2299–2305.
doi:10.1261/rna.035865.112.
- Rubin, A.J., K.R. Parker, A.T. Satpathy, Y. Qi, B. Wu, A.J. Ong, M.R. Mumbach, A.L. Ji, D.S. Kim, S.W. Cho, B.J. Zarnegar, W.J. Greenleaf, H.Y. Chang, and P.A. Khavari. 2018. Coupled Single-Cell CRISPR Screening and Epigenomic Profiling Reveals Causal Gene Regulatory Networks. *Cell*. 176:361-376.e17. doi:10.1016/j.cell.2018.11.022.
- Sansam, C.L., N.M. Cruz, P.S. Danielian, A. Amsterdam, M.L. Lau, N. Hopkins, and J.A. Lees. 2010. A vertebrate gene, *ticrr*, is an essential checkpoint and replication regulator. *Gene Dev*. 24:183–194. doi:10.1101/gad.1860310.
- Schuster, A., H. Erasmus, S. Fritah, P.V. Nazarov, E. van Dyck, S.P. Niclou, and A. Golebiewska. 2018. RNAi/CRISPR Screens: from a Pool to a Valid Hit. *Trends Biotechnol*. 37:38–55. doi:10.1016/j.tibtech.2018.08.002.
- Schwanhäusser, B., D. Busse, N. Li, G. Dittmar, J. Schuchhardt, J. Wolf, W. Chen, and M. Selbach. 2011. Global quantification of mammalian gene expression control. *Nature*. 473:337–342. doi:10.1038/nature10098.
- Shalem, O., N.E. Sanjana, E. Hartenian, X. Shi, D.A. Scott, T.S. Mikkelsen, D. Heckl, B.L. Ebert, D.E. Root, J.G. Doench, and F. Zhang. 2014. Genome-Scale CRISPR-Cas9 Knockout Screening in Human Cells. *Science*. 343:84–87. doi:10.1126/science.1247005.
- Simms, C.L., B.H. Hudson, J.W. Mosior, A.S. Rangwala, and H.S. Zaher. 2014. An Active Role for the Ribosome in Determining the Fate of Oxidized mRNA. *Cell Reports*. 9:1256–1264. doi:10.1016/j.celrep.2014.10.042.

- Sivakumar, S., J.R. Daum, A.R. Tipton, S. Rankin, and G.J. Gorbsky. 2014. The spindle and kinetochore-associated (Ska) complex enhances binding of the anaphase-promoting complex/cyclosome (APC/C) to chromosomes and promotes mitotic exit. *Mol Biol Cell*. 25:594–605. doi:10.1091/mbc.e13-07-0421.
- Sivakumar, S., P.Ł. Janczyk, Q. Qu, C.A. Brautigam, P.T. Stukenberg, H. Yu, and G.J. Gorbsky. 2016. The human SKA complex drives the metaphase-anaphase cell cycle transition by recruiting protein phosphatase 1 to kinetochores. *Elife*. 5:e12902. doi:10.7554/elife.12902.
- Sokolova, M., M. Turunen, O. Mortusewicz, T. Kivioja, P. Herr, A. Vähärautio, M. Björklund, M. Taipale, T. Helleday, and J. Taipale. 2017. Genome-wide screen of cell-cycle regulators in normal and tumor cells identifies a differential response to nucleosome depletion. *Cell Cycle*. 16:189–199. doi:10.1080/15384101.2016.1261765.
- Stumpf, C.R., M.V. Moreno, A.B. Olshen, B.S. Taylor, and D. Ruggero. 2013. The Translational Landscape of the Mammalian Cell Cycle. *Mol Cell*. 52:574–582. doi:10.1016/j.molcel.2013.09.018.
- Sullivan, M.R., and K.A. Bernstein. 2018. RAD-ical New Insights into RAD51 Regulation. *Genes-basel*. 9:629. doi:10.3390/genes9120629.
- Tanenbaum, M.E., L.A. Gilbert, L.S. Qi, J.S. Weissman, and R.D. Vale. 2014. A Protein-Tagging System for Signal Amplification in Gene Expression and Fluorescence Imaging. *Cell*. 159:635–646. doi:10.1016/j.cell.2014.09.039.
- Tanenbaum, M.E., L. Macurek, B. van der Vaart, M. Galli, A. Akhmanova, and R.H. Medema. 2011. A Complex of Kif18b and MCAK Promotes Microtubule Depolymerization and Is

Negatively Regulated by Aurora Kinases. *Curr Biol.* 21:1356–1365.

doi:10.1016/j.cub.2011.07.017.

Tanenbaum, M.E., N. Stern-Ginossar, J.S. Weissman, and R.D. Vale. 2015. Regulation of mRNA translation during mitosis. *Elife.* 4:e07957. doi:10.7554/elife.07957.

Tatavarty, V., M.F. Ifrim, M. Levin, G. Korza, E. Barbarese, J. Yu, and J.H. Carson. 2012. Single-molecule imaging of translational output from individual RNA granules in neurons. *Mol Biol Cell.* 23:918–929. doi:10.1091/mbc.e11-07-0622.

Terada, Y. 2001. Role of Chromosomal Passenger Complex in Chromosome Segregation and Cytokinesis. *Cell Struct Funct.* 26:653–657. doi:10.1247/csf.26.653.

Tholstrup, J., L.B. Oddershede, and M.A. Sørensen. 2012. mRNA pseudoknot structures can act as ribosomal roadblocks. *Nucleic Acids Res.* 40:303–313. doi:10.1093/nar/gkr686.

Vardy, L., and T.L. Orr-Weaver. 2007. Regulating translation of maternal messages: multiple repression mechanisms. *Trends Cell Biol.* 17:547–554. doi:10.1016/j.tcb.2007.09.002.

Verschuren, E.W., K.H. Ban, M.A. Masek, N.L. Lehman, and P.K. Jackson. 2007. Loss of Emi1-Dependent Anaphase-Promoting Complex/Cyclosome Inhibition Deregulates E2F Target Expression and Elicits DNA Damage-Induced Senescence ∇ †. *Mol Cell Biol.* 27:7955–7965. doi:10.1128/mcb.00908-07.

Walter, P., and G. Blobel. 1981. Translocation of Proteins Across the Endoplasmic Reticulum III . Signal Recognition Protein (SRP) Causes Signal Sequence-dependent and Site-specific Arrest of Chain Elongation that is Released by Microsomal Membranes. *The Journal of Cell Biology.* 91:557–561.

- Wang, C., B. Han, R. Zhou, and X. Zhuang. 2016. Real-Time Imaging of Translation on Single mRNA Transcripts in Live Cells. *Cell*. 165:990–1001. doi:10.1016/j.cell.2016.04.040.
- Wang, C., T. Lu, G. Emanuel, H.P. Babcock, and X. Zhuang. 2019. Imaging-based pooled CRISPR screening reveals regulators of lncRNA localization. *Proc National Acad Sci*. 116:201903808. doi:10.1073/pnas.1903808116.
- Wang, T., J.J. Wei, D.M. Sabatini, and E.S. Lander. 2014. Genetic Screens in Human Cells Using the CRISPR-Cas9 System. *Science*. 343:80–84. doi:10.1126/science.1246981.
- Wang, X., B.S. Zhao, I.A. Roundtree, Z. Lu, D. Han, H. Ma, X. Weng, K. Chen, H. Shi, and C. He. 2015. N⁶-methyladenosine Modulates Messenger RNA Translation Efficiency. *Cell*. 161:1388–1399. doi:10.1016/j.cell.2015.05.014.
- Webster, M., K.L. Witkin, and O. Cohen-Fix. 2009. Sizing up the nucleus: nuclear shape, size and nuclear-envelope assembly. *J Cell Sci*. 122:1477–1486. doi:10.1242/jcs.037333.
- Wen, J.-D., L. Lancaster, C. Hodges, A.-C. Zeri, S.H. Yoshimura, H.F. Noller, C. Bustamante, and I. Tinoco. 2008. Following translation by single ribosomes one codon at a time. *Nature*. 452:598–603. doi:10.1038/nature06716.
- Wheeler, E.C., A.Q. Vu, J.M. Einstein, M. DiSalvo, N. Ahmed, E.L.V. Nostrand, A.A. Shishkin, W. Jin, N.L. Allbritton, and G.W. Yeo. 2020. Pooled CRISPR screens with imaging on microarray arrays reveals stress granule-regulatory factors. *Nat Methods*. 17:636–642. doi:10.1038/s41592-020-0826-8.
- Wroblewska, A., M. Dhainaut, B. Ben-Zvi, S.A. Rose, E.S. Park, E.-A.D. Amir, A. Bektesevic, A. Baccarini, M. Merad, A.H. Rahman, and B.D. Brown. 2018. Protein Barcodes Enable High-

Dimensional Single-Cell CRISPR Screens. *Cell*. 175:1141-1155.e16.

doi:10.1016/j.cell.2018.09.022.

Wu, B., A.R. Buxbaum, Z.B. Katz, Y.J. Yoon, and R.H. Singer. 2015. Quantifying Protein-mRNA Interactions in Single Live Cells. *Cell*. 162:211–220. doi:10.1016/j.cell.2015.05.054.

Yanagitani, K., Y. Kimata, H. Kadokura, and K. Kohno. 2011. Translational Pausing Ensures Membrane Targeting and Cytoplasmic Splicing of *XBP1u* mRNA. *Science*. 331:586–589. doi:10.1126/science.1197142.

Yoon, S.-W., D.-K. Kim, K.P. Kim, and K.-S. Park. 2014. Rad51 Regulates Cell Cycle Progression by Preserving G2/M Transition in Mouse Embryonic Stem Cells. *Stem Cells Dev*. 23:2700–2711. doi:10.1089/scd.2014.0129.

Yoshida, M.M., and Y. Azuma. 2016. Mechanisms behind Topoisomerase II SUMOylation in chromosome segregation. *Cell Cycle*. 15:3151–3152. doi:10.1080/15384101.2016.1216928.

Yu, J., J. Xiao, X. Ren, K. Lao, and X.S. Xie. 2006. Probing Gene Expression in Live Cells, One Protein Molecule at a Time. *Science*. 311:1600–1603. doi:10.1126/science.1119623.

Yu, Q., S.-Y. Pu, H. Wu, X.-Q. Chen, J.-J. Jiang, K.-S. Gu, Y.-H. He, and Q.-P. Kong. 2019. TICRR Contributes to Tumorigenesis Through Accelerating DNA Replication in Cancers. *Frontiers Oncol*. 9:516. doi:10.3389/fonc.2019.00516.

Zaher, H.S., and R. Green. 2009. Quality control by the ribosome following peptide bond formation. *Nature*. 457:161–166. doi:10.1038/nature07582.

Publishing Agreement

It is the policy of the University to encourage open access and broad distribution of all theses, dissertations, and manuscripts. The Graduate Division will facilitate the distribution of UCSF theses, dissertations, and manuscripts to the UCSF Library for open access and distribution. UCSF will make such theses, dissertations, and manuscripts accessible to the public and will take reasonable steps to preserve these works in perpetuity.

I hereby grant the non-exclusive, perpetual right to The Regents of the University of California to reproduce, publicly display, distribute, preserve, and publish copies of my thesis, dissertation, or manuscript in any form or media, now existing or later derived, including access online for teaching, research, and public service purposes.

DocuSigned by:

E4BEC2208BAF42B... Author Signature

5/17/2021
Date