

# **Gene expression in sight**

Single-molecule studies of  
gene expression control

**Stijn Sonneveld**

**Utrecht, 2021**

**Colofon**

The work described in this thesis was performed at the Hubrecht Institute for Developmental Biology and Stem Cell Research (Royal Netherlands Academy of Arts and Sciences, KNAW).

Print: Ridderprint | [www.ridderprint.nl](http://www.ridderprint.nl)

ISBN: 978-94-6416-845-7

Cover: Mosaic composition of single-molecule fluorescent foci.

Copyright © 2021 by Stijn Sonneveld. All rights reserved. No part of this book may be reproduced, stored in a retrieval system or transmitted in any form or by any means, without prior permission of the author.

# Gene expression in sight

Single-molecule studies of  
gene expression control

## **Gen expressie in zicht**

Studies naar genregulatie op het niveau  
van enkele moleculen

(met een samenvatting in het Nederlands)

## **Proefschrift**

ter verkrijging van de graad van doctor aan de Universiteit  
Utrecht op gezag van de rector magnificus, prof.dr. H.R.B.M.  
Kummeling, ingevolge het besluit van het college voor promoties  
in het openbaar te verdedigen op woensdag 3 november 2021 des  
middags te 2.15 uur

door

**Stijn Sonneveld**

geboren op 31 mei 1992 te Zaandam

**Promotor:**

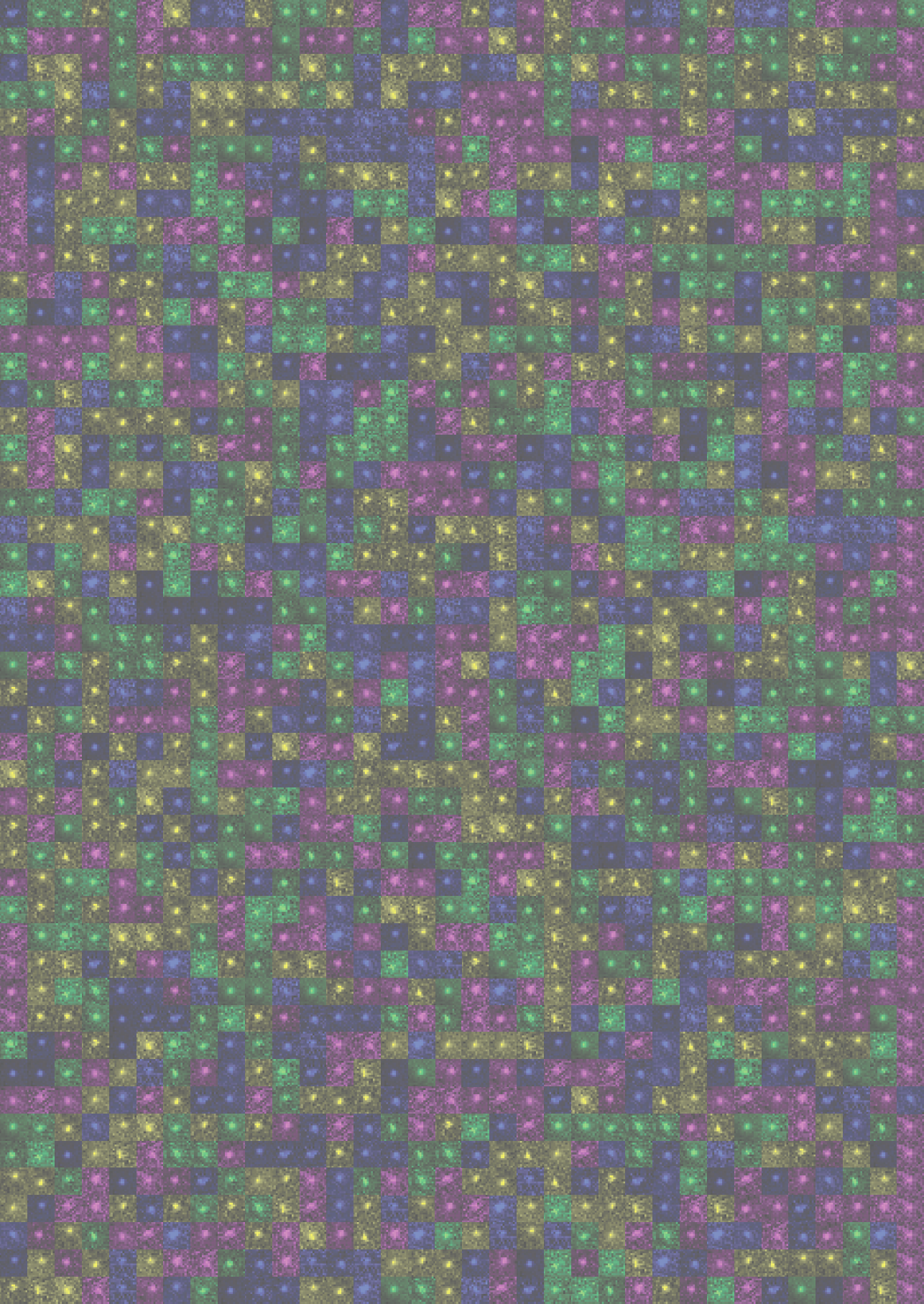
Prof. dr. W.L. de Laat

**Copromotor:**

Dr. M.E. Tanenbaum

# Table of contents

<b>Chapter 1</b>	General introduction	7
<b>Chapter 2</b>	Heterogeneity in mRNA translation	39
<b>Chapter 3</b>	mRNA structural dynamics shape Argonaute-target interactions	63
<b>Chapter 4</b>	Time-resolved single-cell sequencing identifies multiple waves of mRNA decay during mitotic exit	119
<b>Chapter 5</b>	Live-cell imaging of promoter-enhancer communication	151
<b>Chapter 6</b>	General discussion	189
<b>Appendix</b>	Tracking translation with TransTrack	199
<b>Addendum</b>	Nederlandse samenvatting About the author List of publications Acknowledgements	229



# 1

## General introduction

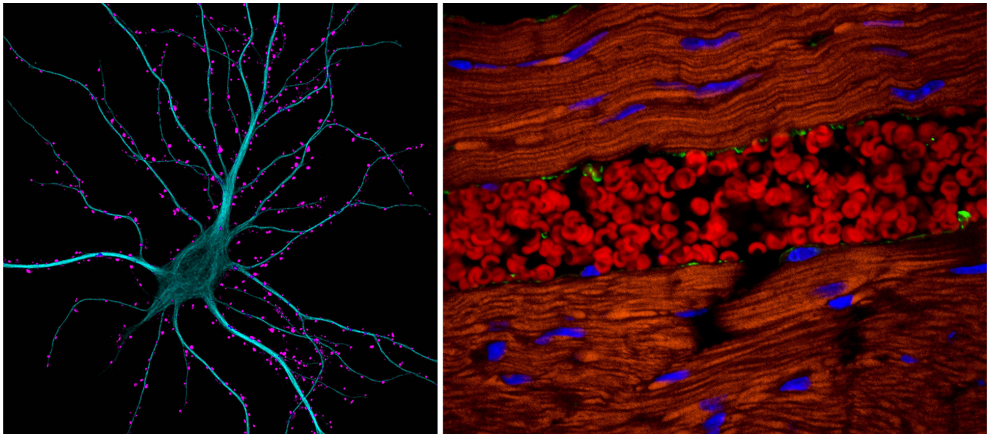
Stijn Sonneveld<sup>1</sup>

<sup>1</sup> Oncode Institute, Hubrecht Institute–KNAW and University Medical Center  
Utrecht, Utrecht, The Netherlands

## 1.1 Gene expression at the single cell level

The average human adult comprises an astounding number of 20-30 trillion cells [1]. Many of those cells perform specialized functions. In fact, it is estimated that the human body requires over 400 different cell types to function normally [2]. This functional diversity is sometimes accompanied with remarkable morphological differences between different cells (Fig. 1). For example, typical neurons have a long and thin projection (i.e. an axon) that is strongly insulated to facilitate rapid transmission of electrical signals between the brain and the body. On the other hand, red blood cells are round disks and lack a cell nucleus (and most other organelles) to increase their hemoglobin capacity. The drastic functional and morphological differences between a neuron and a red blood cell do not arise from differences in their DNA sequence. After all, a neuron and a red blood cell share the same DNA sequence, which is true for most cells in the human body. Instead, the functional and morphological differences found between cells in the body arise from processes that are involved in ‘reading out’ the DNA. Together, we refer to such processes as *gene expression* and it is the regulation of gene expression that explains the differences between a neuron and a red blood cell.

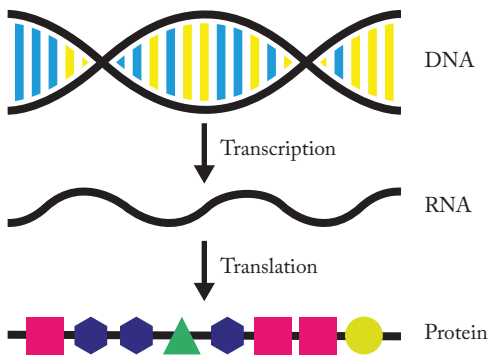
Gene expression are all those processes that are involved in converting the genetic code stored in our DNA into RNA molecules and proteins, the functional building blocks of a cell. Gene expression comprises two steps (Fig. 2) [3]. Firstly, DNA is transcribed by RNA polymerase resulting in the synthesis of RNA molecules [4]. Some



**Fig. 1 | Different cell types in the human body display varying morphologies.**

A neuron (left) and a red blood cell (right) share the same DNA but have altogether different functions and morphologies. The picture of the hippocampal neuron was an entry by Dr. Kieran Boyle into the Nikon small world competition 2013 and the picture of red blood cells inside a blood vessel was an entry by Dr. Tomasz Szul into the Nikon small world competition 2015.





**Fig. 2 | The central dogma of molecular biology.**

The central dogma of molecular biology describes the flow of genetic information. Genetic information is stored in the DNA, which can be read-out by an RNA polymerase to generate RNA molecules in a process called transcription. RNA molecules, in turn, can be converted to protein molecules with the help of the ribosomes in a process called translation.

type of RNA molecules have a function on themselves, for example ribosomal RNA (rRNA) or transfer RNA (tRNA). Other RNA molecules, namely messenger RNA (mRNA), serve merely as an intermediate; their main purpose is to synthesize proteins in a process called translation (i.e. the second step of gene expression) [5,6]. Cells contain a multitude of mechanisms to regulate gene expression, each designed to affect a different step during gene expression [3]. For instance, RNA levels can be affected by tuning RNA synthesis and degradation rates. Alternatively, protein levels can be regulated by control of the protein synthesis and degradation rates. The fact that nearly every aspect

of gene expression is regulated highlights the central role that gene expression plays in determining cell identity. Therefore, studying the mechanisms underlying gene expression regulation is of crucial importance as it will increase our understanding of how a single cell can develop in a full grown organism, and how misregulation can result in disease.

In recent years, the study of gene expression and its regulation is increasingly performed at the single cell and single-molecule level. This development is thanks to the progress in single cell sequencing techniques [7] and single-molecule microscopy-based methods that enable visualization of chromatin architecture, transcription, and translation [8–13]. But why is it interesting to study gene expression regulation at the single cell and single-molecule level in the first place? Firstly, key mechanisms of the processes underlying gene expression can be revealed by studying the distribution and dynamics of gene expression at the single cell or single-molecule level [14,15]. For instance, measuring mRNA levels at the single cell level, using a technique called single-molecule fluorescence in situ hybridization (smFISH) [16], revealed an unexpected amount of variability in mRNA levels between cells [17]. Such variability could not be explained by a simple on/off model for transcription, but required a model in which the promoter of a gene switches between two states and in which transcription can only take place in one of those two states. This model, referred to as the random telegraph model, is now often seen as the standard model for transcription [18].

Secondly, studying gene expression regulation at the single cell and single-molecule level can reveal rare or transient events that would otherwise be averaged out in bulk approaches. For example, single cell RNA sequencing experiments can discover rare cell types within a population [19] and single-molecule translation imaging can reveal rare mRNA molecules that have a high susceptibility for alternative forms of translation [20]. The examples above illustrate how single cell or single-molecule studies contribute to our understanding of the molecular mechanisms that govern gene expression processes. And with the rapid technological advances being made, it is most certain that many more of those questions will be answered in the near future.

### 1.1.1 Scope of this thesis

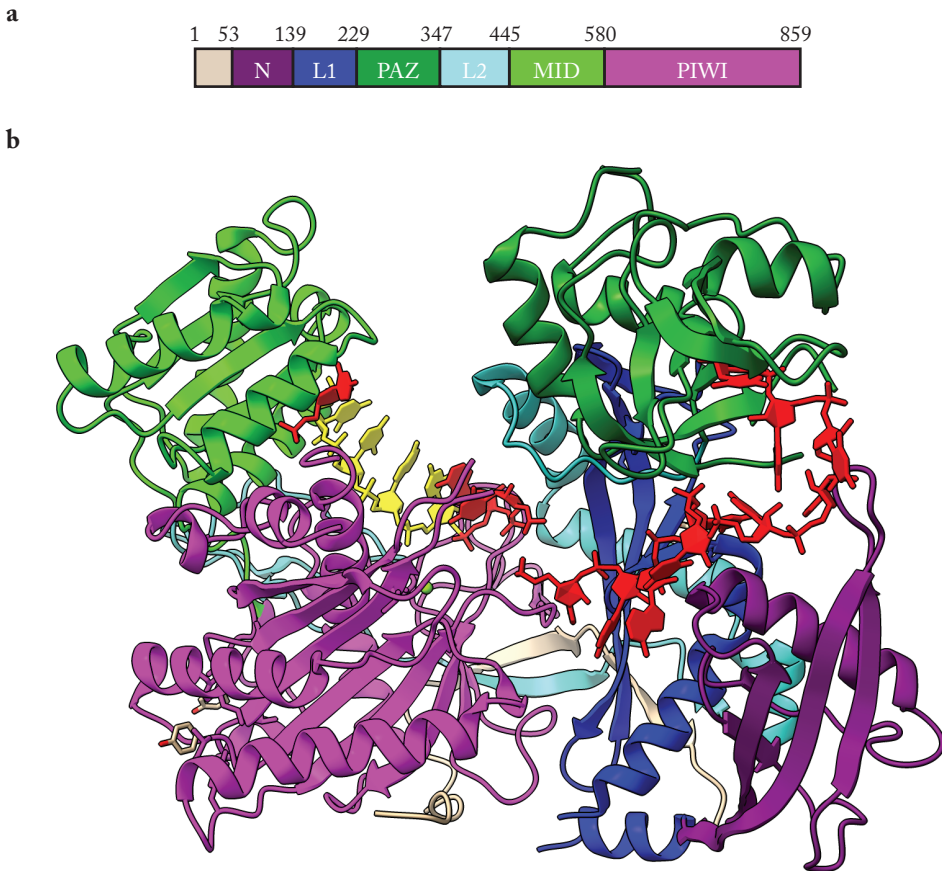
Using single cell and single-molecule techniques, we will investigate three different aspects of gene expression regulation in this thesis. As the topics of the three studies are quite distinct from each other, each study will have its own, separate introduction section below. Firstly, we examine the process of mRNA degradation by small interfering RNAs (siRNAs), revealing how mRNA structural dynamics affects the accessibility of the target mRNAs for the siRNA machinery (section 1.2). Secondly, we discuss how mRNA degradation aids to limit the expression of cell cycle genes to their appropriate cell cycle phase (section 1.3). Finally, we describe a new method to visualize the interaction between a promoter and enhancer in a single cell and we discuss how promoter-enhancer interactions can affect transcriptional output (section 1.4).

## 1.2 In vivo Argonaute-target interactions

An important class of molecules that control gene expression are small RNAs [21–25]. These short RNA molecules are generally between 20-32 nucleotides (nt) in length and are associated with a member of the Argonaute protein family. Together, a small RNA and Argonaute protein form the RNA-induced silencing complex (RISC) [26–30]. RISC is an intricate machinery that combines the programmability of the small RNA to target specific mRNA molecules and the functionality of Argonaute to silence its targets in one of many ways. RISC binding to its target can result in mRNA cleavage or in the recruitment of additional proteins that can induce heterochromatin formation, mRNA deadenylation and decay, or translational repression [27,31–35].

There are different types of small RNAs including microRNAs (miRNAs), small interfering RNAs (siRNAs), and Piwi interacting RNAs (piRNAs) [23]. Of special interest to this thesis are siRNAs. Although siRNAs are not commonly found in mammals, it is thought that siRNAs mainly function to reduce the spreading of

transposable elements in the genome [36–38] and to defend the cell against viral infections [39,40]. In addition, siRNAs are widely used as an exogenous tool to silence the expression of a gene of interest [41], and this application of siRNAs has so far culminated in the development of two ‘siRNA’ drugs that were approved in the United States and the European Union [42]. In human cells, siRNAs are loaded into Argonaute 2 (AGO2), a member of the Argonaute family that, unlike most other Argonaute proteins, cleaves its target RNAs [27,28,34]. A crucial step in the AGO2 cleavage cycle is the selection of its target molecule. While many detailed studies



**Fig. 3 | Crystal structure of AGO2 loaded with a gRNA.**

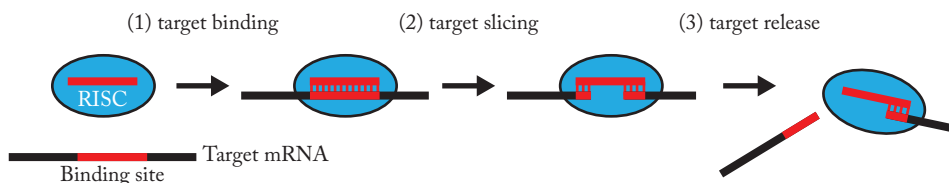
**a**, Primary sequence of human Argonaute 2. **b**, Front view of human Argonaute 2 bound to a defined guide RNA (red and yellow). The seed nucleotides from the guide (g2-g5) are shown in yellow and are exposed to the solvent. Ago2 contains four domains and two linkers. The N (purple), PAZ (dark green), MID (lime green), and PIWI (magenta) domain. And the L1 (dark blue) and L2 (light blue) linker. The Argonaute 2 structure with a defined guide RNA was retrieved from RCSB PDB (PDB ID: 4W5N) and published in [34]. Structure visualization was performed with UCSF Chimera [195].

have uncovered the basic rules of AGO2 target binding *in vitro* [43–48], it remains elusive how the *in vivo* environment affects target binding. In this thesis we will try to understand the *in vivo* environment affects siRNA cleavage efficiency.

### 1.2.1 Argonaute 2 and the RISC pathway

AGO2, a ~97 kD protein, is a member of the Argonaute protein family. Similar to other Argonaute proteins, AGO2 is composed of two lobes, which both consist of two domains (Fig. 3a) [35,49,50]. The N and PAZ domain form the N-terminal lobe, while the C-terminal lobe comprises the MID and PIWI domain. Both lobes are connected by two linker domains, together forming a central cleft. To establish the RISC, AGO2 is initially loaded with an RNA duplex after which the passenger strand is sliced, resulting in destabilization of the duplex and removal of the passenger strand [51]. The remaining guide strand is held in an extended conformation, in which the 5' nucleotide binds to the MID domain and the 3' side is buried in a narrow channel between the N and PAZ domains (Fig. 3b) [34,52–54]. On the other hand, nucleotides 2-7, which are known as the seed region, form a helical structure of which nucleotides 2-4 are directly exposed to the solvent [34].

The RISC pathway can be divided in three phases comprised of (1) target binding, (2) target slicing, and (3) target release (Fig. 4). A plethora of biochemical, structural, and single-molecule studies have provided fundamental insights into these different steps [21,34,35,43–46,55]. Target binding is a complex process in which full target binding is achieved in multiple steps [50]. To optimize the target search, AGO2 is thought to divide its guide into several domains. Nucleotides 2-4, referred to as the sub-seed region, have been shown to be critical for initial target site interactions, which is in accordance with structural studies that have shown that only these nucleotides are initially available for target binding [34,35]. It has also been proposed that the pre-ordered state of the sub-seed lowers the entropic cost of hybridization, possibly explaining why AGO2 binds faster to RNA molecules compared to a naked siRNA



**Fig. 4 | The AGO2 cleavage pathway.**

AGO2 is loaded with an siRNA (together forming RISC) complementary to a sequence in its target mRNA. AGO2 binds to the target mRNA, slices the target between nucleotides 10 and 11, and releases the two resulting cleavage fragments of the mRNA.

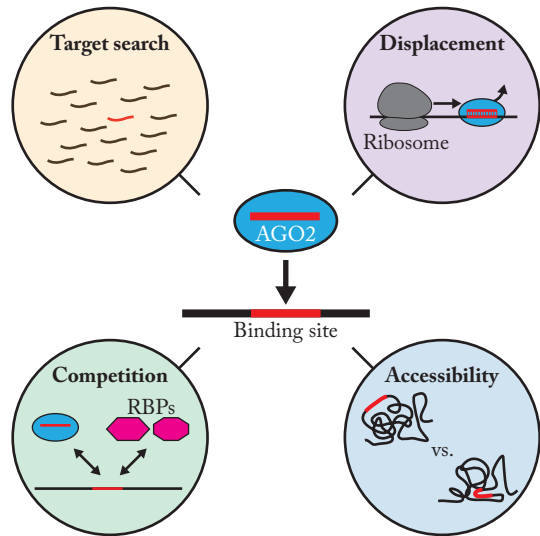
[44,45]. Once sub-seed pairing has occurred a conformational change in AGO2 enables full seed pairing with nucleotides 2-7 [34]. Next, another conformational change enables supplemental pairing to the 3' end of the siRNA (specifically nucleotides 13-16) [34]. Finally, it has been shown that the presence of an adenosine in the target site opposite to the 5' nucleotide of the siRNA can further stabilize siRNA-target interaction by directly binding with AGO2 (as the 5' nucleotide of the siRNA is not available for target interaction) [56]. Taken together, initial target interactions are formed between the target and nucleotides 2-4 of the siRNA after which additional pairing results in a stable siRNA-target interaction.

To proceed to the next step of slicing, the siRNA requires full complementarity binding with its target. Interestingly, while full complementarity is required for the slicing activity of AGO2, it was shown that additional pairing negatively affects the affinity of AGO2 for its target, possibly explaining why many small RNAs that do not cleave their targets show only partial complementarity to their target [34,43]. Full pairing between the siRNA and target RNA is presumably accompanied by conformational changes in AGO2, resulting in a widening of the central cleft and activation of the slicing domain [50]. Next, AGO2 slices the target between nucleotide 10 and 11, resulting in a cleaved RNA target [29]. Finally, the sliced target needs to be released. It was shown that both the 5' and 3' cleavage fragment can be released first, and that this most likely depends on the binding stability of the individual fragments [55].

Target binding is arguably the most crucial step in the RISC pathway and there are many factors that affect efficient binding (Fig 5). Firstly, AGO2 requires to find its target among all mRNA molecules that are present in a cell. As generally the number of non-targets greatly exceeds the number of targets, this search is analogous to the search for a needle in a haystack. A second factor affecting target binding are proteins that displace AGO2 from its target site. Most notably, translation of mRNAs is performed by ribosomes, giant complexes that could potentially displace AGO2 [57]. Interestingly, it was found that miRNA sites are depleted from the open reading frame (ORF), possibly suggesting that AGO2 is displaced by ribosomes and unable to perform its function properly when its target site is located in the ORF [58–60]. Thirdly, RNA molecules in vivo are covered by RNA binding proteins (RBPs) [61]. These proteins may bind to similar or overlapping target sites with AGO2 and thereby hinder AGO2-target interactions [62–64]. Finally, RNA molecules can adopt secondary and tertiary structures through intra-and-intermolecular interactions. These type of structures can potently 'mask' the target site from AGO2, inhibiting AGO2-target interactions [65–68]. In accordance, miRNA sites preferentially reside in GC-poor regions [58], suggesting that regions with weaker structures (as G-C

**Fig. 5 | AGO2-target binding is affected by several cellular factors.**

The binding of AGO2 to its target in cells is affected by several factors. Firstly, cells contain many RNA molecules among which AGO2 needs to find its target. Secondly, once bound to its target, AGO2 can be displaced by other proteins (for example by ribosomes that translate the mRNA) before performing its mechanism of action. Thirdly, one or multiple RNA binding proteins (RBPs) may bind to the same sequence as AGO2. Finally, the AGO2 binding site might be masked from AGO2 binding by RNA structure, hindering AGO2-target binding.



interactions are stronger than A-T interactions) are more favorable for AGO2-target interactions.

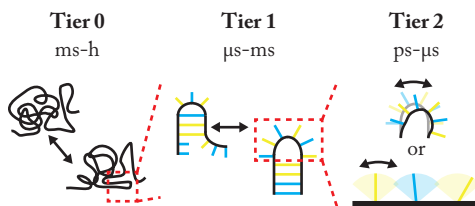
To overcome some of these ‘challenges’ in target binding, AGO2 has equipped itself with some intricate strategies. For instance, to efficiently identify a target site, AGO2 performs a one-dimensional search along an mRNA molecule [45], thereby minimizing its target search time. During this scanning process, AGO2 is not hindered by any cellular obstacles (such as RBPs) as it has the ability to ‘hop’ over them [69]. On the other hand, much less is known how AGO2 can identify target sites that are masked by RNA structure. To understand how AGO2 can reach target sites that are (frequently) inaccessible, we need to know what type of structures limit target site accessibility. This is what we will turn to next.

### 1.2.2 RNA structure

RNA is a linear polymer that is built from four different nucleotides: A, C, U, and G [3]. Similar to protein structure, RNA structure can be hierarchically described. The RNA primary structure is the nucleotide sequence of the RNA molecule. Next, the secondary structure is described by the canonical Watson-Crick (WC) base pairs that are formed within the molecule (WC base pairs are A=U and C=G). Since WC base pairs are the strongest interactions in an RNA molecule, it’s thought that WC base pairing determines the general structure of an RNA molecule and other type of interactions are only refining the final RNA structure [70]. The tertiary structure is set through non-canonical interactions such as non-canonical base interactions or base-sugar interactions. Compared to proteins, which are comprised of (maximally)

20 amino acids that can undergo a multitude of interactions with each other, RNA structure is relatively simple and mostly determined by only two distinct interactions (i.e. the WC base pairs). As a result, many RNA secondary and tertiary structures are redundant in terms of their free energy, meaning that an RNA molecule often lacks a single secondary or tertiary structure that has a much lower free energy compared to the other structures [71]. Therefore, it is expected that two RNA molecules with the same primary sequence may have completely different secondary and tertiary structures [72–77]. This may have far-reaching consequences for AGO2-target interactions as it can very well be that a particular target site is accessible in some secondary and tertiary structures but not in others.

RNA structure is also dynamic, because thermal fluctuations can disrupt existing interactions and result in the formation of new ones. As a result, the accessibility of a particular AGO2 target site may fluctuate over time as an RNA molecule adopts different secondary and tertiary structures. It is known that RNA structure dynamics can take place on timescales spanning 12 orders of magnitude (Fig. 6) [72]. On the very fast time scale (ps-ns), the RNA backbone can change direction and individual bases can undergo small movements. At somewhat slower timescales ( $\mu$ s-ms), individual nucleotide interactions can be broken and formed. Finally, several studies have shown that large scale structural changes can take place at slow timescales ranging from ms to hours [78–80]. Such remarkably slow transitions result from the strong thermodynamic stability of secondary RNA interactions [71]. The fast ps-ns structural changes will most likely have no to little effect on AGO2-target binding as these changes are only taking place locally and don't change any base-pairing interactions. On the other hand, the  $\mu$ s-hour structural changes will very likely affect target site accessibility. Especially slow structural changes (i.e. min-hour) may affect AGO2 functioning as it could happen that target sites remain inaccessible for



**Fig. 6 | RNA structure dynamics at different time scales.**

RNA molecules adapt different structural conformations, but the transition from one conformation to another can occur at different time scales depending on the type of transition. Large scale 3D rearrangements (referred to as tier 0 rearrangements) require the breakage and

formation of many secondary and tertiary interactions, resulting in transition times on the ms to hour time scale. The formation and disruption of individual base pairs, for example resulting in transitions between hairpins with different base-pairing, occurs at the  $\mu$ s to ms time scale (tier 1 rearrangements). Finally, at the very short time scale of ps to ns, the RNA backbone can undergo inter-helical movements and/or the individual bases and sugars can make small movements (tier 2 rearrangements). Figure adapted from [72].

extended durations of time, effectively protecting an RNA molecule against AGO2-mediated cleavage.

In cells, RNA structure is not only subject to change by thermal fluctuations but is also actively modified by cellular factors [81]. As RNA secondary structure can be extremely stable, these cellular factors are necessary to accelerate structural transitions that would otherwise take too long [71]. One important class of proteins that affect RNA structure are RNA helicases [82]. Most helicases bind to a double stranded region in the RNA molecule and unwind the RNA with low processivity, resulting in local unfolding of the double stranded region [83]. Other helicases comprise higher processivity, which may result in more global structural changes [83]. Not only helicases affect RNA structure *in vivo*, also RNA binding proteins (RBPs) without helicase activity may affect RNA secondary and tertiary structure by binding to a specific region, making this region unavailable for base-pairing [71]. Finally, the structure of coding RNAs can also be affected by ribosomes. Ribosomes unwind RNA structure independent of other helicases [57] and several studies have demonstrated that translation generally results in a reduction of RNA structure [84–87]. Taken together, cellular factors such as helicases, RBPs, or ribosomes may strongly impact RNA structure *in vivo*, thereby affecting AGO2 target site accessibility.

### 1.2.3 Outstanding questions

There are many outstanding questions regarding siRNA-Argonaute functioning. Of particular interest are those that address the question why some siRNAs display high efficacy while others do not, since a better understanding of siRNA efficacy may improve siRNA design for therapeutic treatment. In this thesis, we ask these type of questions using single-molecule microscopy in living cells. Some of the questions that we will try to answer are:

- How fast does AGO2-mediated mRNA cleavage occur in cells?
- What factors affect AGO2 functioning?
- What type of RNA structures affect AGO2-target interactions and is target accessibility changing over time?

The answers to these questions will hopefully give us a better understanding of AGO2 functioning inside living cells.



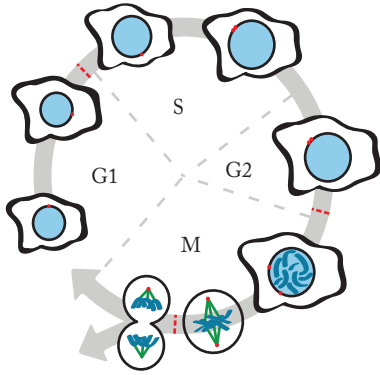
### 1.3 The role of RNA degradation during the cell cycle

Cell division is an essential process in life [88]; for unicellular organisms cell division results in the formation of a new organism and for multicellular organisms numerous cell divisions produce all the cells that encompass a full grown adult. Most cell divisions are accompanied by a duplication of the cellular contents in order to end up with two daughter cells that are (almost) identical to its parent cell. Of special importance is the DNA, which must be duplicated exactly once and of which each daughter cell should receive one copy. Failure to achieve this will result in missing or extra copies of chromosomes, eventually culminating in cell death or carcinogenesis [89,90]. Since an average human body experiences about 10 quadrillion (i.e. 10.000 trillion) cell divisions during its life time [91], cell replication requires to occur with extreme precision. To ensure precise cell replication, a cell undergoes this process in a step-by-step fashion and only proceeds to the next step after the previous step has been fully completed. This order of events is referred to as the cell cycle [88].

To control cell cycle progression (i.e. to make decisions to continue or pause), the cell has geared itself with a plethora of regulatory mechanisms [88]. Central to these mechanisms are a set of genes that form intricate biochemical switches to control cell cycle progression [88]. In addition, many genes, referred to as cell cycle genes, display periodic expression during the cell cycle [92]. Cell cycle genes are only required in a particular phase of the cell cycle, and periodic expression ensures that these genes are only active in their designated period of the cell cycle. Different steps in the gene expression pathway ranging from transcription to protein degradation are used to regulate expression levels of the cell cycle genes. In this thesis we will investigate the role of mRNA degradation in the regulation of these genes. In particular, we will examine the role that mRNA degradation plays in reshaping the transcriptome as cells re-enter a new cell cycle. Additionally, we will try to understand how specific transcripts can be targeted by the mRNA degradation machinery while others remain unaffected.

#### 1.3.1 The cell cycle

During the eukaryotic cell cycle all cellular contents are duplicated and equally distributed between two daughter cells. Most cellular components (such as organelles) are being replicated many times throughout the cell cycle, and they are divided between both daughter cells by simply 'cutting' the cell in half [88]. On the other hand, the chromosomes must be duplicated only once after which they need to be precisely distributed between both daughter cells. To ensure a reliable execution of this process, replication and distribution of the chromosomes is divided over distinct cell cycle phases. First, the DNA is replicated during S phase after which the duplicated



**Fig. 7 | Overview of the eukaryotic cell cycle.**

The eukaryotic cell cycle can be divided in four cell cycle phases. In G1 phase cells grow and prepare for DNA replication. In S phase the DNA is replicated. In G2 phase the cell grows further and prepares for cell division. In M phase (or mitosis) the cell divides into two (genetically) identical daughter cells. After M phase a cell can either re-enter the cell cycle or turn into a quiescent state (sometimes referred to as G0).

chromosomes are equally divided over the two daughter cells in M phase. In most eukaryotic cell cycles, both S and M phase are preceded by a so-called gap phase (G1 and G2 phase, respectively), together comprising the eukaryotic cell cycle (Fig. 7). During G1 and G2 phase, the cell continues duplicating its cellular contents (except for the chromosomes) and prepares for the next phase. Of these two gap phases, especially G1 phase is important as cells can decide here if they re-enter a new cell cycle round or enter a non-dividing state (i.e. quiescence or G0) [88].

To ensure faithful progression through the cell cycle, cells are required to express and activate a set of genes at precisely defined moments [92–96]. These genes are referred to as cell cycle genes. Cell cycle genes can be divided into two categories: the genes that regulate the transitions during the cell cycle (e.g. from G1 to S phase or from G2 to M phase) and ‘secondary’ genes that carry out the tasks that are required during the respective cell cycle phases (e.g. genes involved in DNA replication or in the mitotic spindle). The main regulatory cell cycle genes are the cyclin-dependent kinases (CDKs) and the corresponding cyclins that are required to activate them [88]. Together, CDK and cyclin activity determines when cells progress to the next cell cycle phase, and is responsible for the activation of the required genes in a particular cell cycle phase. CDK activity is mainly regulated by cell cycle dependent phosphorylation events and by fluctuating cyclin levels [88]. Cyclin levels are increased by transcriptional activation [97], for instance by E2F transcription factors during the G1 to S transition, and decreased by proteasomal degradation mediated by ubiquitin ligase complexes such as the SCF and APC/C [98,99]. Together, these regulatory processes form a complex network with intricate feedback mechanisms that ensure proper cell cycle progression.

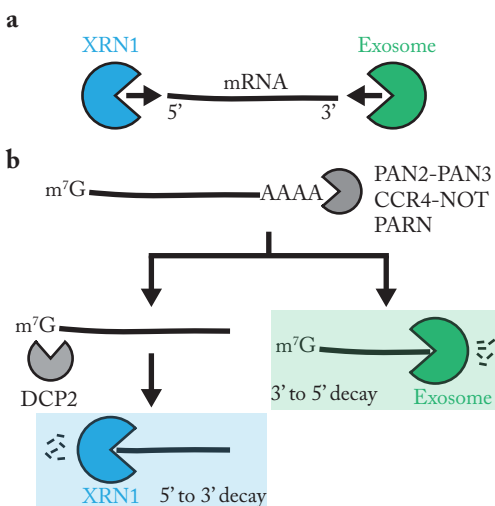
Regulation of the expression levels of cell cycle genes does not only occur through transcriptional activation and protein degradation, but also through translational control [100–102] and mRNA degradation [103–105]. Of special interest to this thesis is the role of mRNA degradation in the cell cycle. If cell cycle gene expression is already controlled by a combination of transcriptional activation and protein

degradation, the question arises what additional role mRNA degradation could play. One answer to this question could be that regulation through mRNA degradation rate offers different response kinetics which cannot be achieved with regulation through mRNA synthesis rate [104,106–108]. Specifically, unstable genes are predicted to respond much faster to changes in mRNA synthesis rate than stable genes as the speed of the transition is only dependent on the degradation rate (see [109] for an explanation of the kinetic model). Somewhat counterintuitively, this also means that unstable genes react faster to transcriptional activation (i.e. they will reach the new steady state levels faster). Such characteristics may be especially important for regulatory cell cycle genes that need to respond fast upon entry of a new cell cycle phase.

The existence of mRNA degradation during the cell cycle also raises the question how this degradation is regulated and how specific mRNAs can be destabilized at particular times in the cell cycle. To answer these questions we need to take a closer look at the mRNA degradation pathways that are present in the cell.

### 1.3.2 mRNA degradation

mRNA degradation has long been regarded merely as a means to get rid of expired or aberrant mRNA molecules, but it is now evident that regulation of mRNA degradation also plays a key role during many processes such as development or the cell cycle [103–107]. Degradation of mRNA molecules is performed by exonucleases, either from a 5' to 3' or from a 3' to 5' direction (Fig. 8) [110]. In human cells the most common cytoplasmic 5' to 3' exonuclease is XRN1 [111], while 3' to 5' decay is carried out by a large complex known as the exosome [112]. Normal mRNA molecules



**Fig. 8 | mRNA degradation pathways in human cells.**

**a**, In the cytoplasm unprotected mRNA molecules (i.e. without a 5' cap and/or a 3' poly(A) tail) are degraded either in a 5' to 3' direction by XRN1 or in a 3' to 5' direction by the exosome. **b**, The most common mRNA degradation pathway in the cytoplasm is deadenylation-dependent decay. First the poly(A) tail is removed by a deadenylase (for example: PAN2-PAN3, CCR4-NOT, or PARN). Next, the mRNA molecule is either directly degraded from the 3' end by the exosome or the cap is removed by DCP2 followed by XRN1-mediated degradation from the 5' end.

are protected against exonucleolytic degradation by features present at both ends of the mRNA molecule. At the 5' end the 7-methylguanosine cap and at the 3' end a poly(A) tail protect against XRN1 and the exosome, respectively [110]. Both the cap and the poly(A) tail also recruit specific proteins; the cap interacts with eIF4E, which is essential for translation initiation, and the poly(A) tail recruits multiple poly(A)-binding proteins (PABPs), which are thought to be crucial in the protection against exosome-mediated decay [110]. To degrade an mRNA molecule, either the cap or the poly(A) tail have to be removed. Alternatively, the cell also contains a number of endonucleases that are able to internally cleave an mRNA molecule, ultimately resulting in rapid degradation of both cleavage fragments (as the 5' cleavage fragment is unprotected at its 3' end and the 3' cleavage fragment is unprotected at its 5' end). Here, we will describe the most common cytoplasmic mRNA decay pathway which is initiated by deadenylation of the 3' poly(A) tail.

Deadenylation-dependent decay starts with the degradation of the poly(A) tail by a deadenylase. Several deadenylases have been found in human cells, most notably the PAN2-PAN3 complex, the CCR4-NOT complex and PARN [113–115]. It is still unclear if each deadenylase has a distinctive function or if they are partially redundant, most likely it is a combination of both. After deadenylation, an mRNA is decapped by Dcp2 resulting in 5' to 3' degradation by XRN1 [116]. Alternatively, a deadenylated mRNA molecule can also be directly degraded by the exosome. Both 5' to 3' and 3' to 5' degradation after deadenylation are likely to happen as knock out of any of the two pathways only results in minor transcriptome changes, indicative of redundancy [117,118]. In contrast, other studies suggest that under physiological circumstances XRN1 is responsible for the bulk of cytoplasmic mRNA decay and that the exosome is mostly active in degradation of mRNAs that are marked for decay by translation-associated mRNA quality control pathways [119,120]. In summary, deadenylation-dependent decay is the main cytoplasmic mRNA decay pathway and ultimately results in degradation of the mRNA in a 5' to 3' or a 3' to 5' fashion.

How are mRNA molecules selected for deadenylation-dependent decay? This is still an area of intensive research but it seems that deadenylases are generally recruited to mRNA molecules through other RBPs. These RBPs are recruited to elements that are usually present in the 3' UTR, presumably because of the close proximity to the poly(A) tail and the lack of translating ribosomes which may otherwise displace RBPs. One of the most studied mRNA-stability elements is the so-called AU-rich element (ARE) [121]. All ARE-sequences contain an AUUUA sequence, however no ARE sequence is exactly identical as they vary in the number of AUUUA repeats or in surrounding sequence [122]. Potentially, this allows the RNA decay machinery to distinguish between different mRNAs [110]. In line with this, it was shown for

another mRNA stability element (a Pumilio-response element) that its RBP (Puf3) could distinguish between RNAs containing very similar sequences [123], suggesting that Puf3 is able to induce degradation of specific target mRNAs. Taken together, it is thought that the deadenylase machinery is directed to its target mRNAs through a wide variety of mRNA-stability elements, but more research is required to identify all elements and their interplay.

### 1.3.3 Outstanding questions

In this thesis we will examine the role of mRNA degradation during the cell cycle. Specifically, we want to investigate if mRNA degradation of (some) cell cycle genes is regulated and what the mechanisms are behind cell cycle dependent mRNA degradation. We will focus on a specific part of the cell cycle, namely the M-G1 phase transition. As cells enter a new cell cycle, the activity of many G2/M-specific genes need to be reversed, and this is therefore a likely time when mRNA degradation might play a role. While examining this cell cycle phase transition, we will ask the following questions:

- Which cell cycle genes are subject to mRNA degradation during the M-G1 phase transition?
- What is the timing of degradation, are all genes degraded simultaneously or are there specific timings for each gene?
- How are cell cycle genes selected for mRNA degradation?

Together, this may answer some of the remaining questions regarding the role of mRNA degradation during the cell cycle.

## 1.4 How enhancers control transcription

Gene expression starts with transcription during which the information stored in the DNA is decoded to synthesize mRNA molecules. Years of intensive research have greatly increased our understanding of the key players and processes in transcription (for a comprehensive review on the eukaryotic transcription cycle see [4]). Briefly, transcription of mRNA molecules is initiated by recruitment of RNA polymerase II (Pol II) to the transcription start site. On Pol II binding and the formation of the preinitiation complex, the double-stranded DNA helix locally melts to provide a single-stranded DNA template to Pol II. Next, Pol II undergoes a series of abortive elongation cycles in the first 50-100 nt (i.e. synthesizing and releasing short RNA oligonucleotides) until phosphorylation of Pol II and transcription factor TFIIB initiates productive elongation. After the start of productive elongation, Pol II

transcribes the remainder of the gene body until it encounters a polyadenylation sequence (usually present at the 3' end of the gene body) resulting in cleavage of the nascent mRNA and dissociation of Pol II from the DNA.

Transcription is also a dynamic process and to understand transcription and its regulation it is paramount to measure its kinetics [18]. While initial (biochemical) studies provided crucial insights into the processes underlying transcription, these studies were not well suited to study the dynamics of transcription as the measurements they are based on mostly lack dynamic information and require averaging over many molecules. Around the turn of the millennium, progress in fluorescent microscopy sparked the development of new microscopy-based techniques that enabled (live) single-molecule imaging of transcription [16,124,125], allowing single-molecule measurements of transcription. Surprisingly, it was found that transcription of most genes occurs in a burst-like fashion [17,126]: rather than a constant production of mRNA molecules, most genes display periods of active transcription alternated with periods of inactivity. As a consequence the transcription levels can widely vary from cell to cell or in a single cell over time. Nowadays, it is generally accepted that the variability of transcription is a reflection of its stochastic nature, since transcription requires the stochastic recruitment of multiple transcription factors (TFs) and Pol II [127].

This new and dynamic view on transcription also increased the understanding of transcriptional regulation. Measuring transcription for many genomic positions revealed that cells control the transcription rate (i.e. the amount of mRNA molecules produced in a particular time span) by regulating either the burst size (how many Pol II molecules initiate in a period of active transcription) or burst frequency (how often does a period of active transcription occur) [128,129]. Interestingly, several examples have shown that burst size and frequency can be regulated either cooperatively or independently [130,131]. For example, the  $\beta$ -globin gene is transcriptionally upregulated during erythroid maturation by an increase in burst frequency and size [130]. On the other hand, changes in cell size increase the burst size but not frequency, while DNA duplication (e.g. during certain phases in the cell cycle) decreases burst frequency but not size [131]. While these different types of transcriptional regulation are becoming increasingly clear, much less is known about the molecular mechanisms that regulate burst size and frequency. Different mechanisms such as DNA supercoiling, TF (un)binding, and the chromatin state have all been proposed to play a role [18], and it is most likely that a complex interplay of all those factors regulate transcription.

An important player in transcriptional regulation are DNA regulatory elements. These

elements often contain multiple binding sites for TFs and chromatin remodelers, thereby regulating transcriptional activity. Two important types of such regulatory elements are promoters and enhancers. Promoters are located directly upstream of a gene and are pivotal for transcription initiation by recruiting TFs and Pol II [3]. However, promoters don't function autonomously and their activation often requires interactions with enhancer elements. Here we will discuss what these enhancer elements are and how they function to control transcription.

### 1.4.1 Enhancers

Intriguingly, less than 2% of the DNA actually contains protein coding genes. The other ~98% of the DNA, generally referred to as the non-coding DNA, is thought to regulate the expression levels of many of the protein coding genes by controlling the transcriptional activity. One class of non-coding DNA elements that are thought to be particularly important in transcriptional regulation are enhancers [132,133]. Several lines of evidence suggest that enhancer elements are one of the main primary regulatory elements of transcription in metazoans. Firstly, enhancers are often active in specific cell types and rewiring of enhancer interactions is an essential process during development [133,134]. Furthermore, disruption of promoter-enhancer contacts can result in disease. For example, structural variant (SV) mutations that alter promoter-enhancer contacts can result in developmental defects such as polydactyly (the addition of fingers or toes) or in diseases such as cancer [135,136]. Enhancers are also thought to play an important role during evolution as they have less constraints on the nucleotide composition compared to coding genes and rewiring of promoter-enhancer contacts can result in new functions of genes [134,137].

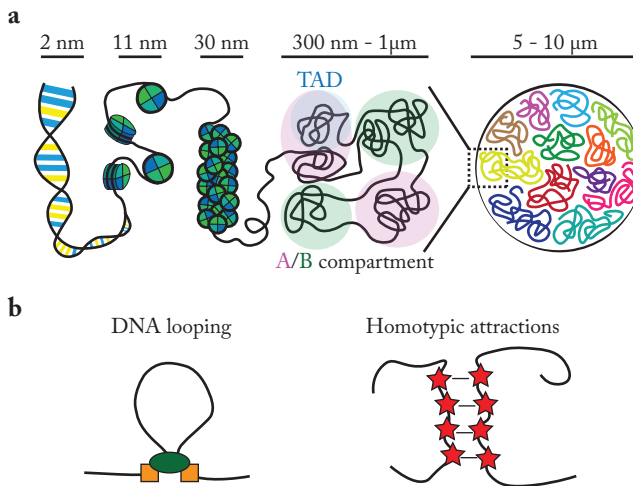
When first identified, enhancers were defined as regulatory elements that increase transcriptional activity irrespective of the distance or orientation towards its target gene [138]. In fact, enhancers are often located far away from their target gene, and can be found up to several mB upstream or downstream. Enhancers were initially identified using reporter-based assays and genome-wide searches based on sequence conservation [138,139]. Later these approaches were complemented with techniques that searched for specific enhancer marks such as DNase I hypersensitivity or the presence of specific histone modifications (e.g. H3K4me1) [140–143]. Together, this has revealed that the human genome contains hundreds of thousands putative enhancers, thereby largely outnumbering the number of genes (the human genome is estimated to contain ~20.000 genes) [144,145]. It must be noted, however, that the functionality of enhancers requires to be tested individually and likely not all of these enhancers will be functionally active [146,147]. Nevertheless, it's assumed that most genes are regulated by multiple enhancers and this apparent redundancy may

fulfill several purposes. First, transcription of a gene is more robust if it's regulated by multiple enhancers [148]. As enhancer-gene contacts are made over large distances, it is not evident that they will always be able to do so. The existence of multiple enhancers would in this case increase the chance of a functional promoter-enhancer interaction, ultimately resulting in more robust transcription. Secondly, multiple enhancers allows complex transcriptional regulation [149]. For example, Sonic hedgehog (Shh) transcription is regulated by more than 11 enhancers, resulting in complex spatiotemporal control of Shh transcription [150].

The example of the Shh gene highlights a key challenge for most enhancers: how to find their target genes and how to communicate with the correct gene at the correct time. In the next section we will focus on the current model, which proposes that enhancer-target interactions are mediated by the 3D genome organization.

### 1.4.2 The 3D organization of the genome shapes enhancer contacts

Many years of extensive research has established that DNA has a strong hierarchical organization (Fig. 9). This strict organization is not only important for regulating the communication between distal genomic elements such as enhancers and their target genes, but also to compact the DNA, which is necessary to fit the DNA inside



**Fig. 9 | Genome organization at different length scales.**

**a**, The genome is organized to compact the DNA inside the cell nucleus and to regulate interactions between distal genomic elements. At the smallest length scale (11 nm), the DNA is wrapped around nucleosomes forming a structure referred to as 'beads-on-a-string'. The DNA is further compacted in the chromatin fiber (30 nm). At higher organizational levels, DNA contacts are regulated by the formation of different 3D structures such as TADs (~300

nm) and compartments (~1  $\mu$ m). Finally, at the nuclear level chromosomes have preferred positions, referred to as chromosome territories. **b**, Two of the main mechanisms responsible for higher-order genome organization (at the 300 nm - 1  $\mu$ m length scale). DNA looping is performed by different looping factors and is thought to result in the formation of TADs. Homotypic attractions can arise from a plethora of mechanisms such as DNA binding proteins and DNA or histone modifications. Homotypic attractions are thought to be responsible for the formation of the A and B compartments.



a nucleus as stretched out DNA has a length of approximately 2 meters. While DNA compaction is mostly facilitated by processes at smaller scales such as the formation of nucleosomes and the chromatin fiber (Fig. 9), enhancer communication is regulated by processes at larger scales. Current evidence suggests that in particular two processes shape enhancer contacts: loop formation and homotypic attraction. Below, we will describe both processes and their implications for DNA organization and enhancer interactions.

Loop formation is the process in which two distal DNA sequences are brought together. Although the mechanism behind loop formation in human cells has not been definitively proven yet, the current model is that loop formation happens in a process called loop extrusion [151,152]. In this model, cohesin is loaded on the DNA together with several co-factors, followed by one-sided or two-sided DNA extrusion [153]. If unobstructed, loop extrusion would only be limited by the chromosome length or the processivity of the cohesin complex. However, the genome also contains several roadblocks that stop loop extrusion. One of the most important roadblocks are CTCF boundary sites, which comprise three repeats of a CCCTC sequence that is recognized by the CTCF protein [154]. CTCF roadblocks result in insulation as loop extrusion is now limited between two roadblocks. Insulation strength strongly depends on the CTCF boundary site. For instance, CTCF boundary sites may have different binding affinities for CTCF, resulting in different binding occupancies for distinct CTCF boundary sites. In addition, CTCF has a residence time of approximately 1-2 minutes at a boundary site [155]. Thus, it could be that CTCF binding sites are not always occupied (i.e. depending on the CTCF concentration).

Together, 'bounded loop extrusion' is thought to result on larger scales in the formation of topologically associated domains (TADs). TADs were first visualized using chromosome conformation capture technologies [156–159] and are regions in the genome, usually between 185 kB and 1 mB, in which DNA sequences preferentially interact with other sequences that lie in the same TAD. Some of the strongest evidence that bounded loop extrusion is the driving force behind TAD formation comes from studies in which key factors such as cohesin, CTCF, or cohesin co-factors were depleted, resulting in the disappearance of TADs [160–165]. While TADs at first sight seem to be static structures with fixed boundaries, several recent studies using single cell Hi-C or microscopy techniques have demonstrated that TAD boundaries are highly variable between single cells and that most loops reform many times during a single interphase [155,166–169].

Loop formation and TADs have been proposed to play an important role in shaping enhancer interactions. Direct looping between a target gene and enhancer can ensure

specific interactions between an enhancer and its target promoter, while the strong insulator sites at the boundaries of a TAD limit the search space for an enhancer and prevents formation of interactions with non-target promoters that lie outside of the TAD. The importance of looping and TADs has also been demonstrated by studies in which deletion or inversion of CTCF boundary elements disrupted enhancer communication, ultimately affecting transcription of the target gene [170,171]. However, several lines of evidence have demonstrated that loop formation and TADs are likely not the full explanation for enhancer interactions. First, while enhancer contacts are extensively rewired during differentiation, TADs appear very similar between distinct cell types [156,157,172,173]. Secondly, knock-down of essential proteins involved in loop extrusion such as cohesin and CTCF do affect TAD formation, but often only have a very mild effect on transcription [160,161,165], suggesting that enhancer communication is still (partially) intact. Taken together, looping and TAD formation cannot fully explain how promoter-enhancer contacts are established and other processes must contribute to enhancer communication.

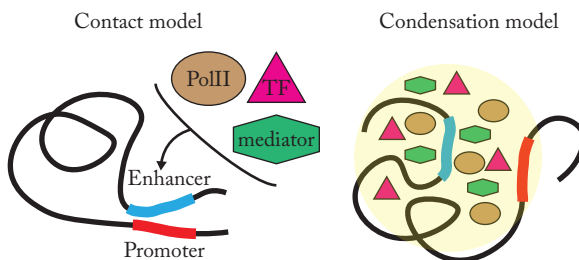
A second process that appears to play an important role in promoter-enhancer contacts is homotypic attraction [174]. Although the molecular mechanisms driving homotypic attraction remain to a large extent obscure, it results in clustering of DNA sequences with a comparable chromatin state and ultimately in the formation of an active (A) and inactive (B) compartment (i.e. compartmentalization) [175]. Many studies have investigated what type of forces could be responsible for the A and B compartmentalization. For example, it was suggested that specific proteins such as LDB1, YY1, and PRC1 can link two DNA regions and that compartmentalization is driven by differential binding of these proteins [176–178]. Alternatively, liquid-liquid phase separation was proposed to play a role in compartmentalization [179]. Together, intrinsically disordered proteins, the transcription machinery and the DNA can form condensates [180–184] to which DNA sequences with a similar chromatin state would be attracted to, possibly resulting in compartmentalization.

Loop formation and homotypic attraction underlie the formation of TADs and compartments, but how do these two processes operate together? A recent computational model argued that loop formation and homotypic attraction are competing with each other [185]. This results in a more complex picture of genome organization, in which active loop extrusion ‘disrupts’ the chromosome compartments. One could speculate that this could be necessary to bring enhancers and promoters together that would otherwise reside in distinct compartments. Taken together, loop formation and homotypic attraction seem to be two important forces in shaping the 3D genome organization and promoter-enhancer contacts, but further research will be needed to understand their intricate interplay.

### 1.4.3 Different models for transcriptional activation by enhancers

Transcriptional activation requires the communication between an enhancer and its target gene, which is likely regulated by the 3D genome organization (as we discussed in the previous section). The molecular events that follow and ultimately result in transcriptional activation are also still an area of intense debate. There are two models that explain how enhancers communicate with their target genes (Fig. 10). The first model, here referred to as the ‘contact’ model, proposes that transcriptional activation requires direct physical contact between an enhancer and its target gene. Direct contact between an enhancer and its target gene would result in delivery of essential TFs or Pol II molecules to the promoter of the target gene and increase transcriptional activity. The second model, here referred to as the ‘condensation’ model, proposes a more indirect mode of communication in which enhancers form phase-separated assemblies that comprise high concentrations of TFs and Pol II molecules [186]. Phase-separated assemblies create a favorable environment for transcription and entering of target genes into these assemblies would result in transcriptional activation.

In recent years, several observations have challenged the contact model, while further support has been found for the condensation model. First, it was demonstrated that a single enhancer can activate multiple genes simultaneously [187,188]. In the condensation model multiple target genes can be recruited to a single phase-separated assembly, explaining how a single enhancer could activate multiple genes. In contrast, in the contact model a single enhancer can only activate a single gene unless an enhancer rapidly switches between multiple target genes. Secondly, several imaging studies directly visualized the location of the enhancer and its target gene during periods of active and inactive transcription [189–192]. Remarkably, all studies found a significant distance (around 300 nm), between the enhancer and target gene during periods of active transcription further suggesting that enhancers and



**Fig. 10 | Models for transcriptional activation.**

Two distinct models for transcriptional activation of a gene by its enhancer. In the direct contact model (left), the enhancer forms a direct interaction with the promoter. Recruitment of Pol II, mediator, and different transcription factors (TFs) by the enhancer results in

activation of the promoter. In the phase separation model (right), the enhancer creates a local environment with high concentrations of Pol II, mediator, and different TFs. The local high accumulation of these molecules results in transcriptional activation of the promoter.

their target genes don't form direct contacts. Finally, a number of studies recently discovered that many TFs and Pol II itself form local accumulations in the nucleus [180–183,193,194]. These accumulations have the properties of liquid droplets, and they are highly dynamic and constantly reforming. It is tempting to speculate that these are the phase-separated assemblies that are generated at enhancers.

So, while a picture is emerging in which enhancer communication requires target genes in close proximity but not in direct contact, the question remains what then triggers transcriptional activation? One popular model suggests that transcriptional activation is regulated by the entering and leaving of the promoter in the phase separated assembly. Indeed, visualization of the *eve* locus did demonstrate that transcriptional activation was preceded by a reduction in the distance between the enhancer and reporter gene [189]. On the other hand, other studies that visualized the location of an enhancer and target gene together with its transcriptional output challenge this view [190,191]. One study showed that an enhancer and promoter didn't show any increased proximity upon transcriptional activation [191], while another study even found an increased distance between the enhancer and its target gene [190]. Taken together, the trigger for transcriptional activation seems to be complex and while some genes seem to be activated when enhancers move in close proximity, other genes seem to reside already in close proximity to their enhancers before activation. It is currently an open question what triggers transcriptional activation in these cases.

### 1.4.4 Outstanding questions

In this thesis we investigate how the LCR enhancer controls the transcriptional activity of its target genes. To study this we aim to simultaneously visualize the location of the enhancer and promoter of a reporter gene as well as its transcriptional activity. Using this approach, we want to address the following questions:

- Is promoter-enhancer proximity required for transcriptional activation?
- Do the promoter and enhancer interact transiently or are stable interactions formed?
- How do distance between promoter and enhancer, and other structural elements such as CTCF binding sites affect promoter-enhancer interactions?

## 1.5 Outline of this thesis

Regulation of gene expression determines the proteome composition of a cell, making

it a crucial process for cell identity and function. In this thesis, we will investigate three different processes that are all part of the gene expression regulatory machinery. We will study those processes using both single cell and single-molecule techniques. Although the three processes studied in this thesis are distinct, one common theme is that they demonstrate that at the single cell and single-molecule level gene expression is often heterogeneous.

In **chapter 2**, we will set out to explore the concept of heterogeneity on the single cell level in more detail. Specifically, we review recent advances made in studying translational heterogeneity of single mRNA molecules that originate from the same gene. We will see that translation can, in fact, be extensively heterogeneous as different mRNA molecules in a single cell can produce different proteins or can display different translation efficiencies.

In **chapter 3** we investigate the mechanisms of AGO2 targeting in human cells. Using a live microscopy-based assay, we visualize the cleavage of individual mRNA molecules. We find that ribosomes are important for AGO2 targeting as ribosomes remove mRNA structure. Without translating ribosomes (e.g. for AGO2 target sites present in the 3' UTR), mRNA structure often occludes the siRNA target site and target site masking can protect individual mRNA molecules for minutes from AGO2 binding and degradation.

Next, in **chapter 4** we examine the role of mRNA degradation as cells progress from mitosis into the G1 cell cycle phase. Gene expression control is essential for correct cell cycle progression as it limits the expression of certain cell cycle genes to their appropriate phase during the cell cycle. However, not much is known for a role of mRNA degradation in the cell cycle. In this work, we reveal that a large group of genes is actively degraded when cells enter the G1 cell cycle phase.

In **chapter 5** we study how the interaction between the  $\mu$ LCR enhancer and the  $\gamma$ -globin promoter influences the transcriptional output. This research is still in progress, and here we will mainly discuss the development of new tools that enable simultaneous visualization of promoter-enhancer contacts and the transcriptional status of the promoter. Together, these tools can hopefully provide new insights in the mechanism of promoter-enhancer looping and how this links to transcriptional activation.

In **chapter 6**, we will discuss the different findings of this thesis and try to put them together in a wider perspective. Finally, as a microscopy-based lab, we always look for new methods to analyze the rich data that microscopy images provide. In the

**appendix** we will describe a custom written Graphical User Interface (GUI), called TransTrack, that was developed to monitor the status of single mRNA molecules over time.

## 1.6 References

- 1 Sender, R. et al. (2016) Revised estimates for the number of human and bacteria cells in the body. *PLoS Biol.* 14, e1002533
- 2 Vickaryous, M.K. and Hall, B.K. (2006) Human cell type diversity, evolution, development, and classification with special reference to cells derived from the neural crest. *Biol. Rev. Camb. Philos. Soc.* 81, 425–455
- 3 Alberts, B. et al. (2002) *Molecular biology of the cell*, 4th edn. Garland Publishing.
- 4 Shandilya, J. and Roberts, S.G.E. (2012) The transcription cycle in eukaryotes: from productive initiation to RNA polymerase II recycling. *Biochim. Biophys. Acta - Gene Regul. Mech.* 1819, 391–400
- 5 Jackson, R.J. et al. (2010) The mechanism of eukaryotic translation initiation and principles of its regulation. *Nat. Rev. Mol. Cell Biol.* 11, 113–127
- 6 Schuller, A.P. and Green, R. (2018) Roadblocks and resolutions in eukaryotic translation. *Nat. Rev. Mol. Cell Biol.* 19, 526–541
- 7 Shapiro, E. et al. (2013) Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nat. Rev. Genet.* 14, 618–630
- 8 Brandão, H.B. et al. (2021) Tracking and interpreting long-range chromatin interactions with super-resolution live-cell imaging. *Curr. Opin. Cell Biol.* 70, 18–26
- 9 Darzacq, X. et al. (2009) Imaging transcription in living cells. *Annu. Rev. Biophys.* 38, 173–196
- 10 Vera, M. et al. (2016) Single-cell and single-molecule analysis of gene expression regulation. *Annu. Rev. Genet.* 50, 267–291
- 11 Biswas, J. et al. (2019) Fluorescence imaging methods to investigate translation in single cells. *Cold Spring Harb. Perspect. Biol.* 11, a032722
- 12 Cialek, C.A. et al. (2020) Lighting up single-mRNA translation dynamics in living cells. *Curr. Opin. Genet. Dev.* 61, 75–82
- 13 Pichon, X. et al. (2018) A growing toolbox to image gene expression in single cells: sensitive approaches for demanding challenges. *Mol. Cell* 71, 468–480
- 14 Munsky, B. et al. (2012) Using gene expression noise to understand gene regulation. *Science* 336, 183–187
- 15 Larson, D.R. (2011) What do expression dynamics tell us about the mechanism of transcription? *Curr. Opin. Genet. Dev.* 21, 591–599
- 16 Femino, A.M. et al. (1998) Visualization of single RNA transcripts in situ. *Science* 280, 585–590
- 17 Raj, A. et al. (2006) Stochastic mRNA synthesis in mammalian cells. *PLoS Biol.* 4, 1707–1719
- 18 Lenstra, T.L. et al. (2016) Transcription dynamics in living cells. *Annu. Rev. Biophys.* 45, 25–47
- 19 Grün, D. et al. (2015) Single-cell messenger RNA sequencing reveals rare intestinal cell types. *Nature* 525, 251–255
- 20 Boersma, S. et al. (2019) Multi-color single-molecule imaging uncovers extensive heterogeneity in mRNA decoding. *Cell* 178, 458–472
- 21 Bartel, D.P. (2018) Metazoan microRNAs. *Cell* 173, 20–51
- 22 Gebert, L.F.R. and MacRae, I.J. (2019) Regulation of microRNA function in animals. *Nat. Rev. Mol. Cell Biol.* 20, 21–37

- 23 Ghildiyal, M. and Zamore, P.D. (2009) Small silencing RNAs: an expanding universe. *Nat. Rev. Genet.* 10, 94–108
- 24 Malone, C.D. and Hannon, G.J. (2009) Small RNAs as guardians of the genome. *Cell* 136, 656–668
- 25 Ozata, D.M. et al. (2019) PIWI-interacting RNAs: small RNAs with big functions. *Nat. Rev. Genet.* 20, 89–108
- 26 Hammond, S.M. et al. (2000) An RNA-directed nuclease mediates post-transcriptional gene silencing in *Drosophila* cells. *Nature* 404, 293–296
- 27 Liu, J. et al. (2004) Argonaute2 is the catalytic engine of mammalian RNAi. *Science* 305, 1437–1441
- 28 Meister, G. et al. (2004) Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Mol. Cell* 15, 185–197
- 29 Rivas, F. V. et al. (2005) Purified Argonaute2 and an siRNA form recombinant human RISC. *Nat. Struct. Mol. Biol.* 12, 340–349
- 30 Martinez, J. et al. (2002) Single-stranded antisense siRNAs guide target RNA cleavage in RNAi. *Cell* 110, 563–574
- 31 Jonas, S. and Izaurralde, E. (2015) Towards a molecular understanding of microRNA-mediated gene silencing. *Nat. Rev. Genet.* 16, 421–433
- 32 Iwasaki, Y.W. et al. (2015) PIWI-interacting RNA: its biogenesis and functions. *Annu. Rev. Biochem.* 84, 405–433
- 33 Meister, G. (2013) Argonaute proteins: functional insights and emerging roles. *Nat. Rev. Genet.* 14, 447–459
- 34 Schirle, N.T. et al. (2014) Structural basis for microRNA targeting. *Science* 346, 608–613
- 35 Song, J.J. et al. (2004) Crystal structure of Argonaute and its implications for RISC slicer activity. *Science* 305, 1434–1437
- 36 Czech, B. et al. (2008) An endogenous small interfering RNA pathway in *Drosophila*. *Nature* 453, 798–802
- 37 Ghildiyal, M. et al. (2008) Endogenous siRNAs derived from transposons and mRNAs in *Drosophila* somatic cells. *Science* 320, 1077–1081
- 38 Kawamura, Y. et al. (2008) *Drosophila* endogenous small RNAs bind to Argonaute 2 in somatic cells. *Nature* 453, 793–797
- 39 Ding, S.W. and Voinnet, O. (2007) Antiviral immunity directed by small RNAs. *Cell* 130, 413–426
- 40 Maillard, P. V. et al. (2013) Antiviral RNA interference in mammalian cells. *Science* 342, 235–238
- 41 Elbashir, S.M. et al. (2001) Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells. *Nature* 411, 494–498
- 42 (2020) Second RNAi drug approved. *Nat. Biotechnol.* 38, 385
- 43 Wee, L.M. et al. (2012) Argonaute divides its RNA guide into domains with distinct functions and RNA-binding properties. *Cell* 151, 1055–1067
- 44 Salomon, W.E. et al. (2015) Single-molecule imaging reveals that Argonaute reshapes the binding properties of its nucleic acid guides. *Cell* 162, 84–95
- 45 Chandradoss, S.D. et al. (2015) A dynamic search process underlies microRNA targeting. *Cell* 162, 96–107
- 46 Jo, M.H. et al. (2015) Human Argonaute2 has diverse reaction pathways on target RNAs. *Mol. Cell* 59, 117–124
- 47 Becker, W.R. et al. (2019) High-throughput analysis reveals rules for target RNA binding and cleavage by AGO2. *Mol. Cell* 75, 741–755
- 48 McGeary, S.E. et al. (2019) The biochemical basis of microRNA targeting efficacy. *Science* 366, eaav1741
- 49 Schirle, N.T. and MacRae, I.J. (2012) The crystal structure of human Argonaute2. *Science* 336, 1037–1040
- 50 Sheu-Gruttadauria, J. and MacRae, I.J. (2017) Structural foundations of RNA silencing

- by Argonaute. *J. Mol. Biol.* 429, 2619–2639
- 51 Kobayashi, H. and Tomari, Y. (2016) RISC assembly: coordination between small RNAs and Argonaute proteins. *Biochim. Biophys. Acta* 1859, 71–81
- 52 Frank, F. et al. (2010) Structural basis for 5'-nucleotide base-specific recognition of guide RNA by human AGO2. *Nature* 465, 818–822
- 53 Parker, J.S. et al. (2005) Structural insights into mRNA recognition from a PIWI domain-siRNA guide complex. *Nature* 434, 663–666
- 54 Ma, J.B. et al. (2005) Structural basis for 5' -end-specific recognition of guide RNA by the *A. fulgidus* Piwi protein. *Nature* 434, 666–670
- 55 Yao, C. et al. (2015) Single-molecule analysis of the target cleavage reaction by the drosophila RNAi enzyme complex. *Mol. Cell* 59, 125–132
- 56 Schirle, N.T. et al. (2015) Water-mediated recognition of t1-adenosine anchors Argonaute2 to microRNA targets. *Elife* 4, e07646
- 57 Qu, X. et al. (2011) The ribosome uses two active mechanisms to unwind messenger RNA during translation. *Nature* 475, 118–121
- 58 Grimson, A. et al. (2007) MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol. Cell* 27, 91–105
- 59 Gu, S. et al. (2009) Biological basis for restriction of microRNA targets to the 3' untranslated region in mammalian mRNAs. *Nat. Struct. Mol. Biol.* 16, 144–150
- 60 Zhang, C. et al. (2012) Roles of target site location and sequence complementarity in trans-acting siRNA formation in *Arabidopsis*. *Plant J.* 69, 217–226
- 61 Hentze, M.W. et al. (2018) A brave new world of RNA-binding proteins. *Nat. Rev. Mol. Cell Biol.* 19, 327–341
- 62 Jankowsky, E. and Harris, M.E. (2015) Specificity and nonspecificity in RNA-protein interactions. *Nat. Rev. Mol. Cell Biol.* 16, 533–544
- 63 Bhattacharyya, S.N. et al. (2006) Relief of microRNA-mediated translational repression in human cells subjected to stress. *Cell* 125, 1111–1124
- 64 Kedde, M. et al. (2007) RNA-binding protein Dnd1 inhibits microRNA access to target mRNA. *Cell* 131, 1273–1286
- 65 Brown, K.M. et al. (2005) Target accessibility dictates the potency of human RISC. *Nat. Struct. Mol. Biol.* 12, 469–470
- 66 Kertesz, M. et al. (2007) The role of site accessibility in microRNA target recognition. *Nat. Genet.* 39, 1278–1284
- 67 Tafer, H. et al. (2008) The impact of target site accessibility on the design of effective siRNAs. *Nat. Biotechnol.* 26, 578–583
- 68 Ameres, S.L. et al. (2007) Molecular basis for target RNA recognition and cleavage by human RISC. *Cell* 130, 101–112
- 69 Cui, T.J. et al. (2019) Argonaute bypasses cellular obstacles without hindrance during target search. *Nat. Commun.* 10, 4390
- 70 Sponer, J. et al. (2018) RNA structural dynamics as captured by molecular simulations: a comprehensive overview. *Chem. Rev.* 118, 4177–4338
- 71 Herschlag, D. (1995) RNA chaperones and the RNA folding problem. *J. Biol. Chem.* 270, 20871–20874
- 72 Ganser, L.R. et al. (2019) The roles of structural dynamics in the cellular functions of RNAs. *Nat. Rev. Mol. Cell Biol.* 20, 474–489
- 73 Abelson, J. et al. (2010) Conformational dynamics of single pre-mRNA molecules during in vitro splicing. *Nat. Struct. Mol. Biol.* 17, 504–512
- 74 Woods, C.T. et al. (2017) Comparative visualization of the RNA suboptimal conformational ensemble in vivo. *Biophys. J.* 113, 290–301
- 75 Spasic, A. et al. (2018) Modeling RNA secondary structure folding ensembles using SHAPE mapping data. *Nucleic Acids Res.* 46, 314–323



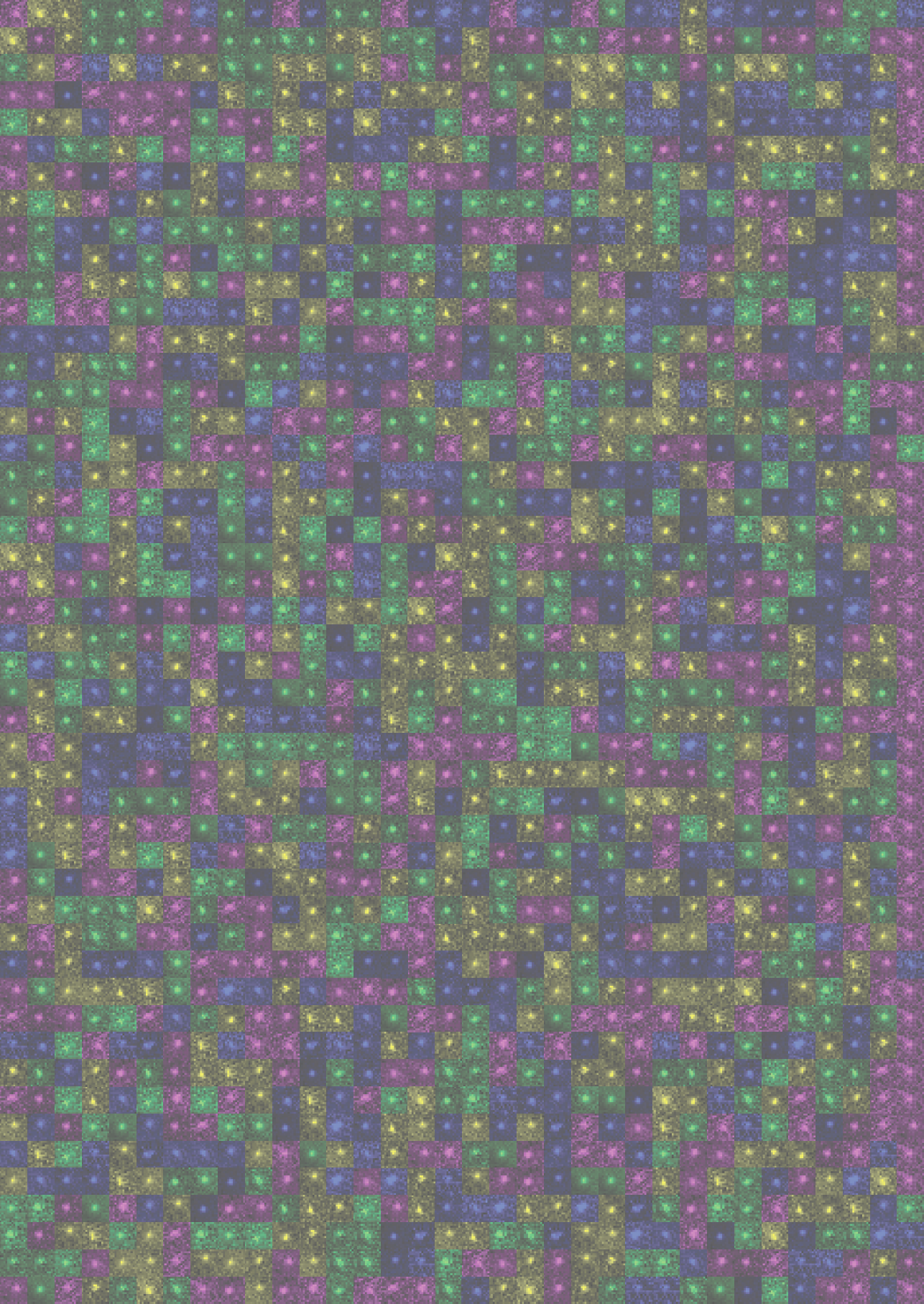
- 76 Li, H. and Aviran, S. (2018) Statistical modeling of RNA structure profiling experiments enables parsimonious reconstruction of structure landscapes. *Nat. Commun.* 9, 606
- 77 Frank, A.T. et al. (2009) Constructing RNA dynamical ensembles by combining MD and motionally decoupled NMR RDCs: new insights into RNA dynamics and adaptive ligand recognition. *Nucleic Acids Res.* 37, 3670–3679
- 78 Solomatin, S. V. et al. (2010) Multiple native states reveal persistent ruggedness of an RNA folding landscape. *Nature* 463, 681–684
- 79 Ditzler, M.A. et al. (2008) A rugged free energy landscape separates multiple functional RNA folds throughout denaturation. *Nucleic Acids Res.* 36, 7088–7099
- 80 Rinaldi, A.J. et al. (2016) The Shine-Dalgarno sequence of riboswitch-regulated single mRNAs shows ligand-dependent accessibility bursts. *Nat. Commun.* 7, 8976
- 81 Rouskin, S. et al. (2014) Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature* 505, 701–705
- 82 Jankowsky, E. (2011) RNA helicases at work: binding and rearranging. *Trends Biochem. Sci.* 36, 19–29
- 83 Bourgeois, C.F. et al. (2016) The multiple functions of RNA helicases as drivers and regulators of gene expression. *Nat. Rev. Mol. Cell Biol.* 17, 426–438
- 84 Beaudoin, J.-D. et al. (2018) Analyses of mRNA structure dynamics identify embryonic gene regulatory programs. *Nat. Struct. Mol. Biol.* 25, 677–686
- 85 Mustoe, A.M. et al. (2018) Pervasive regulatory functions of mRNA structure revealed by high-resolution SHAPE probing. *Cell* 173, 181–195
- 86 Adivarahan, S. et al. (2018) Spatial organization of single mRNPs at different stages of the gene expression pathway. *Mol. Cell* 72, 727–738
- 87 Mizrahi, O. et al. (2018) Virus-induced changes in mRNA secondary structure uncover cis-regulatory elements that directly control gene expression. *Mol. Cell* 72, 862–874
- 88 Morgan, D.O. (2007) *The cell cycle: principles of control*, New Science Press.
- 89 Hanahan, D. and Weinberg, R.A. (2011) *Hallmarks of cancer: the next generation*. *Cell* 144, 646–674
- 90 Malumbres, M. and Barbacid, M. (2001) To cycle or not to cycle: a critical decision in cancer. *Nat. Rev. Cancer* 1, 222–231
- 91 Bansbach, C.E. and Cortez, D. (2011) Defining genome maintenance pathways using functional genomic approaches. *Crit. Rev. Biochem. Mol. Biol.* 46, 327–341
- 92 Bar-Joseph, Z. et al. (2008) Genome-wide transcriptional analysis of the human cell cycle identifies genes differentially regulated in normal and cancer cells. *Proc. Natl. Acad. Sci. U. S. A.* 105, 955–960
- 93 Whitfield, M.L. et al. (2002) Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Mol. Biol. Cell* 13, 1977–2000
- 94 Cho, R.J. et al. (2001) Transcriptional regulation and function during the human cell cycle. *Nat. Genet.* 27, 48–54
- 95 Schwabe, D. et al. (2020) The transcriptome dynamics of single cells during the cell cycle. *Mol. Syst. Biol.* 16, e9946
- 96 Xia, C. et al. (2019) Spatial transcriptome profiling by MERFISH reveals subcellular RNA compartmentalization and cell cycle-dependent gene expression. *Proc. Natl. Acad. Sci. U. S. A.* 116, 19490–19499
- 97 Bertoli, C. et al. (2013) Control of cell cycle transcription during G1 and S phases. *Nat. Rev. Mol. Cell Biol.* 14, 518–528
- 98 Nakayama, K.I. and Nakayama, K. (2006) Ubiquitin ligases: cell-cycle control and cancer. *Nat. Rev. Cancer* 6, 369–381
- 99 Vodermaier, H.C. (2004) APC/C and SCF: controlling each other and the cell cycle. *Curr. Biol.* 14, 787–796

- 100 Kronja, I. and Orr-Weaver, T.L. (2011) Translational regulation of the cell cycle: when, where, how and why? *Philos. Trans. R. Soc. B Biol. Sci.* 366, 3638–3652
- 101 Stumpf, C.R. et al. (2013) The translational landscape of the mammalian cell cycle. *Mol. Cell* 52, 574–582
- 102 Tanenbaum, M.E. et al. (2015) Regulation of mRNA translation during mitosis. *Elife* 4, e07957
- 103 Eser, P. et al. (2014) Periodic mRNA synthesis and degradation co-operate during cell cycle gene expression. *Mol. Syst. Biol.* 10, 717
- 104 Battich, N. et al. (2020) Sequencing metabolically labeled transcripts in single cells reveals mRNA turnover strategies. *Science* 367, 1151–1156
- 105 Trcek, T. et al. (2011) Single-molecule mRNA decay measurements reveal promoter-regulated mRNA stability in yeast. *Cell* 147, 1484–1497
- 106 Rabani, M. et al. (2014) High-resolution sequencing and modeling identifies distinct dynamic RNA regulatory strategies. *Cell* 159, 1698–1710
- 107 Elkon, R. et al. (2010) Major role for mRNA stability in shaping the kinetics of gene induction. *BMC Genomics* 11, 259
- 108 Shalem, O. et al. (2008) Transient transcriptional responses to stress are generated by opposing effects of mRNA production and degradation. *Mol. Syst. Biol.* 4, 223
- 109 Pérez-Ortín, J.E. et al. (2007) Genomics and gene transcription kinetics in yeast. *Trends Genet.* 23, 250–257
- 110 Garneau, N.L. et al. (2007) The highways and byways of mRNA decay. *Nat. Rev. Mol. Cell Biol.* 8, 113–126
- 111 Nagarajan, V.K. et al. (2013) XRN 5'-3' exoribonucleases: structure, mechanisms and functions. *Biochim. Biophys. Acta - Gene Regul. Mech.* 1829, 590–603
- 112 Chlebowski, A. et al. (2013) RNA decay machines: the exosome. *Biochim. Biophys. Acta - Gene Regul. Mech.* 1829, 552–560
- 113 Brown, C.E. et al. (1996) PAN3 encodes a subunit of the Pab1p-dependent poly(A) nuclease in *Saccharomyces cerevisiae*. *Mol. Cell Biol.* 16, 5744–5753
- 114 Tucker, M. et al. (2001) The transcription factor associated Ccr4 and Caf1 proteins are components of the major cytoplasmic mRNA deadenylase in *Saccharomyces cerevisiae*. *Cell* 104, 377–386
- 115 Dehlin, E. et al. (2000) Cap-dependent deadenylation of mRNA. *EMBO J.* 19, 1079–1086
- 116 Steiger, M. et al. (2003) Analysis of recombinant yeast decapping enzyme. *RNA* 9, 231–238
- 117 He, F. et al. (2003) Genome-wide analysis of mRNAs regulated by the nonsense-mediated and 5' to 3' mRNA decay pathways in yeast. *Mol. Cell* 12, 1439–1452
- 118 Houalla, R. et al. (2006) Microarray detection of novel nuclear RNA substrates for the exosome. *Yeast* 23, 439–454
- 119 Parker, R. (2012) RNA degradation in *Saccharomyces cerevisiae*. *Genetics* 191, 671–702
- 120 Tuck, A.C. et al. (2020) Mammalian RNA decay pathways are highly specialized and widely linked to translation. *Mol. Cell* 77, 1222–1236
- 121 Khabar, K.S.A. (2005) The AU-rich transcriptome: more than interferons and cytokines, and its role in disease. *J. Interf. Cytokine Res.* 25, 1–10
- 122 Spasic, M. et al. (2012) Genome-wide assessment of AU-rich elements by the AREScore algorithm. *PLoS Genet.* 8, e1002433
- 123 Webster, M.W. et al. (2019) RNA-binding proteins distinguish between similar sequence motifs to promote targeted deadenylation by Ccr4-Not. *Elife* 8, e40670
- 124 Bertrand, E. et al. (1998) Localization of ASH1 mRNA particles in living yeast. *Mol. Cell* 2, 437–445
- 125 Larson, D.R. et al. (2011) Real-time

- observation of transcription initiation and elongation on an endogenous yeast gene. *Science* 332, 475–478
- 126 Chubb, J.R. et al. (2006) Transcriptional pulsing of a developmental gene. *Curr. Biol.* 16, 1018–1025
- 127 Hager, G.L. et al. (2009) Transcription dynamics. *Mol. Cell* 35, 741–753
- 128 Suter, D.M. et al. (2011) Mammalian genes are transcribed with widely different bursting kinetics. *Science* 332, 472–474
- 129 Dar, R.D. et al. (2012) Transcriptional burst frequency and burst size are equally modulated across the human genome. *Proc. Natl. Acad. Sci. U. S. A.* 109, 17454–17459
- 130 Bartman, C.R. et al. (2016) Enhancer regulation of transcriptional bursting parameters revealed by forced chromatin looping. *Mol. Cell* 62, 237–247
- 131 Padovan-Merhar, O. et al. (2015) Single mammalian cells compensate for differences in cellular volume and DNA copy number through independent global transcriptional mechanisms. *Mol. Cell* 58, 339–352
- 132 Furlong, E.E.M. and Levine, M. (2018) Developmental enhancers and chromosome topology. *Science* 361, 1341–1345
- 133 Long, H.K. et al. (2016) Ever-changing landscapes: transcriptional enhancers in development and evolution. *Cell* 167, 1170–1187
- 134 Levine, M. (2010) Transcriptional enhancers in animal development and evolution. *Curr. Biol.* 20, R754–R763
- 135 Spielmann, M. et al. (2018) Structural variation in the 3D genome. *Nat. Rev. Genet.* 19, 453–467
- 136 Smith, E. and Shilatifard, A. (2014) Enhancer biology and enhanceropathies. *Nat. Struct. Mol. Biol.* 21, 210–219
- 137 Wittkopp, P.J. and Kalay, G. (2012) Cis-regulatory elements: molecular mechanisms and evolutionary processes underlying divergence. *Nat. Rev. Genet.* 13, 59–69
- 138 Banerji, J. et al. (1981) Expression of a  $\beta$ -globin gene is enhanced by remote SV40 DNA sequences. *Cell* 27, 299–308
- 139 Atchison, M.L. (1988) Enhancers: mechanisms of action and cell specificity. *Annu. Rev. Cell Biol.* 4, 127–153
- 140 Crawford, G.E. et al. (2006) Genome-wide mapping of DNase hypersensitive sites using massively parallel signature sequencing (MPSS). *Genome Res.* 16, 123–131
- 141 Creighton, M.P. et al. (2010) Histone H3K27ac separates active from poised enhancers and predicts developmental state. *Proc. Natl. Acad. Sci. U. S. A.* 107, 21931–21936
- 142 Rada-Iglesias, A. et al. (2011) A unique chromatin signature uncovers early developmental enhancers in humans. *Nature* 470, 279–285
- 143 Zentner, G.E. et al. (2011) Epigenetic signatures distinguish multiple classes of enhancers with distinct cellular functions. *Genome Res.* 21, 1273–1283
- 144 The ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74
- 145 Shen, Y. et al. (2012) A map of the cis-regulatory sequences in the mouse genome. *Nature* 488, 116–120
- 146 Kwasnieski, J.C. et al. (2014) High-throughput functional testing of ENCODE segmentation predictions. *Genome Res.* 24, 1595–1602
- 147 Catarino, R.R. and Stark, A. (2018) Assessing sufficiency and necessity of enhancer activities for gene expression and the mechanisms of transcription activation. *Genes Dev.* 32, 202–223
- 148 Osterwalder, M. et al. (2018) Enhancer redundancy provides phenotypic robustness in mammalian development. *Nature* 554, 239–243
- 149 Andrey, G. and Mundlos, S. (2017) The three-dimensional genome: regulating gene expression during pluripotency and development. *Dev.* 144, 3646–3658

- 150 Anderson, E. et al. (2014) Mapping the Shh long-range regulatory domain. *Dev.* 141, 3934–3943
- 151 Sanborn, A.L. et al. (2015) Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes. *Proc. Natl. Acad. Sci. U. S. A.* 112, E6456–E6465
- 152 Fudenberg, G. et al. (2016) Formation of chromosomal domains by loop extrusion. *Cell Rep.* 15, 2038–2049
- 153 Davidson, I.F. et al. (2019) Human cohesin compacts DNA by loop extrusion. *Science* 366, 1338–1345
- 154 Ghirlando, R. and Felsenfeld, G. (2016) CTCF: making the right connections. *Genes Dev.* 30, 881–891
- 155 Hansen, A.S. et al. (2017) CTCF and cohesin regulate chromatin loop stability with distinct dynamics. *Elife* 6, e25776
- 156 Dixon, J.R. et al. (2012) Topological domains in mammalian genomes identified by analysis of chromatin interactions. *Nature* 485, 376–380
- 157 Nora, E.P. et al. (2012) Spatial partitioning of the regulatory landscape of the X-inactivation centre. *Nature* 485, 381–385
- 158 Sexton, T. et al. (2012) Three-dimensional folding and functional organization principles of the *Drosophila* genome. *Cell* 148, 458–472
- 159 Rao, S.S.P. et al. (2014) A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping. *Cell* 159, 1665–1680
- 160 Rao, S.S.P. et al. (2017) Cohesin loss eliminates all loop domains. *Cell* 171, 305–320
- 161 Schwarzer, W. et al. (2017) Two independent modes of chromatin organization revealed by cohesin removal. *Nature* 551, 51–56
- 162 Wutz, G. et al. (2017) Topologically associating domains and chromatin loops depend on cohesin and are regulated by CTCF, WAPL, and PDS5 proteins. *EMBO J.* 36, 3573–3599
- 163 Gassler, J. et al. (2017) A mechanism of cohesin-dependent loop extrusion organizes zygotic genome architecture. *EMBO J.* 36, 3600–3618
- 164 Haarhuis, J.H.I. et al. (2017) The cohesin release factor WAPL restricts chromatin loop extension. *Cell* 169, 693–707
- 165 Nora, E.P. et al. (2017) Targeted degradation of CTCF decouples local insulation of chromosome domains from genomic compartmentalization. *Cell* 169, 930–944
- 166 Cattoni, D.I. et al. (2017) Single-cell absolute contact probability detection reveals chromosomes are organized by multiple low-frequency yet specific interactions. *Nat. Commun.* 8, 1753
- 167 Nagano, T. et al. (2013) Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* 502, 59–64
- 168 Finn, E.H. et al. (2019) Extensive heterogeneity and intrinsic variation in spatial genome organization. *Cell* 176, 1502–1515
- 169 Bintu, B. et al. (2018) Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science* 362, eaau1783
- 170 Guo, Y. et al. (2015) CRISPR inversion of CTCF sites alters genome topology and enhancer/promoter function. *Cell* 162, 900–910
- 171 Lupiáñez, D.G. et al. (2015) Disruptions of topological chromatin domains cause pathogenic rewiring of gene-enhancer interactions. *Cell* 161, 1012–1025
- 172 Smith, E.M. et al. (2016) Invariant TAD boundaries constrain cell-type-specific looping interactions between promoters and distal elements around the CFTR locus. *Am. J. Hum. Genet.* 98, 185–201
- 173 Harmston, N. et al. (2017) Topologically associating domains are ancient features that coincide with Metazoan clusters of extreme noncoding conservation. *Nat. Commun.* 8, 441
- 174 Robson, M.I. et al. (2019) Regulatory landscaping: how enhancer-promoter communication is sculpted in 3D. *Mol. Cell* 74,

- 1110–1122
- 175 Lieberman-Aiden, E. et al. (2009) Comprehensive mapping of long-range interactions reveals folding principles of the Human genome. *Science* 326, 289–294
- 176 Weintraub, A.S. et al. (2017) YY1 Is a structural regulator of enhancer–promoter loops. *Cell* 171, 1573–1579
- 177 Isono, K. et al. (2013) SAM domain polymerization links subnuclear clustering of PRC1 to gene silencing. *Dev. Cell* 26, 565–577
- 178 Pombo, A. and Dillon, N. (2015) Three-dimensional genome architecture: players and mechanisms. *Nat. Rev. Mol. Cell Biol.* 16, 245–257
- 179 Shin, Y. et al. (2018) Liquid nuclear condensates mechanically sense and restructure the genome. *Cell* 175, 1481–1491
- 180 Boija, A. et al. (2018) Transcription factors activate genes through the phase-separation capacity of their activation domains. *Cell* 175, 1842–1855
- 181 Sabari, B.R. et al. (2018) Coactivator condensation at super-enhancers links phase separation and gene control. *Science* 361, eaar3958
- 182 Cho, W.K. et al. (2018) Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* 361, 412–415
- 183 Chong, S. et al. (2018) Imaging dynamic and selective low-complexity domain interactions that control gene transcription. *Science* 361, eaar2555
- 184 Strom, A.R. et al. (2017) Phase separation drives heterochromatin domain formation. *Nature* 547, 241–245
- 185 Nuebler, J. et al. (2018) Chromatin organization by an interplay of loop extrusion and compartmental segregation. *Proc. Natl. Acad. Sci.* 115, E6697–E6706
- 186 Hnisz, D. et al. (2017) A phase separation model for transcriptional control. *Cell* 169, 13–23
- 187 Fukaya, T. et al. (2016) Enhancer control of transcriptional bursting. *Cell* 166, 358–368
- 188 Lim, B. et al. (2018) Visualization of transvection in living *Drosophila* embryos. *Mol. Cell* 70, 287–296
- 189 Chen, H. et al. (2018) Dynamic interplay between enhancer–promoter topology and gene activity. *Nat. Genet.* 50, 1296–1303
- 190 Benabdallah, N.S. et al. (2019) Decreased enhancer–promoter proximity accompanying enhancer activation. *Mol. Cell* 76, 473–484
- 191 Alexander, J.M. et al. (2019) Live-cell imaging reveals enhancer-dependent Sox2 transcription in the absence of enhancer proximity. *Elife* 8, e41769
- 192 Heist, T. et al. (2019) Large distances separate coregulated genes in living *Drosophila* embryos. *Proc. Natl. Acad. Sci. U. S. A.* 116, 15062–15067
- 193 Han, X. et al. (2020) Roles of the BRD4 short isoform in phase separation and active gene transcription. *Nat. Struct. Mol. Biol.* 27, 333–341
- 194 Cisse, I.I. et al. (2013) Real-time dynamics of RNA Polymerase II clustering in live human cells. *Science* 341, 664–667
- 195 Pettersen, E.F. et al. (2004) UCSF Chimera: a visualization system for exploratory research and analysis. *J. Comput. Chem.* 25, 1605–1612



# 2

## Heterogeneity in mRNA translation

Stijn Sonneveld<sup>1,2</sup>, Bram M. P. Verhagen<sup>1,2</sup>, Marvin E. Tanenbaum<sup>1,\*</sup>

<sup>1</sup> Oncode Institute, Hubrecht Institute–KNAW and University Medical Center Utrecht, Utrecht, The Netherlands

<sup>2</sup> These authors contributed equally to this work.

\* Corresponding author and lead contact. Email address for correspondence: [m.tanenbaum@hubrecht.eu](mailto:m.tanenbaum@hubrecht.eu).

Adapted from: Sonneveld, S., Verhagen, B.M.P., Tanenbaum, M.E.. Heterogeneity in mRNA translation. *Trends Cell Biol* 30, 606-618 (2020).  
<https://doi.org/10.1016/j.tcb.2020.04.008>

Published online 25 May 2020.

## 2.1 Abstract

During mRNA translation, the genetic information stored in mRNA is translated into protein sequence. It is imperative that the genetic information is translated with high precision. Surprisingly however, recent experimental evidence has demonstrated that translation can be highly heterogeneous, even among different mRNA molecules derived from a single gene in an individual cell; multiple different polypeptides can be produced from a single mRNA molecule, and the rate of translation can vary in both space and time. However, whether translational heterogeneity serves an important cellular function, or rather predominantly represents gene expression ‘noise’ remains an open question. In this review, we will discuss the molecular basis and potential functions of such translational heterogeneity.

## 2.2 Highlights

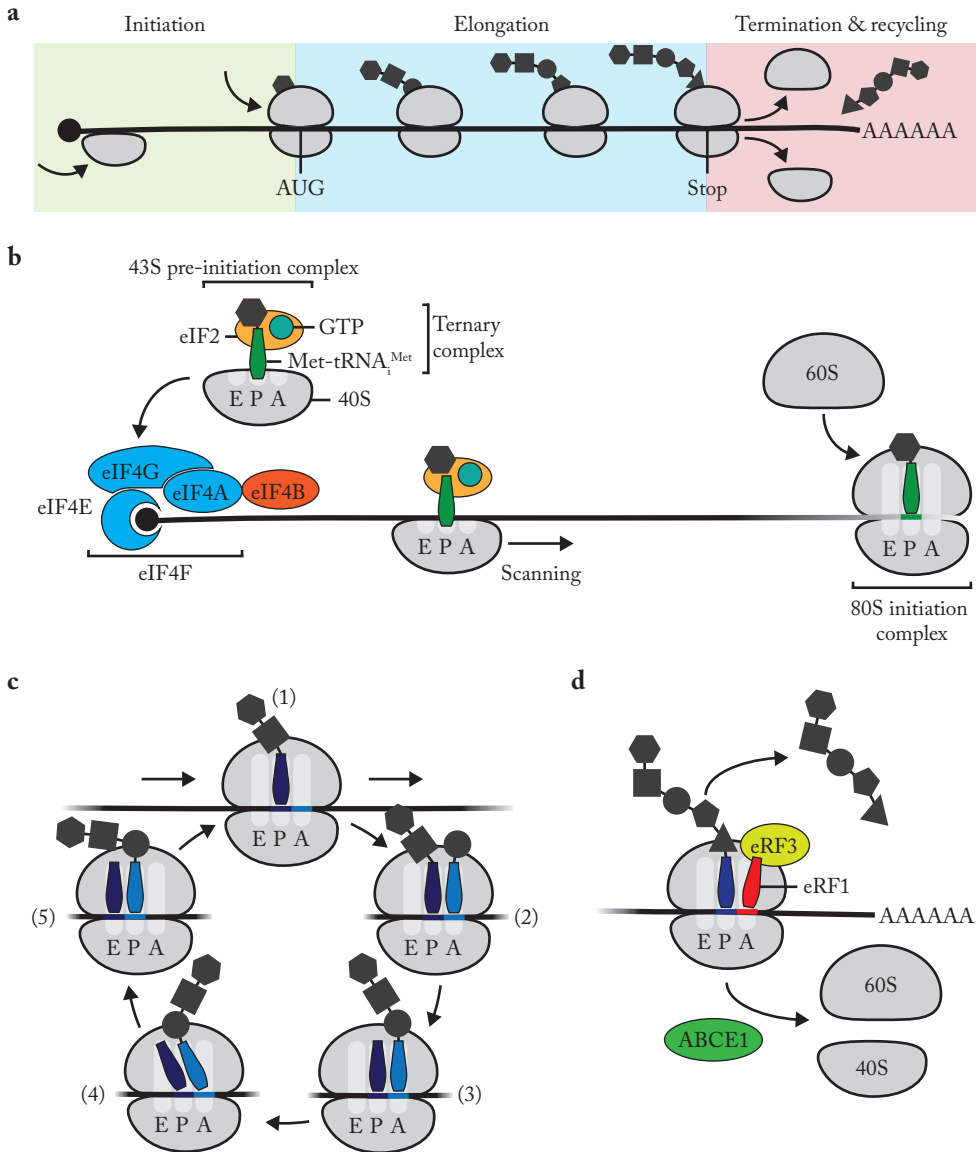
- Different types of translational heterogeneity can be distinguished: cell or tissue heterogeneity (translation of a gene is different in distinct cell or tissue types), ‘intergenic’ heterogeneity (mRNAs derived from different genes are translated differentially), or ‘intragenic’ heterogeneity (different mRNAs derived from one gene in a single cell show translational heterogeneity).
- Recent studies using single-molecule imaging have revealed widespread occurrence of intragenic translational heterogeneity.
- Intragenic translational heterogeneity can have multiple origins, including heterogeneity in primary mRNA sequence, RNA binding proteins, RNA modifications, mRNA structure and ribosome composition.
- Translational heterogeneity likely represents both ‘noise’, as well as spatiotemporal regulation of translation.

## 2.3 Mechanisms of translational control

The fate and function of each cell depends on its protein composition. Therefore, accurate gene expression control is critical for proper cell functioning. A key step in gene expression is mRNA translation (Box 1 and Fig. 1), during which the genetic information stored in mRNA is decoded. Regulation of translation can affect both protein sequence and abundance [1]. Moreover, translational regulation is fast, reversible, and provides spatial control, making it a unique regulatory mechanism of gene expression. In a simplistic view, each mRNA molecule encodes a single protein and mRNAs derived from the same gene are decoded in the same way. However, it is



now clear that translation is far more heterogeneous. Different types of translational heterogeneity can be distinguished. First, a single mRNA species can be translated differentially in different cell types, resulting in ‘cell-to-cell’ translational heterogeneity, which is important during different cellular processes such as differentiation [2].



**Fig. 1 | Mechanism of translation initiation, elongation and termination.**

**a**, Overview of the different phases of translation. **b-d**, Schematic of translation initiation (b), translation elongation (c), and translation termination and recycling (d).

**Box 1 | Mechanism of translation initiation, elongation and termination.**

**Initiation.** The 43S pre-initiation complex (PIC) is composed of the small (40S) ribosomal subunit, GTP-bound eIF2 and the initiator methionine tRNA (Met-tRNA<sub>i</sub>) along with several other translation initiation factors (Fig. 1b). The PIC is recruited to the 7-methylguanosine cap of an mRNA, a process that is mediated by an interaction of cap-bound eIF4E, eIF4G, and eIF4A translation initiation factors and the PIC. After recruitment to the cap, the PIC scans the mRNA from 5' to 3' in search of a start codon (Fig. 1b). Upon base pairing between the start codon and the Met-tRNA<sub>i</sub>, the PIC undergoes a conformational change, resulting in the recruitment of the large (60S) ribosomal subunit and the start of the translation elongation phase (Fig. 1b).

**Elongation.** During elongation, the open reading frame (ORF) is decoded in steps of three nucleotides (one codon) to synthesize a polypeptide chain. The elongation cycle starts with a peptidyl-tRNA in the P site, and an empty aminoacyl and exit site (A and E site, respectively) (Fig. 1c, step 1). First, an aminoacyl-tRNA is selected at the empty A site by base pairing of the tRNA anticodon and the mRNA codon (Fig. 1c, step 2). Next, a peptide bond is formed in the peptidyl transferase center between the nascent peptide and the aminoacyl-tRNA, resulting in a deacylated-tRNA and a new peptidyl-tRNA, respectively (Fig. 1c, step 3). Finally, the tRNAs adopt a hybrid state (Fig. 1c, step 4), followed by eEF2-mediated translocation of the deacylated-tRNA and the new peptidyl-tRNA to the E and P site, respectively (Fig. 1c, step 5). The elongation cycle is repeated until a stop codon is encountered.

**Termination and recycling.** Translation is terminated upon recognition of a stop codon in the A site by release factors eRF1 and eRF3 (Fig. 1d). After binding to the stop codon in the A site, eRF1 induces release of the polypeptide and stimulates recruitment of the ribosome recycling factor ABCE1, which, in turn, results in dissociation of the ribosome subunits from the mRNA.

Second, mRNA molecules originating from different genes can be differentially translated in a single cell, resulting in 'intergenic' translational heterogeneity. Third, mRNA molecules originating from a single gene in a single cell can also display translational heterogeneity, which we will refer to as 'intragenic' heterogeneity. Finally, a single mRNA molecule can also be translated differentially over time, representing a special case of intragenic heterogeneity. While regulatory functions for cell-to-cell or intergenic translational heterogeneity are well known, potential functions and mechanisms of regulation for intragenic translational heterogeneity are less evident.

In this review, we will focus on intragenic translational heterogeneity, a field that has rapidly emerged in recent years, in part due to the development of new methods that provide sufficient detection sensitivity to study translation of single mRNA molecules. First, we will describe experimental evidence supporting the existence of intragenic translational heterogeneity, as well as its potential functions. Next, we will discuss the molecular mechanisms underlying intragenic translational heterogeneity. We will also discuss whether intragenic translational heterogeneity represents (functional) translational regulation or whether it is a consequence of the variability inherent to translational regulation and mainly reflects 'noise' in the system.

## 2.4 Experimental evidence for heterogeneity in translation

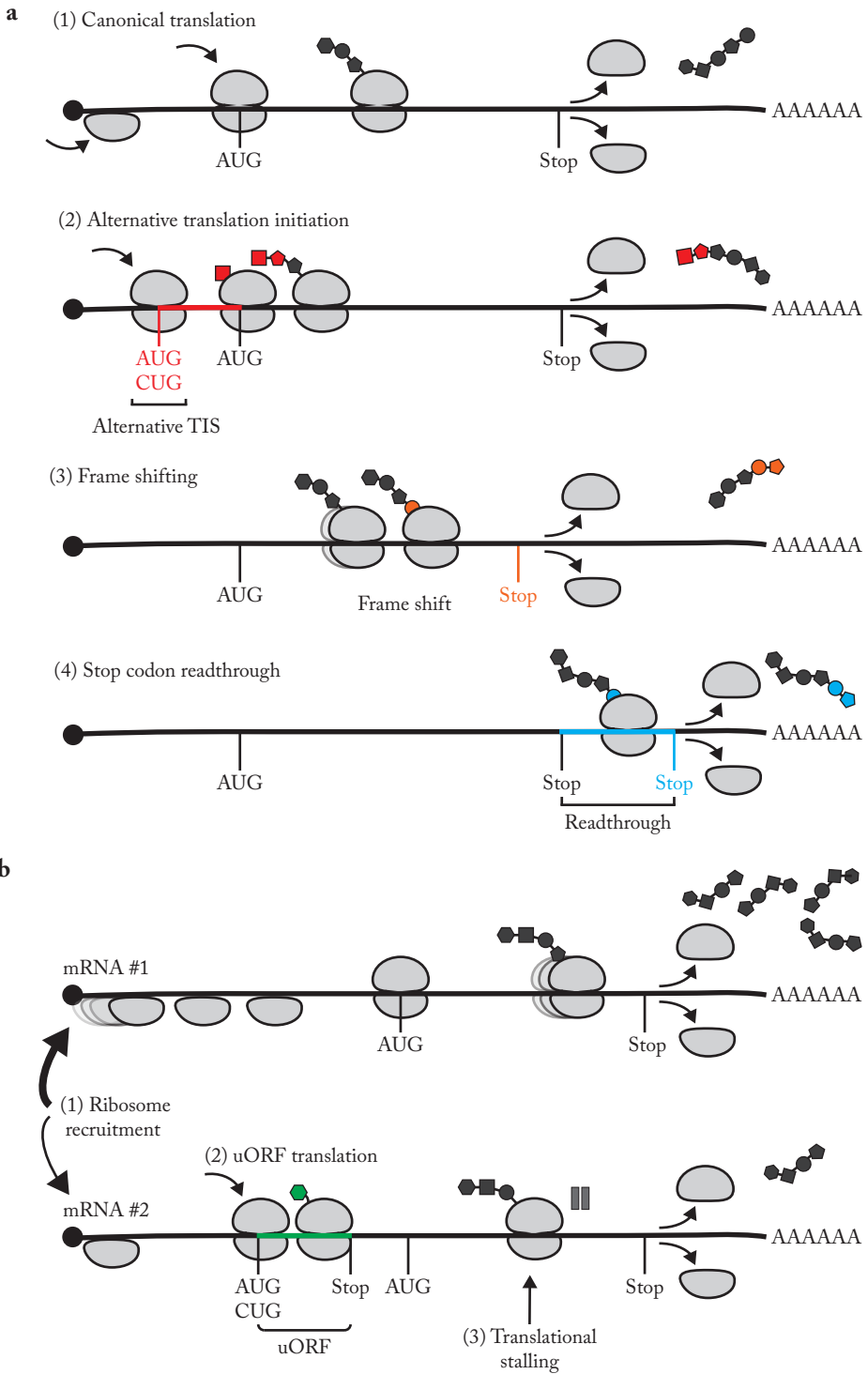
Two types of intragenic translational heterogeneity can be distinguished: heterogeneity in the amino acid sequence of newly synthesized proteins (Fig. 2a) and heterogeneity in the protein synthesis rate (Fig. 2b). In this section, we will discuss the experimental evidence and function of both types of translational heterogeneity.

### 2.4.1 Heterogeneity in protein sequence

Heterogeneity in protein sequence can be caused by alternative translation start site selection, ribosome frameshifting during translation elongation or stop codon read-through (Fig. 2a). Here, we will focus on translation start site heterogeneity and stop codon read-through, as these processes are best described and likely represent the predominant mechanisms underlying translational heterogeneity in eukaryotic cells.

5'-to-3' directed scanning along the mRNA by the 43S pre-initiation complex (PIC) generally results in translation initiation at the first (i.e. most 5') AUG codon (Box 1 and Fig. 1b). Detection efficiency of AUG codons depends on the surrounding sequences, and AUG codons are most effectively recognized as a translation initiation site (TIS) when surrounded by the Kozak Consensus Sequence (GCCACCAUGG) [3]. However, translation initiation can also occur at non-AUG codons, such as GUG or CUG [4]. Moreover, a PIC can fail to initiate at an AUG codon when encountering it during scanning (a process termed leaky scanning). Several lines of evidence have demonstrated that heterogeneity in start site usage is indeed apparent in cells (Box 2 for further information on the type of methods to assess translational heterogeneity). Ribosomal profiling revealed that at least half of the human mRNA transcripts contain more than one TIS [4,5], and in a small number of cases proteins synthesized from an alternative TIS have been detected by mass spectrometry [6–10].

Alternative TISes can be employed by the cells to drive isoform specific expression of a protein. In-frame upstream and downstream initiation from alternative start sites leads to protein N-terminal extensions and N-terminal truncations, respectively, which may affect protein localization or function in the cell [9,11]. The N-terminal extended protein isoform of the c-myc gene, for example, that stems from an upstream alternative initiation event, has enhanced DNA-binding capacity that drives the expression of cell cycle repressors [11]. While the example of c-myc clearly demonstrates the possibility of precise regulation of gene expression through alternative TIS usage, it is unclear whether the widespread occurrence of alternative TIS usage contributes to functional proteome diversification. Alternative TIS usage could also act as part of a regulatory mechanism to deflect ribosomes away from the main open reading frame (ORF). Translation from alternative TISes may result



in ‘junk’ polypeptides that are produced as a result of infidelity of the translation initiation machinery. This is especially the case for alternative TISEs that are out-of-frame with respect to the main protein coding ORF, which generally result in the synthesis of short peptides with an amino acid sequence unrelated to the main protein coding ORF. Additional work is required to catalog the expression of such alternative protein products and identify functions for these (poly)peptides. In this context, it is interesting to note that prokaryotes use the Shine-Dalgarno sequence, which is located directly upstream of the main TIS, to guide ribosomes directly to the correct site of initiation, rather than using a scanning mechanism for TIS identification (reviewed in [12]). It is possible that TIS identification by scanning provides more flexibility for alternative TIS selection, and, if so, would have evolved only if alternative TIS selection is functionally important (and thus not only ‘noise’).

A second process that can cause protein sequence heterogeneity is translation stop codon read-through (RT). Stop codon recognition is generally efficient and results in translation termination. However, in rare cases the stop codon can be decoded as a sense codon, resulting in a C-terminally extended protein. Estimates of stop codon RT vary from 0.01% to 0.1% in mammalian cells for most genes (with the UGA stop codon being the most prone to induce RT [13]), although in some examples RT frequency can exceed 30% of translation termination events [14]. Recent studies in both mammalian cells and bacteria used live-cell imaging to directly visualize translation of the 3' UTR. Surprisingly, both studies found that 3' UTR translation was highly heterogeneous between [15] and even within individual cells [16]. Furthermore, a genome-wide study revealed extensive changes in RT for many mRNAs during drosophila development [17]. While there is limited understanding of mechanisms controlling RT, one recent paper on the RT of the AGO1 mRNA stop codon revealed that RT is enhanced by miRNA binding downstream of the stop codon [18]. It will be interesting to learn whether this or a similar mechanism also acts on other genes. Together, these studies show that RT is highly heterogeneous and is likely a regulated process.

**Fig. 2 (opposite page) | Intragenic translational heterogeneity.**

Intragenic translational heterogeneity can be divided into heterogeneity in protein sequence (a) and heterogeneity in protein synthesis rate (b). **a**, Different mRNA molecules are shown that produce proteins with different amino acid sequences due to distinct mechanisms: (1) canonical translation; (2) alternative translation initiation; (3) ribosome frame shifting; (4) stop codon readthrough. **b**, Two mRNAs are depicted that have distinct synthesis rates of the protein encoded in the canonical ORF, resulting in translational heterogeneity between the two mRNA molecules (high and low synthesis rates for top and bottom mRNAs respectively). Protein synthesis rates of the canonical ORF are affected by: (1) 43S ribosome recruitment rates; (2) the presence of uORFs; (3) ribosomal stalling in the transcript.

**Box 2 | Methods for measuring translational heterogeneity.**

**Mass Spectrometry.** Mass spectrometry allows direct detection of distinct protein isoforms, for example those arising from alternative translation initiation sites, stop codon read-through or through translation of alternatively spliced isoforms. Identification of distinct protein isoforms using mass spectrometry is limited by the detection sensitivity of the instrument, which can be challenging for the detection of rare and/or unstable translation products. To enhance the detection of alternative translation products, multiple enrichment strategies have been developed, such as enrichment of protein N-termini for identification of alternative translation initiation sites [96,97]. While distinct translation products can be identified by mass spectrometry, single cell analysis has not yet been achieved, making it challenging to study intragenic translational heterogeneity in single cells using these methods.

**Ribosome Profiling.** Ribosome profiling is a sequencing-based method used to identify the precise position of ribosomes along mRNA by sequencing the ribosome-protected mRNA fragments [19]. Ribosome profiling is well suited to study intergenic translational heterogeneity in a genome-wide fashion, as it enables the identification of translation initiation sites, translation of the non-coding regions of the transcript and measurements of translation efficiency. Since ribosome profiling is an ensemble method, it requires averaging of thousands of mRNAs and cells, so measurements of translation of different mRNA molecules in single cells are challenging.

**Single-molecule imaging methods.** Recently, several methods have been developed that enable visualization of translation of single mRNA molecules [21–26,98]. Most of these methods rely on the introduction of multiple copies of a short peptide sequence (e.g. the SunTag [99]) into the coding sequence of a gene of interest. Upon translation, the short epitope sequence is synthesized and bound by a fluorescently-labeled antibody that is stably expressed in the cell, resulting in a bright fluorescence signal that reports on the translation efficiency of individual mRNA molecules [100]. Through direct visualization of translation, heterogeneity has been observed in many aspects of mRNA translation and decay [16,21–26,101–104]. Furthermore, using multiple tags (e.g. SunTag and MoonTag or Frankenbody [16,104]) allows quantification of translation of multiple open reading frames for single mRNA molecules. Therefore, single-molecule imaging methods are uniquely suited to study intragenic translational heterogeneity in single cells. A drawback of these methods is the low throughput (only one or a couple of genes can be analyzed at a time) and the need to introduce bulky tags into the gene of interest, which has so far mostly limited analysis to reporter genes, although endogenously tagged genes are becoming available [23].

### 2.4.2 Heterogeneity in protein synthesis rate

A second type of translational heterogeneity is heterogeneity in the translation rate (i.e. the number of proteins synthesized from a single mRNA molecule per unit of time) (Fig. 2b). Transcriptome-wide methods, such as ribosome profiling [19], are widely used to study the translation rate of individual genes and can readily uncover cell-to-cell and intergenic heterogeneity in the translation rate (Box 2). Furthermore, a recent genome-wide study using polysome profiling demonstrated that different transcript isoforms can often have distinct translation rates [20], suggestive of intragenic translational heterogeneity. More recently, new single-molecule imaging approaches (Box 2) have revealed intragenic translation rate heterogeneity for individual mRNAs, as well as temporal fluctuations in the translation rate of single

mRNAs [21–26], further confirming the existence of intragenic translation rate heterogeneity.

Since translation initiation is generally the rate-limiting step for protein synthesis, heterogeneity in the translation rate likely originates predominantly at the initiation step. Differential recruitment rates of the PIC to the 5' cap can cause translation rate heterogeneity. In addition, the translation rate of the main ORF can be affected by usage of alternative TISEs. For example, (short) upstream ORFs (uORFs) often reduce translation of the main ORF by promoting translation initiation of the uORF followed by ribosome recycling after termination at the uORF stop codon, preventing initiation at the main ORF TIS [27]. Importantly, differential usage of uORFs between distinct mRNA molecules may result in intragenic translation rate heterogeneity, providing a direct link between translation start site heterogeneity and translation rate heterogeneity [16]. In addition to the translation initiation step, variability in the translation elongation rate (for example, due to ribosome pausing) may also contribute to intragenic translation rate heterogeneity, although this remains a largely unexplored question.

Intragenic translation rate heterogeneity can represent spatial control over the translation rate. For example, in neuronal cells certain transcripts are only translated locally in axons (see [28] for a review). In addition, translation rate heterogeneity could also be a consequence of cell-wide translation rate regulation; to achieve a cell-wide reduction in translation rate, either the translation rate of each mRNA molecule is reduced partially, or alternatively the translation of a sub-set of mRNAs is reduced severely while other mRNAs are translated normally. Experimental evidence suggests that the latter mechanism may indeed occur, at least under some conditions. For example, upon inhibition of mTOR signaling, the translational repressor 4E-BP binds to eIF4E associated with individual mRNAs [29], which likely results in complete inhibition of translation of those mRNAs, while mRNAs that are not bound by 4E-BP are likely unaffected (at least at short time-scales). In these examples intragenic translation rate heterogeneity has important functional consequences and is tightly regulated. However, it is also possible that intragenic translation rate heterogeneity is simply a consequence of the stochastic nature of translation initiation. For example, the translation rate may fluctuate as translation factors bind and release stochastically from single mRNAs. Dissecting the contribution of active regulation and stochastic events to translation rate heterogeneity is an important future goal.

## 2.5 Molecular mechanisms of translational heterogeneity

In the previous sections we discussed emerging evidence of intragenic translational

heterogeneity. In the following sections, we will provide an overview of different possible mechanisms that could cause translational heterogeneity (Fig. 3).

### 2.5.1 mRNA primary sequence

Many eukaryotic genes encode multiple mRNA isoforms that differ in their primary nucleotide sequence. A median of six transcript isoforms has been detected per human protein coding gene (GENCODE, Release 33) and different transcript isoforms can coexist within a single cell [30–33]. Primary mRNA sequence variations can originate from variable transcription start site (TSS) usage, alternative usage of polyadenylation sites or through alternative splicing. The possible contribution of these types of sequence variability to intragenic translational heterogeneity is discussed below.

TSS mapping using CAGE in over nine hundred human cell lines and tissues uncovered more than one million TSSes [34]. An alternative TSS creates an alternative 5' untranslated region (UTR), and therefore may affect the translation rate of the transcript isoform [35]. Indeed, several studies identified hundreds of genes for which transcripts with alternative 5' UTRs showed distinct translation rates [20,36]. How differences in 5' UTRs result in different translation rates is not completely understood. Differential inclusion of an upstream ORF (uORF) or alternative TISes in a subset of transcript isoforms could contribute to the observed differences in translation rate. However, additional mechanisms may also exist; one study combined a massively parallel reporter assay and advanced computing and identified multiple new sequence motifs that regulate translation initiation efficiency of 5' UTRs [37], although the mechanism by which these motifs affect translation remains to be explored. Alternative TSS usage can also affect protein sequence. For example, if the TSS is located downstream of the TIS, then the mRNA will generate a truncated protein. Similarly, an upstream TSS could introduce additional TISes in the 5' UTR that result in either short peptides or N-terminally extended proteins [36]. Thus, 5' UTR heterogeneity occurs for many, if not most, genes due to alternative TSS usage, and may be an important driver of intragenic translational heterogeneity.

Sequence analysis of exon-exon boundaries in human cells revealed that most, if not all, multi-exon genes undergo alternative splicing [38], and in many cases, differentially spliced transcripts appear to co-exist in individual cells [30]. Differential splicing in a single cell may be caused by imperfect recognition of the splice sites by the spliceosome or weak/partial regulation of alternative splicing, such that a subset of mRNA molecules escapes the regulation. Most events of alternative splicing occur in the coding sequence and thus mainly cause protein sequence heterogeneity, rather



than translation rate heterogeneity. Nonetheless, a subset of splicing events, especially those in the 5' UTR which impact the 5' UTR sequence, can also cause translation rate heterogeneity [20].

During transcription, the 3' end of mRNAs is processed through cleavage of the pre-mRNA followed by addition of a polyA tail. Pre-mRNA cleavage occurs downstream of a cleavage and polyadenylation sequence (e.g. AAUAAA) and is regulated by a number of RNA binding proteins [39]. Interestingly, over half of the human and mouse genes express multiple transcript isoforms that differ in their 3' UTR sequence, due to cleavage and polyadenylation at alternative sites [40]. Even within single cells, multiple transcript isoforms are often detected that differ in the site of polyadenylation [32]. Why different pre-mRNA molecules originating from a single gene are processed differently is largely unclear. It is possible that differentially polyadenylated mRNAs were produced at different times in the cell's life cycle, when distinct regulatory processes were active. Alternatively, determination of the polyadenylation site may be stochastic, with different probabilities of usage for different polyadenylation sites. Alternative usage of polyadenylation sites generally alters the sequence of the 3' UTR of mRNAs. Since many RBPs and microRNAs, both of which can affect mRNA fate, bind to the 3' UTR, an alternative 3' UTR sequence may lead to an altered translation rate. Moreover, interactions between RBPs and microtubule motor proteins can control localization of an mRNA in the cell [40], and a recent study found that hundreds of transcript isoform pairs with different polyadenylation site usage showed distinct localizations in the brain [41]. Since localization of a transcript can affect its translation (especially in neurons), alternative localization of specific mRNA isoforms may also affect their translation.

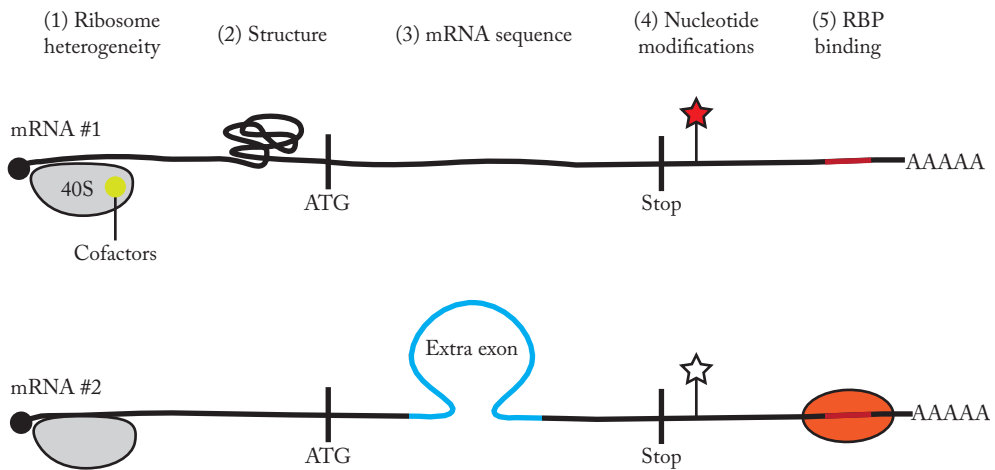
In summary, most genes express multiple transcript isoforms with distinct primary sequences, often within a single cell. Distinct transcript isoforms frequently show different translation rates due to inclusion or exclusion of specific regulatory sequences or alternative translation start sites, and, when the ORF sequence is affected, protein sequence is likely altered as well.

### 2.5.2 mRNA modifications

Information in mRNA is not only stored in the primary sequence, but also through post-transcriptional modifications of RNA nucleotides. Eleven distinct nucleotide modifications have been described for mRNAs, which collectively have been termed the 'epitranscriptome' [42]. The most prevalent modification is methylation of the adenosine base at the nitrogen-6 position ( $m^6A$ ), with an average occurrence of 1-3 modified adenosines per mRNA [42]. While other modifications are involved in

translational regulation as well, we will focus our discussion on the m<sup>6</sup>A modification, because it is the most abundant and best characterized modification on mRNAs. A plethora of studies has shown that m<sup>6</sup>A modifications affect mRNA translation. First, m<sup>6</sup>A modifications can increase the translation initiation rate by directly recruiting eIF3, METTL3 or YTHDF1 to the mRNA [43–46]. Second, it has been reported that m<sup>6</sup>A present in the coding sequence inhibits translation [47–49], possibly by slowing down translation elongation [50]. Thus, m<sup>6</sup>A plays an important regulatory role during translation.

Several studies have performed transcriptome-wide mapping of m<sup>6</sup>A sites to uncover the positions of m<sup>6</sup>A modifications in all mRNAs [51–56]. Interestingly, detailed quantitative analysis of individual m<sup>6</sup>A sites revealed that for most individual m<sup>6</sup>A sites only a subset of mRNA molecules contain the modification [57]. This ‘m<sup>6</sup>A stoichiometry’ (i.e. the fraction of mRNA molecules that contains an m<sup>6</sup>A modification at a specific site) can range from 7–77% depending on the modification site and cell line [57]. Recently, a new technique, MAZTER-seq [58], has further expanded the analysis of m<sup>6</sup>A stoichiometry to allow measurement of m<sup>6</sup>A stoichiometry for a substantial fraction of all m<sup>6</sup>A sites in the transcriptome. This study confirmed that in most cases only a subset of mRNA molecules is modified at any given site (‘intra-site heterogeneity’). In addition, a substantial variability exists in



**Fig. 3 | Molecular mechanisms of translational heterogeneity.**

Overview of potential origins of intragenic translational heterogeneity. Two mRNA molecules are shown that differ in: (1) composition of the translating ribosome, (2) mRNA structure, (3) mRNA sequence, (4) nucleotide modifications, and (5) RBPs bound to the mRNA molecule. Together these different features drive translational heterogeneity.

m<sup>6</sup>A stoichiometry for different sites ('inter-site heterogeneity') [58]. Inter-site m<sup>6</sup>A heterogeneity (i.e. different stoichiometries for different m<sup>6</sup>A sites) depends on local sequence context and RNA structure, suggesting that the specific stoichiometry of an m<sup>6</sup>A site is, at least in part, 'hard coded' in the mRNA sequence [58]. In contrast, intra-site m<sup>6</sup>A heterogeneity is likely caused by variability in m<sup>6</sup>A deposition and/or removal. For example, some mRNAs show decreased m<sup>6</sup>A methylation levels after transport to axons, where these mRNA molecules are locally translated [49]. Taken together, these studies reveal that m<sup>6</sup>A modifications are highly heterogeneous and are a potential source of intragenic translational heterogeneity.

### 2.5.3 mRNA structure

RNA can form intricate higher order structures, adding another layer of information to RNA molecules. RNA structure is formed by Watson-Crick base-pairing and can be further stabilized through other RNA interactions, such as sugar-backbone interactions. Structure in mRNAs can affect translation in multiple ways. First, mRNA structures can inhibit translation initiation by physically blocking the PIC during scanning [59]. Second, mRNA structure can also promote translation initiation, for example through recruitment of the translation initiation factor eIF3 to specific RNA hairpin structures [60]. Moreover, mRNA structure can stimulate translation initiation from non-canonical start codons through stalling of the PIC during scanning [61,62], or induce non-canonical translation initiation through direct recruitment of ribosomes to internal ribosome entry sites (IRESes) [63]. Finally, RNA structures in the coding sequence can stall ribosomes during translation elongation [64,65]. Collectively, these examples illustrate that mRNA structures have a profound impact on translation.

In recent years, many methods have been developed that use chemical probing combined with deep sequencing to investigate the structure of mRNA molecules *in vivo* (see [66] for an comprehensive review). While such methods have provided a wealth of new information on mRNA structures *in vivo*, these methods often provide only an ensemble mRNA structure. To understand if mRNA structural heterogeneity contributes to translational heterogeneity, methods are required that resolve all the different unique (full length) mRNA structures, preferably even in a single cell. Although challenging, new computational methods, analogous to methods used for NMR data analysis, have been developed to determine individual mRNA structures in chemical probing data sets [67–69]. Through these approaches, a recent study suggested that ACTB mRNA adopts multiple structural conformations *in vivo*, affecting the accessibility of the binding site of a protein [70]. Thus, mRNA molecules likely adopt different mRNA structures *in vivo*, which may be an important

contributor to translational heterogeneity.

RNA structure is likely highly dynamic, such that a single mRNA molecule can adopt multiple different conformations over time *in vivo*, which may further contribute to structural heterogeneity. Several studies have shown that mRNA molecules can adopt different conformations during their lifetime, for example, as mRNAs translocate from the nucleus to the cytoplasm, or during their translation by ribosomes [71–74]. Moreover, our recent work using single-molecule imaging uncovered substantial structural dynamics at short timescales (seconds to minutes) as well [75]. Such dynamic changes in mRNA structure can affect binding site accessibility of regulatory proteins with the mRNA, possibly affecting translation. Similarly, an *in vitro* study found that bacterial mRNA molecules constantly re-fold into different conformations, affecting the accessibility of the Shine-Dalgarno (SD) sequence over time, which, *in vivo*, may result in temporal translational heterogeneity [76]. Thus, to understand the contribution of mRNA structure on translational heterogeneity, it is necessary to assess both the suite of mRNA structures that are adopted by different mRNA molecules, and the dynamics of these mRNA structures.

#### **2.5.4 RNA binding proteins (RBPs)**

RBPs regulate the fate of mRNAs in multiple ways, including regulation of mRNA stability, localization and translation. Most RBPs associate with the mRNA through one or multiple well-defined RNA-binding domains (RBDs) to form ribonucleoprotein complexes (RNPs). RBDs often bind a relatively short sequence motif of around 2 to 8 nucleotides. In a number of cases the binding affinities of an RBP and RNA have been measured *in vitro*, which revealed dissociation constants ranging from a low nanomolar (strong binder) to micromolar [77–80]. Affinities in the nanomolar to micromolar range will generally result in interaction half-lives in the (mili)seconds to minutes range [78,81,82]. Since a typical mRNA molecule has a half-life of several hours [1], the duration of these interaction half-lives suggest that the protein composition of every mRNP is constantly changing, resulting in a high intragenic mRNP heterogeneity. The degree of mRNP heterogeneity depends on RBP concentration as well, as at high RBP concentrations binding sites may become saturated. Additional intragenic heterogeneity in mRNP composition is expected as a result of competition between RBPs that bind the same, or overlapping, binding sites in an mRNA. While experimental evidence for intragenic heterogeneity in mRNP composition is limited due to technical limitations in detecting interactions between a single mRNA and protein molecules, the existence of mRNP compositional heterogeneity can be deduced from RBP-mRNA interaction durations and mRNA half-life, and may have a major effect on translational heterogeneity.

While heterogeneity in mRNP composition likely exists for all mRNAs due to stochastic binding and unbinding of RBPs to individual mRNAs, mRNA compositional heterogeneity can also be due to active regulation. For example, during early development caudal mRNA is translationally active only at the posterior side of the embryo due to the asymmetric distribution of the translational repressor bicoid [83]. Even in non-polarized somatic cells spatial distribution of mRNAs can cause translational heterogeneity; for example, a recent study identified a new membraneless organelle, called a TIS granule (assemblies of the protein TIS11B), which resides adjacent to the ER [84]. The localization of mRNAs to TIS granules depends on RBPs, including HuR, that selectively bind specific mRNA transcripts [85]. Thus, translational heterogeneity can arise through heterogeneous localization of the mRNA transcripts, which can be mediated by RBPs.

### 2.5.5 Ribosome heterogeneity

The protein composition of the ribosome is generally considered invariant. However, several studies have suggested that some degree of heterogeneity may exist in the composition of the ribosome, and that distinct ribosomes may be capable of performing unique functions (reviewed in [86,87]).

Heterogeneity in ribosome composition is supported by gene expression analysis of ribosomal genes across various tissues and cell lines [88,89], which revealed that up to 25% of the ribosomal genes are differentially expressed [89]. Moreover, several paralogs of ribosomal proteins (RPs) are exclusively expressed in one tissue or cell type [88]. Given the structural similarity between RPs and their paralogs, RP paralogs may substitute canonical RPs during ribosome assembly [90,91]. Exclusion of canonical RPs or incorporation of an RP paralog may result in an 'alternative ribosome', potentially capable of performing specific functions (e.g. translating a subset of mRNA molecules differently). Indeed, four RPs were recently identified to be substoichiometric in mouse embryonic stem cells, suggesting that even within single cells some ribosomes contain these RPs, while others do not. Ribosome profiling using ribosomes containing these RPs identified hundreds of mRNAs that are preferentially translated by ribosomes containing these RPs [92]. In another example, haploinsufficiency of RPL38 causes severe developmental defects, originating from impaired IRES-mediated translation of a specific set of HOX genes [92,93]. While alternative ribosomes may preferentially translate specific transcripts, some genes may also be more sensitive than others to the availability of ribosomes (i.e. when the ribosome number is reduced), resulting in gene specific changes in translation efficiency [94]. Ribosomes were also found to associate with hundreds of additional proteins [95], providing another source of potential heterogeneity in ribosome

function. One of these ribosome-associated proteins, PKM, was identified as a factor specifically involved in the translation of ER-associated mRNAs, suggesting that the compositional heterogeneity of ribosome-associated proteins can also contribute to spatial heterogeneity in translation [95].

## 2.6 Concluding Remarks

The main function of translation is to decode the genetic information that is stored in the DNA. In this light, translation might be expected to be highly homogeneous, as the information stored in mRNA should be faithfully translated to protein. However, recent experimental evidence has demonstrated that mRNA molecules originating from the same gene can produce different amounts of protein and can synthesize entirely different polypeptides, suggesting that translation is, in fact, highly heterogeneous. Development of new tools to visualize mRNA translation of individual mRNA molecules in space and time provides exciting new opportunities to study translational heterogeneity and will hopefully help to address the many questions that remain unanswered (see 2.7 Outstanding Questions).

To what extent translational heterogeneity is functionally important remains a central question. Translational heterogeneity may be a consequence of temporal or spatial regulation of translation. Additionally, translational heterogeneity may contribute to functional proteome diversification [9]. On the other hand, translational heterogeneity could be a side-effect of the stochastic nature of the processes underlying mRNA translation. While translational control and functional proteome diversification are important for cellular function, stochasticity in translation may result in 'unwanted' heterogeneity that can result in the production of aberrant or even toxic proteins. We speculate that the flexibility required for complex translational regulation, for example the synthesis of multiple functional protein isoforms from a single mRNA, may have inadvertently resulted in unwanted translational heterogeneity as well. For example, regulation of the translation start site requires 'flexible' rules for translation initiation, which may also induce stochastic translation initiation at non-functional sites and thus production of aberrant proteins. It will be of great interest to further study how maximal regulatability is achieved, while unwanted heterogeneity is minimized.

As discussed in this review, many different mechanisms can underlie intragenic translational heterogeneity and it is currently unclear what the contribution of individual processes (e.g. RNA structural dynamics, m6a stoichiometry) is to the overall levels of heterogeneity. Moreover, it is unclear whether the heterogeneity caused by each of these processes individually mostly results in functional heterogeneity or mostly represents unwanted noise. Finally, a direct, causal link between the observed

intragenic translational heterogeneity and its potential molecular origins requires new methods that can quantify translation of single mRNA molecules and simultaneously measure different features (e.g. m<sup>6</sup>A modification, primary sequence, etc.) of the same mRNA molecule. A better understanding of the origins and potential functions of translational heterogeneity will provide a full picture of mechanisms underlying the decoding process of genetic information, a central process in life.

## 2.7 Outstanding Questions

- While different mechanisms can contribute to translational heterogeneity, the relative contribution of each mechanism to overall translational heterogeneity remains to be determined.
- How widespread is intragenic translational heterogeneity on endogenous genes? Several studies have found extensive translational heterogeneity on reporter genes. However, the extent of translational heterogeneity on native mRNAs has not been explored. It will be interesting to determine whether different mRNA species show distinct levels of intragenic translational heterogeneity.
- mRNA molecules can have a different mRNA sequence, RBP composition, nucleotide modification status, or mRNA structure. Does the large number of heterogeneous parameters mean that (almost) every mRNA molecule is unique? Or do sub-groups of mRNA molecules exist due to co-regulation of different parameters?
- What are the time scales on which translation of individual mRNA molecules change? The answer to this question likely varies for each type of translational heterogeneity.
- To what extent are heterogeneous mRNA features actively regulated? Regulation of mRNA features may result in spatiotemporal regulation of translation, while unregulated mRNA features may contribute to translation 'noise'.

## 2.8 Acknowledgements

We would like to thank members of the Tanenbaum lab for helpful discussions. This work was financially supported by an ERC starting grant (ERC-STG 677936-RNAREG), and by the Howard Hughes Medical Institute through an International Research Scholar grant to MET (HHMI/IRS 55008747). MET was also financially supported by Oncode Institute.

## 2.9 References

- 1 Schwanhäusser, B. et al. (2011) Global quantification of mammalian gene expression control. *Nature* 473, 337–342
- 2 Buszczak, M. et al. (2014) Cellular differences in protein synthesis regulate tissue homeostasis. *Cell* 159, 242–251
- 3 Kozak, M. (1986) Point mutations define a sequence flanking the AUG initiator codon that modulates translation by eukaryotic ribosomes. *Cell* 44, 283–292
- 4 Ingolia, N.T. et al. (2011) Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. *Cell* 147, 789–802
- 5 Lee, S. et al. (2012) Global mapping of translation initiation sites in mammalian cells at single-nucleotide resolution. *Proc. Natl. Acad. Sci. U. S. A.* 109, E2424–E2432
- 6 Na, C.H. et al. (2018) Discovery of noncanonical translation initiation sites through mass spectrometric analysis of protein N termini. *Genome Res.* 28, 25–36
- 7 Van Damme, P. et al. (2014) N-terminal proteomics and ribosome profiling provide a comprehensive view of the alternative translation initiation landscape in mice and men. *Mol. Cell. Proteomics* 13, 1245–1261
- 8 Yeom, J. et al. (2017) Comprehensive analysis of human protein N-termini enables assessment of various protein forms. *Sci. Rep.* 7, 6599
- 9 Chen, J. et al. (2020) Pervasive functional translation of noncanonical human open reading frames. *Science* 367, 1140–1146
- 10 Martinez, T.F. et al. (2020) Accurate annotation of human protein-coding small open reading frames. *Nat. Chem. Biol.* 16, 458–468
- 11 Hann, S.R. et al. (1992) Translational activation of the non-AUG-initiated c-myc 1 protein at high cell densities due to methionine deprivation. *Genes Dev.* 6, 1229–1240
- 12 Rodnina, M. V. (2018) Translation in prokaryotes. *Cold Spring Harb. Perspect. Biol.* 10, a032664
- 13 Dabrowski, M. et al. (2015) Translational readthrough potential of natural termination codons in eucaryotes – The impact of RNA sequence. *RNA Biol.* 12, 950–958
- 14 Loughran, G. et al. (2014) Evidence of efficient stop codon readthrough in four mammalian genes. *Nucleic Acids Res.* 42, 8928–8938
- 15 Fan, Y. et al. (2017) Heterogeneity of stop codon readthrough in single bacterial cells and implications for population fitness. *Mol. Cell* 67, 826–836
- 16 Boersma, S. et al. (2019) Multi-color single-molecule imaging uncovers extensive heterogeneity in mRNA decoding. *Cell* 178, 458–472
- 17 Dunn, J.G. et al. (2013) Ribosome profiling reveals pervasive and regulated stop codon readthrough in *Drosophila melanogaster*. *Elife* 2, e01179
- 18 Singh, A. et al. (2019) Let-7a-regulated translational readthrough of mammalian AGO1 generates a micro RNA pathway inhibitor. *EMBO J.* 38, e100727
- 19 Ingolia, N.T. (2014) Ribosome profiling: new views of translation, from single codons to genome scale. *Nat. Rev. Genet.* 15, 205–213
- 20 Floor, S.N. and Doudna, J.A. (2016) Tunable protein synthesis by transcript isoforms in human cells. *Elife* 5, e10921
- 21 Yan, X. et al. (2016) Dynamics of translation of single mRNA molecules in vivo. *Cell* 165, 976–989
- 22 Wu, B. et al. (2016) Translation dynamics of single mRNAs in live cells and neurons. *Science* 352, 337–342
- 23 Pichon, X. et al. (2016) Visualization of single polysomes reveals translation dynamics in living human cells. *J. Cell Biol.* 214, 769–781
- 24 Morisaki, T. et al. (2016) Real-time



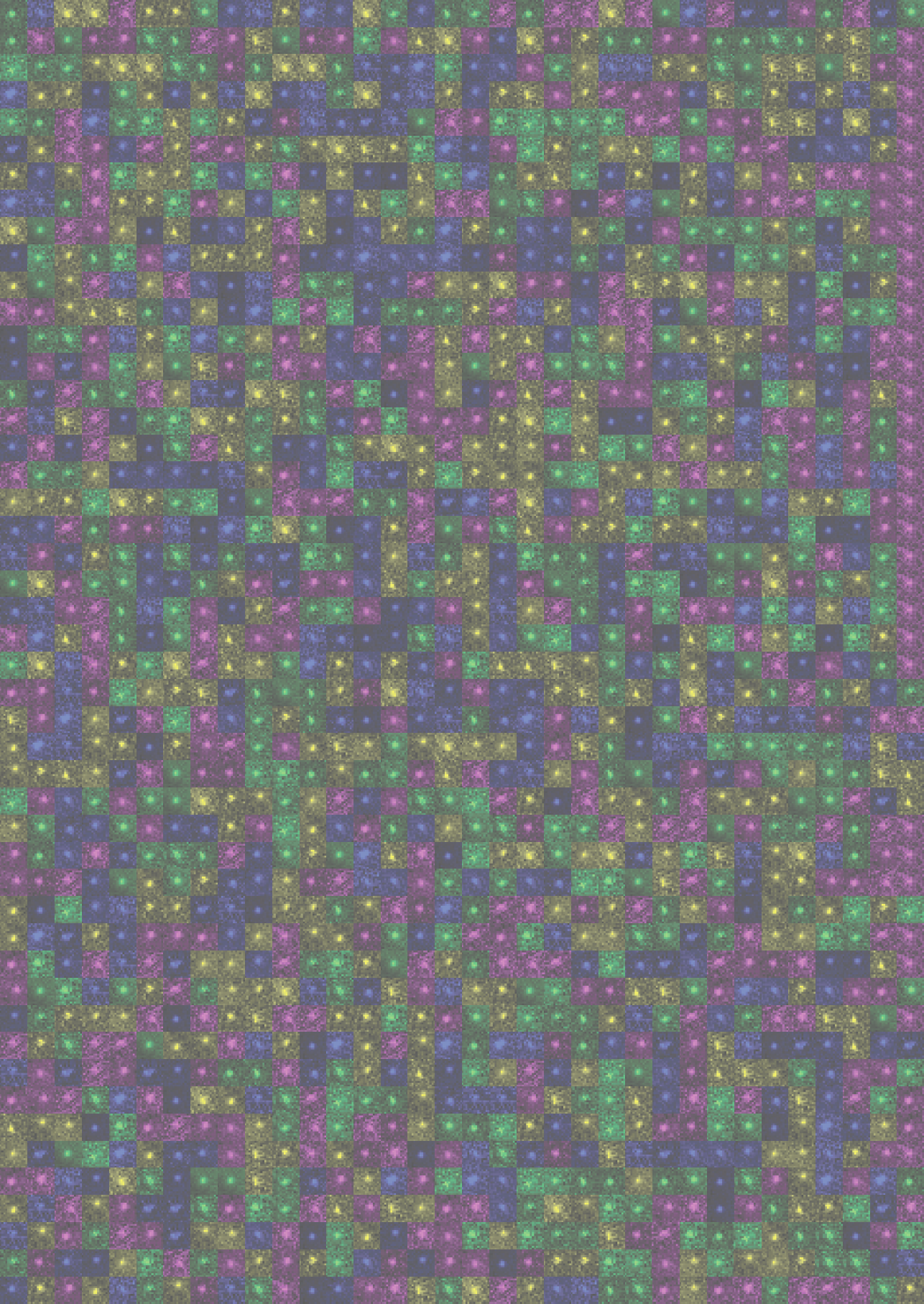
- quantification of single RNA translation dynamics in living cells. *Science* 352, 1425–1429
- 25 Wang, C. et al. (2016) Real-time imaging of translation on single mRNA transcripts in live cells. *Cell* 165, 990–1001
- 26 Halstead, J.M. et al. (2015) An RNA biosensor for imaging the first round of translation from single cells to living animals. *Science* 347, 1367–1370
- 27 Johnstone, T.G. et al. (2016) Upstream ORFs are prevalent translational repressors in vertebrates. *EMBO J.* 35, 706–723
- 28 Glock, C. et al. (2017) mRNA transport & local translation in neurons. *Curr. Opin. Neurobiol.* 45, 169–177
- 29 Peter, D. et al. (2015) Molecular architecture of 4E-BP translational inhibitors bound to eIF4E. *Mol. Cell* 57, 1074–1087
- 30 Song, Y. et al. (2017) Single-cell alternative splicing analysis with expedition reveals splicing dynamics during neuron differentiation. *Mol. Cell* 67, 148–161
- 31 Shalek, A.K. et al. (2013) Single-cell transcriptomics reveals bimodality in expression and splicing in immune cells. *Nature* 498, 236–240
- 32 Velten, L. et al. (2015) Single-cell polyadenylation site mapping reveals 3' isoform choice variability. *Mol. Syst. Biol.* 11, 812
- 33 Waks, Z. et al. (2011) Cell-to-cell variability of alternative RNA splicing. *Mol. Syst. Biol.* 7, 506
- 34 Kanamori-Katayama, M. et al. (2011) Unamplified cap analysis of gene expression on a single-molecule sequencer. *Genome Res.* 21, 1150–1159
- 35 Hinnebusch, A.G. et al. (2016) Translational control by 5'-untranslated regions of eukaryotic mRNAs. *Science* 352, 1413–1416
- 36 Wang, X. et al. (2016) Pervasive isoform-specific translational regulation via alternative transcription start sites in mammals. *Mol. Syst. Biol.* 12, 875
- 37 Sample, P.J. et al. (2019) Human 5' UTR design and variant effect prediction from a massively parallel translation assay. *Nat. Biotechnol.* 37, 803–809
- 38 Wang, E.T. et al. (2008) Alternative isoform regulation in human tissue transcriptomes. *Nature* 456, 470–476
- 39 Tian, B. and Manley, J.L. (2017) Alternative polyadenylation of mRNA precursors. *Nat. Rev. Mol. Cell Biol.* 18, 18–30
- 40 Mayr, C. (2016) Evolution and biological roles of alternative 3'UTRs. *Trends Cell Biol.* 26, 227–237
- 41 Tushev, G. et al. (2018) Alternative 3' UTRs modify the localization, regulatory potential, stability, and plasticity of mRNAs in neuronal compartments. *Neuron* 98, 495–511
- 42 Zaccara, S. et al. (2019) Reading, writing and erasing mRNA methylation. *Nat. Rev. Mol. Cell Biol.* 20, 608–624
- 43 Meyer, K.D. et al. (2015) 5' UTR m6A promotes cap-independent translation. *Cell* 163, 999–1010
- 44 Zhou, J. et al. (2015) Dynamic m6A mRNA methylation directs translational control of heat shock response. *Nature* 526, 591–594
- 45 Choe, J. et al. (2018) mRNA circularization by METTL3-eIF3h enhances translation and promotes oncogenesis. *Nature* 561, 556–560
- 46 Wang, X. et al. (2015) N6-methyladenosine modulates messenger RNA translation efficiency. *Cell* 161, 1388–1399
- 47 Slobodin, B. et al. (2017) Transcription impacts the efficiency of mRNA translation via co-transcriptional N6-adenosine methylation. *Cell* 169, 326–337
- 48 Qi, S.T. et al. (2016) N6-methyladenosine sequencing highlights the involvement of mRNA methylation in oocyte meiotic maturation and embryo development by regulating translation in *Xenopus laevis*. *J. Biol. Chem.* 291, 23020–23026
- 49 Yu, J. et al. (2018) Dynamic m6A modification regulates local translation of mRNA in axons. *Nucleic Acids Res.* 46, 1412–1423

- 50 Choi, J. et al. (2016) N6-methyladenosine in mRNA disrupts tRNA selection and translation-elongation dynamics. *Nat. Struct. Mol. Biol.* 23, 110–115
- 51 Meyer, K.D. et al. (2012) Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell* 149, 1635–1646
- 52 Dominissini, D. et al. (2012) Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature* 485, 201–206
- 53 Linder, B. et al. (2015) Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. *Nat. Methods* 12, 767–772
- 54 Ke, S. et al. (2015) A majority of m6A residues are in the last exons, allowing the potential for 3' UTR regulation. *Genes Dev.* 29, 2037–2053
- 55 Meyer, K.D. (2019) DART-seq: an antibody-free method for global m6A detection. *Nat. Methods* 16, 1275–1280
- 56 Zhang, Z. et al. (2019) Single-base mapping of m6A by an antibody-independent method. *Sci. Adv.* 5, eaax0250
- 57 Liu, N. et al. (2013) Probing N6-methyladenosine RNA modification status at single nucleotide resolution in mRNA and long noncoding RNA. *RNA* 19, 1848–1856
- 58 Garcia-Campos, M.A. et al. (2019) Deciphering the “m6A Code” via antibody-independent quantitative profiling. *Cell* 178, 731–747
- 59 Mortimer, S.A. et al. (2014) Insights into RNA structure and function from genome-wide studies. *Nat. Rev. Genet.* 15, 469–479
- 60 Lee, A.S.Y. et al. (2015) EIF3 targets cell-proliferation messenger RNAs for translational activation or repression. *Nature* 522, 111–114
- 61 Guenther, U.P. et al. (2018) The helicase Ded1p controls use of near-cognate translation initiation codons in 5' UTRs. *Nature* 559, 130–134
- 62 Kozak, M. (1990) Downstream secondary structure facilitates recognition of initiator codons by eukaryotic ribosomes. *Proc. Natl. Acad. Sci. U. S. A.* 87, 8301–8305
- 63 Weingarten-Gabbay, S. et al. (2016) Systematic discovery of cap-independent translation sequences in human and viral genomes. *Science* 351, aad4939
- 64 Jungfleisch, J. et al. (2017) A novel translational control mechanism involving RNA structures within coding sequences. *Genome Res.* 27, 95–106
- 65 Burkhardt, D.H. et al. (2017) Operon mRNAs are organized into ORF-centric structures that predict translation efficiency. *Elife* 6, e22037
- 66 Strobel, E.J. et al. (2018) High-throughput determination of RNA structures. *Nat. Rev. Genet.* 19, 615–634
- 67 Spasic, A. et al. (2018) Modeling RNA secondary structure folding ensembles using SHAPE mapping data. *Nucleic Acids Res.* 46, 314–323
- 68 Li, H. and Aviran, S. (2018) Statistical modeling of RNA structure profiling experiments enables parsimonious reconstruction of structure landscapes. *Nat. Commun.* 9, 606
- 69 Frank, A.T. et al. (2009) Constructing RNA dynamical ensembles by combining MD and motionally decoupled NMR RDCs: new insights into RNA dynamics and adaptive ligand recognition. *Nucleic Acids Res.* 37, 3670–3679
- 70 Woods, C.T. et al. (2017) Comparative visualization of the RNA suboptimal conformational ensemble in vivo. *Biophys. J.* 113, 290–301
- 71 Beaudoin, J.-D. et al. (2018) Analyses of mRNA structure dynamics identify embryonic gene regulatory programs. *Nat. Struct. Mol. Biol.* 25, 677–686
- 72 Adivarahan, S. et al. (2018) Spatial organization of single mRNPs at different stages of the gene expression pathway. *Mol. Cell* 72, 727–738
- 73 Mustoe, A.M. et al. (2018) Pervasive regulatory functions of mRNA structure revealed by high-resolution SHAPE probing. *Cell* 173,

- 181–195
- 74 Mizrahi, O. et al. (2018) Virus-induced changes in mRNA secondary structure uncover cis-regulatory elements that directly control gene expression. *Mol. Cell* 72, 862–874
- 75 Ruijtenberg, S. et al. (2020) mRNA structural dynamics shape Argonaute-target interactions. *Nat. Struct. Mol. Biol.* 27, 790–801
- 76 Rinaldi, A.J. et al. (2016) The Shine-Dalgarno sequence of riboswitch-regulated single mRNAs shows ligand-dependent accessibility bursts. *Nat. Commun.* 7, 8976
- 77 Jens, M. and Rajewsky, N. (2015) Competition between target sites of regulators shapes post-transcriptional gene regulation. *Nat. Rev. Genet.* 16, 113–126
- 78 Buenrostro, J.D. et al. (2014) Quantitative analysis of RNA-protein interactions on a massively parallel array reveals biophysical and evolutionary landscapes. *Nat. Biotechnol.* 32, 562–568
- 79 Webster, M.W. et al. (2019) RNA-binding proteins distinguish between similar sequence motifs to promote targeted deadenylation by Ccr4-Not. *Elife* 8, e40670
- 80 Benhalevy, D. et al. (2017) The human CCHC-type zinc finger nucleic acid-binding protein binds G-rich elements in target mRNA coding sequences and promotes translation. *Cell Rep.* 18, 2979–2990
- 81 D'Agostino, V.G. et al. (2013) A novel high throughput biochemical assay to evaluate the HuR protein-RNA complex formation. *PLoS One* 8, e72426
- 82 Nicastro, G. et al. (2017) Mechanism of  $\beta$ -actin mRNA recognition by ZBP1. *Cell Rep.* 18, 1187–1199
- 83 Niessing, D. et al. (2002) Bicoid associates with the 5'-cap-bound complex of caudal mRNA and represses translation. *Genes Dev.* 16, 2576–2582
- 84 Ma, W. and Mayr, C. (2018) A membraneless organelle associated with the endoplasmic reticulum enables 3'UTR-mediated protein-protein interactions. *Cell* 175, 1492–1506
- 85 Berkovits, B.D. and Mayr, C. (2015) Alternative 3' UTRs act as scaffolds to regulate membrane protein localization. *Nature* 522, 363–367
- 86 Gerst, J.E. (2018) Pimp my ribosome: ribosomal protein paralogs specify translational control. *Trends Genet.* 34, 832–845
- 87 Genuth, N.R. and Barna, M. (2018) The discovery of ribosome heterogeneity and its implications for gene regulation and organismal life. *Mol. Cell* 71, 364–374
- 88 Gupta, V. and Warner, J.R. (2014) Ribosome-omics of the human ribosome. *RNA* 20, 1004–1013
- 89 Guimaraes, J.C. and Zavolan, M. (2016) Patterns of ribosomal protein expression specify normal and malignant human cells. *Genome Biol.* 17, 236
- 90 Jiang, L. et al. (2017) RPL10L is required for male meiotic division by compensating for RPL10 during meiotic sex chromosome inactivation in mice. *Curr. Biol.* 27, 1498–1505
- 91 O'Leary, M.N. et al. (2013) The ribosomal protein Rpl22 controls ribosome composition by directly repressing expression of its own paralog, Rpl2211. *PLoS Genet.* 9, e1003708
- 92 Shi, Z. et al. (2017) Heterogeneous ribosomes preferentially translate distinct subpools of mRNAs genome-wide. *Mol. Cell* 67, 71–83
- 93 Xue, S. et al. (2015) RNA regulons in Hox 5' UTRs confer ribosome specificity to gene regulation. *Nature* 517, 33–38
- 94 Mills, E.W. and Green, R. (2017) Ribosomopathies: there's strength in numbers. *Science* 358, eaan2755
- 95 Simsek, D. et al. (2017) The mammalian ribo-interactome reveals ribosome functional diversity and heterogeneity. *Cell* 169, 1051–1065
- 96 Staes, A. et al. (2008) Improved recovery of proteome-informative, protein N-terminal peptides by combined fractional diagonal chromatography (COFRADIC). *Proteomics* 8, 1362–1370

- 97 Kleinfeld, O. et al. (2011) Identifying and quantifying proteolytic events and the natural N terminome by terminal amine isotopic labeling of substrates. *Nat. Protoc.* 6, 1578–1611
- 98 Na, Y. et al. (2016) Real-Time imaging reveals properties of glutamate-induced Arc/Arg 3.1 translation in neuronal dendrites. *Neuron* 91, 561–573
- 99 Tanenbaum, M.E. et al. (2014) A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell* 159, 635–646
- 100 Khuperkar, D. et al. (2020) Quantification of mRNA translation in live cells using single-molecule imaging. *Nat. Protoc.* 15, 1371–1398
- 101 Wilbertz, J.H. et al. (2019) Single-molecule imaging of mRNA localization and regulation during the integrated stress response. *Mol. Cell* 73, 946–958
- 102 Hoek, T.A. et al. (2019) Single-molecule imaging uncovers rules governing nonsense-mediated mRNA decay. *Mol. Cell* 75, 324–339
- 103 Moon, S.L. et al. (2019) Multicolour single-molecule tracking of mRNA interactions with RNP granules. *Nat. Cell Biol.* 21, 162–168
- 104 Lyon, K. et al. (2019) Live-cell single RNA imaging reveals bursts of translational frameshifting. *Mol. Cell* 75, 172–183





# 3

## mRNA structural dynamics shape Argonaute-target interactions

Suzan Ruijtenberg<sup>1,4</sup>, Stijn Sonneveld<sup>1,4</sup>, Tao Ju Cui<sup>2</sup>, Ive Logister<sup>1</sup>, Dion de Steenwinkel<sup>1</sup>, Yao Xiao<sup>3</sup>, Ian J. MacRae<sup>3</sup>, Chirlmin Joo<sup>2</sup>, Marvin E. Tanenbaum<sup>1\*</sup>

<sup>1</sup> Oncode Institute, Hubrecht Institute–KNAW and University Medical Center Utrecht, Utrecht, The Netherlands

<sup>2</sup> Kavli Institute of NanoScience, Department of BioNanoScience, Delft University of Technology, 2629 HZ Delft, The Netherlands

<sup>3</sup> Department of Integrative Structural and Computational Biology, The Scripps Research Institute, La Jolla, CA, USA

<sup>4</sup> These authors contributed equally to this work.

\* Corresponding author and lead contact. Email address for correspondence: [m.tanenbaum@hubrecht.eu](mailto:m.tanenbaum@hubrecht.eu).

Adapted from: Ruijtenberg, S., Sonneveld, S., Cui, T.J. et al. mRNA structural dynamics shape Argonaute-target interactions. *Nat Struct Mol Biol* 27, 790–801 (2020).  
<https://doi.org/10.1038/s41594-020-0461-1>

Published online 13 July 2020.

### 3.1 Abstract

Small interfering RNAs (siRNAs) promote RNA degradation in a variety of processes and have important clinical applications. siRNAs direct cleavage of target RNAs by guiding Argonaute2 (AGO2) to its target site. Target site accessibility is critical for AGO2-target interactions, but how target site accessibility is controlled *in vivo* is poorly understood. Here, we use live-cell single-molecule imaging in human cells to determine rate constants of the AGO2 cleavage cycle *in vivo*. We find that the rate-limiting step in mRNA cleavage frequently involves unmasking of target sites by translating ribosomes. Target site masking is caused by heterogeneous intramolecular RNA-RNA interactions, which can conceal target sites for many minutes in the absence of translation. Our results uncover how dynamic changes in mRNA structure shape AGO2-target recognition, provide estimates of mRNA (un)folding rates *in vivo*, and provide experimental evidence for the role of mRNA structural dynamics in control of mRNA-protein interactions.

### 3.2 Introduction

A family of small RNAs of 20–32 nucleotides (nt), including microRNAs (miRNAs), small interfering RNAs (siRNAs), and Piwi interacting RNAs (piRNAs), is an important class of molecules that regulate RNA and protein levels in cells [1–5]. Small RNAs guide Argonaute (AGO) proteins to target RNAs via Watson-Crick base pairing, resulting in either target cleavage or recruitment of additional effector proteins to induce other types of target repression [6–11].

*In vitro* single-molecule imaging, as well as biochemical and structural approaches [1,10,12–16] have shown that initial AGO-target interactions are established by nt 2–4 of the small RNA, and are subsequently extended to nt 2–8 (referred to as the seed region), which further stabilizes the interaction between AGO and the target site [12–15]. Further base pairing beyond the seed region is required for target cleavage by AGO2, the major human AGO family member with endonucleolytic cleavage activity in somatic cells [8,10,17–19]. However, AGO-target interaction dynamics *in vivo* are likely more complex. First, the cytoplasm contains many different RNA species, providing a far more complex environment for the target search process by AGO. Second, hundreds of RNA binding proteins (RBPs) exist *in vivo*, which may undergo kinetic competition with AGO for target site binding [20–23]. Third, *in vivo*, RNA targets are often translated by ribosomes, which may actively displace AGO proteins bound within the open reading frame (ORF) [1,24,25]. Finally, *in vivo* RNA targets are typically at least an order of magnitude longer than the RNA oligonucleotides that are frequently used as targets for *in vitro* studies. Long RNA targets have a far



greater potential to adopt one or more secondary and tertiary structures, and RNA structures inhibit target recognition by AGO [26–32]. Importantly, RNA structural dynamics can take place on timescales spanning several orders of magnitude (ranging from ms to hrs) depending on the type of structural rearrangement (e.g. opening of single nucleotide interactions or large scale tertiary rearrangements) [33–35]. It is, however, currently unclear which types of RNA dynamics are functionally relevant for processes like AGO–target interactions.

### 3.3 Results

#### 3.3.1 Single-molecule live-cell imaging of AGO2-dependent mRNA target silencing

To study AGO2 activity on single translated mRNA molecules in living cells, we adapted a microscopy-based live-cell imaging method that we and others recently developed to visualize translation of individual mRNA molecules [36–40] (Fig. 1a, left). We designed an siRNA with full complementarity to a site in the coding sequence (CDS) of a reporter mRNA (Fig. 1b, methods 3.5.2 and 3.5.4 for plasmid design and siRNA sequences, respectively). Analysis by northern blot, qPCR, and single-molecule FISH (smFISH) revealed that siRNA transfection induced a reduction in reporter mRNA levels, and the formation of 3' and 5' cleavage fragments (Fig. 1c and Supplemental Fig. 1a–f), indicating that the reporter mRNA was targeted for endonucleolytic cleavage. The 5' cleavage fragment comprises most of the CDS and is thus likely associated with the majority of ribosomes translating the SunTag epitope, which recruits GFP signal (Fig. 1a, right). The 3' mRNA cleavage fragment contains a small part of the CDS, as well as the PCP binding sites, which bind to and concentrate mCherry-labeled molecules embedded in the plasma membrane. Upon cleavage, the mCherry-labeled 3' cleavage fragment is thus expected to remain in the field-of-view (until it is degraded by an RNA exonuclease) while the GFP positive 5' fragment is expected to diffuse out of the field-of-view (where it is likely degraded through the non-stop decay pathway) (Fig. 1a, right). Thus, we reasoned that in live-cell imaging experiments, mRNA cleavage would result in a separation of GFP and mCherry foci.

Upon induction of transcription of the reporter mRNA in human U2OS cells, new mRNAs rapidly appeared in the field-of-view and initiated translation (Figure 1d). Strikingly, in siRNA-transfected cells, GFP and mCherry foci frequently separated within minutes of translation initiation (92% of mRNAs in 10 min) (Fig. 1d, bottom, Fig. 1e, methods 3.5.18). Separation of GFP and mCherry foci was due to AGO2-dependent endonucleolytic cleavage as foci separation was largely eliminated by mutation of the siRNA binding site or depletion of AGO2 (Fig. 1f and Supplemental

Fig. 1g). Furthermore, the fluorescence intensity of most (97%) GFP foci at the time of separation was greater than the intensity of a single SunTag polypeptide, indicative of endonucleolytic cleavage rather than translation termination (Supplemental Fig. 1h, methods 3.5.20). In contrast, transcription of the reporter mRNA (Supplemental

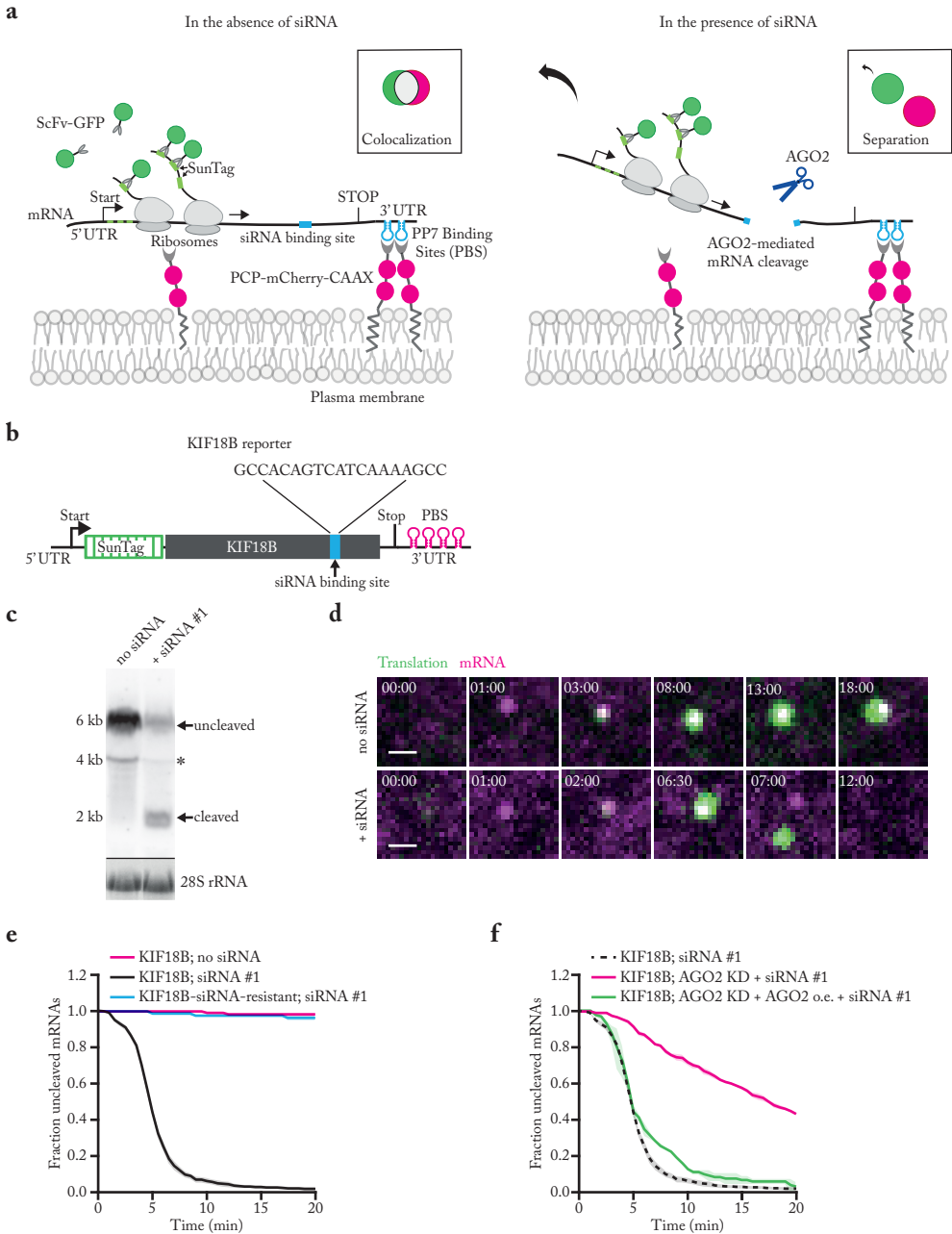


Fig. 1i) and nuclear mRNA levels (Supplemental Fig. 1d,f) were unaffected by AGO2-siRNA complex, consistent with previous studies [41]. Similarly, no significant effect on translation rates was observed (Supplemental Fig. 1j,k, and methods 3.5.26). Together, these results show that endonucleolytic cleavage in the cytoplasm is the predominant mechanism of action of AGO2-siRNA complexes.

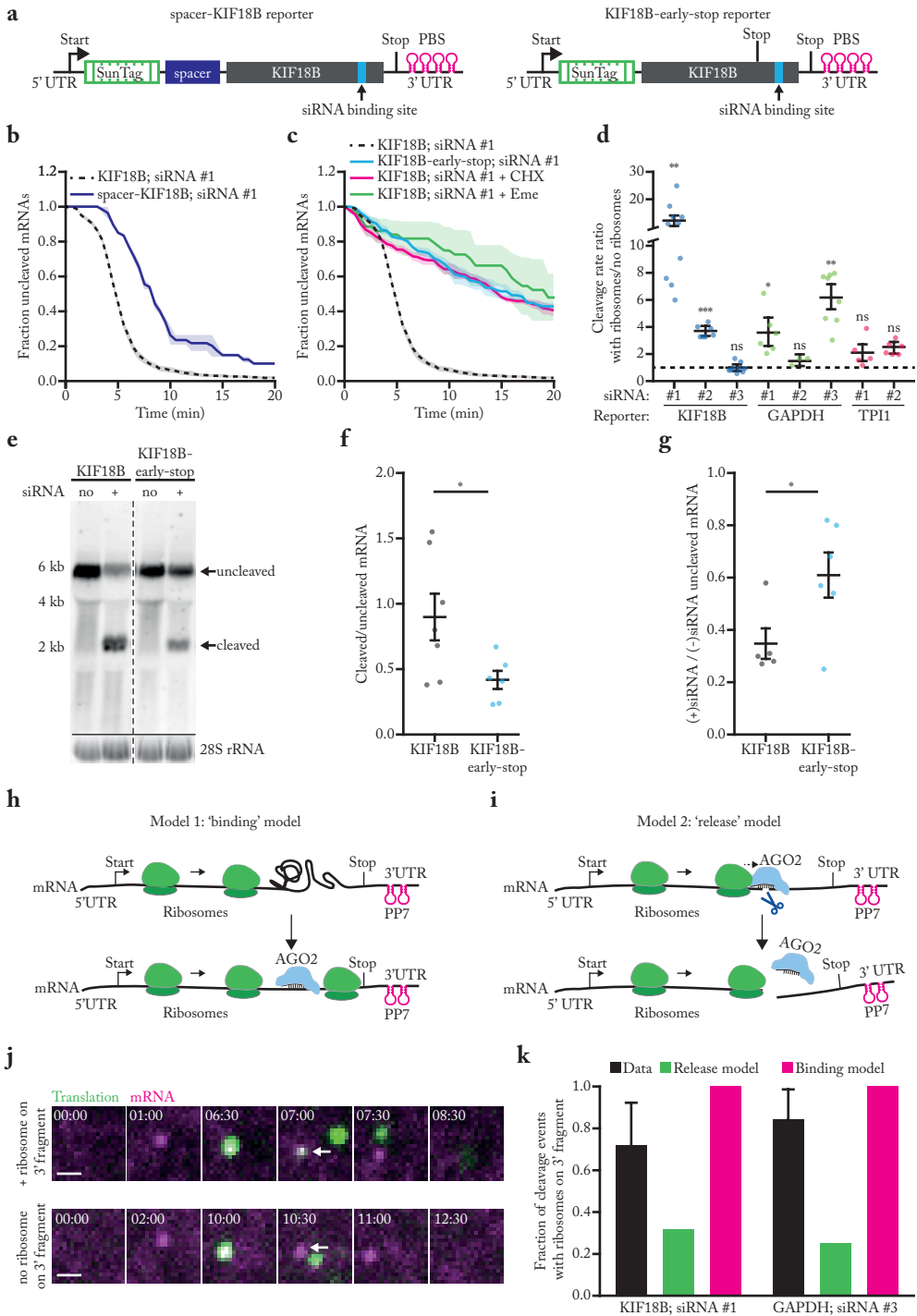
### 3.3.2 Ribosomes stimulate AGO2-dependent mRNA cleavage

Surprisingly, many mRNAs (56%) were cleaved between 4–6 minutes after the start of translation (i.e. after first appearance of GFP signal on an mRNA) (see Fig. 1e). Intriguingly, this time window represents the approximate time at which the first ribosome arrives at the AGO2 cleavage site [37] (see also Supplemental Fig. 2a, pink bars, and methods 3.5.25). Introduction of a spacer sequence between the SunTag and AGO2 binding site (Fig. 2a, left), substantially delayed cleavage relative to the start of translation (Fig. 2b), suggesting that ribosomes arriving at the AGO2 binding site stimulate AGO2-dependent mRNA cleavage. Furthermore, treatment of cells with the ribosome translocation inhibitors cycloheximide (CHX) or Emetine (Eme), or introduction of a stop codon upstream of the AGO2 binding site all strongly inhibited AGO2-dependent mRNA cleavage (Fig. 2a, right, Fig. 2c and Supplemental Fig. 2b). Together, these results show that arrival of translocating ribosomes at the siRNA binding site stimulates mRNA cleavage by AGO2.

Analysis of additional siRNAs and mRNAs revealed that other siRNAs (4/8) also showed a higher cleavage rate by AGO2 in the presence of translating ribosomes (Fig. 2d, Supplemental Fig. 2c-i, and methods 3.5.23). Ribosome-stimulated cleavage by

**Fig. 1 (opposite page) | Observing AGO2-dependent mRNA target silencing by single-molecule live-cell imaging.**

**a**, Schematic of the single-molecule imaging assay used to visualize AGO2-mediated mRNA silencing in the absence (left) or presence (right) of siRNA. Green and magenta spots (insets) show nascent polypeptides (translation) and mRNA, respectively, as observed by microscopy. **b**, Schematic of the mRNA reporter. **c**, Northern blot of cells expressing the reporter mRNA shown in (b) either without siRNA or transfected with KIF18B siRNA #1. (top) Upper band (uncleaved) represents full-length reporter mRNA, lower band (cleaved) represents the 3' cleavage fragment. Asterisk indicates an additional 4 kb band that may represent a shorter isoform of the reporter mRNA. (bottom) 28S rRNA acts as a loading control. **d**, Representative images of mRNA molecules of the reporter shown in (b) expressed in SunTag-PP7 cells without (top) or with siRNA (bottom). Scale bar, 1  $\mu$ m. Time is shown in min:sec. **e-f**, SunTag-PP7 cells expressing indicated reporters were transfected with 10 nM KIF18B siRNA #1, where indicated. The time from first detection of translation until separation of GFP and mCherry foci (i.e. mRNA cleavage) is shown. Solid lines and corresponding shaded regions represent mean  $\pm$  SEM. **f**, Cells expressing dCas9-KRAB were infected with sgRNA targeting endogenous AGO2 (AGO2 KD), or with full length AGO2 (AGO2 o.e. (overexpression)), where indicated. Dotted lines indicate that the data is replotted from panel e for comparison.



AGO2 was further confirmed by northern blot analysis (Fig. 2e-g and Supplemental Fig. 2j). Together, these results suggest that ribosome-stimulated cleavage by AGO2 may be a common phenomenon in living cells.

### 3.3.3 Ribosomes promote AGO2–target interactions

We considered two models explaining how a translating ribosome could stimulate AGO2-dependent mRNA cleavage. First, ribosomes could promote AGO2–mRNA target interactions (‘binding’ model; Fig. 2h). For example, ribosomes may clear the AGO2 binding site of RBPs or unfold RNA structures that mask the AGO2 binding site. Second, it is possible that ribosome collisions with AGO2 stimulate release of the 5' and 3' cleavage fragments from AGO2 after endonucleolytic cleavage has occurred (‘release’ model; Fig. 2i) (note that our imaging approach cannot distinguish between mRNA cleavage and fragment release). When imaging mRNA cleavage at higher time-resolution, we frequently found a ribosome on the 3' cleavage fragment, which is consistent with a model in which the ribosome clears the AGO2 binding site (binding model), but not with the release model (Fig. 2h and 2j upper panel, and methods 3.5.28–3.5.30). After normalizing the data (see methods 3.5.29) we found that one or more ribosomes was present on the 3' cleavage fragment in 76% and 85%, for KIF18B and GAPDH reporters, respectively (Fig. 2k, black bars, and

**Fig. 2 (opposite page) | Ribosomes stimulate AGO2-dependent mRNA cleavage by promoting AGO2–target interactions.**

**a**, Schematic of indicated reporters. **b–c**, SunTag-PP7 cells expressing indicated reporters were transfected with 10 nM KIF18B siRNA #1 and treated with CHX or Emetine, where indicated. The time from first detection of translation (b) or from CHX or Eme addition (c) (+ Eme, + CHX), until separation of GFP and mCherry foci (i.e. mRNA cleavage) is shown. Solid lines and corresponding shaded regions represent mean  $\pm$  SEM. Dotted lines indicate that the data is replotted from Fig. 1e for comparison. **d**, Ratio of the cleavage rates in the presence and absence of translating ribosomes (translation was inhibited by CHX addition) is shown for the indicated siRNAs and reporters. **e**, Northern blot of cells expressing the KIF18B or KIF18B-early-stop reporter, either non-transfected (no siRNA) or transfected with 10nM KIF18B siRNA #1 (+ siRNA). (top) Upper band (uncleaved) represents the full-length reporter mRNA; lower band (cleaved) represents the 3' cleavage fragment. (bottom) 28S rRNA acts as a loading control. **f**, Ratio of the northern blot band intensity for bands representing cleaved and uncleaved mRNAs for the + siRNA condition. **g**, Ratio of the intensity of the + siRNA and - siRNA uncleaved bands. **f–g**, Each dot represents a single experiment and lines with error bars indicate the mean  $\pm$  SEM. **h–i**, Schematics for h, ‘binding’ and i, ‘release’ models explaining how ribosomes could stimulate AGO2-mediated mRNA cleavage. **j**, Representative images of mRNA molecules in SunTag-PP7 cells expressing the KIF18B-ext reporter showing cleavage events with (top) or without (bottom) a ribosome on the 3' cleavage fragment. Arrows indicate 3' cleavage fragments. Time is indicated as min:sec. **k**, The fraction of mRNAs that contains a ribosome on the 3' cleavage fragment is shown for the data (black bars) and for the indicated models (green and pink bars). P-values in d, f–g are based on a two-tailed Student's t-test. P-values are indicated as \* ( $p < 0.05$ ), \*\* ( $p < 0.01$ ), \*\*\* ( $p < 0.001$ ), ns = not significant.

Supplemental Fig. 2k). Using computational modeling, we found that these values were indeed most consistent with the ‘binding’ model (Fig. 2k, Supplemental Fig. 2l and methods 3.5.30).

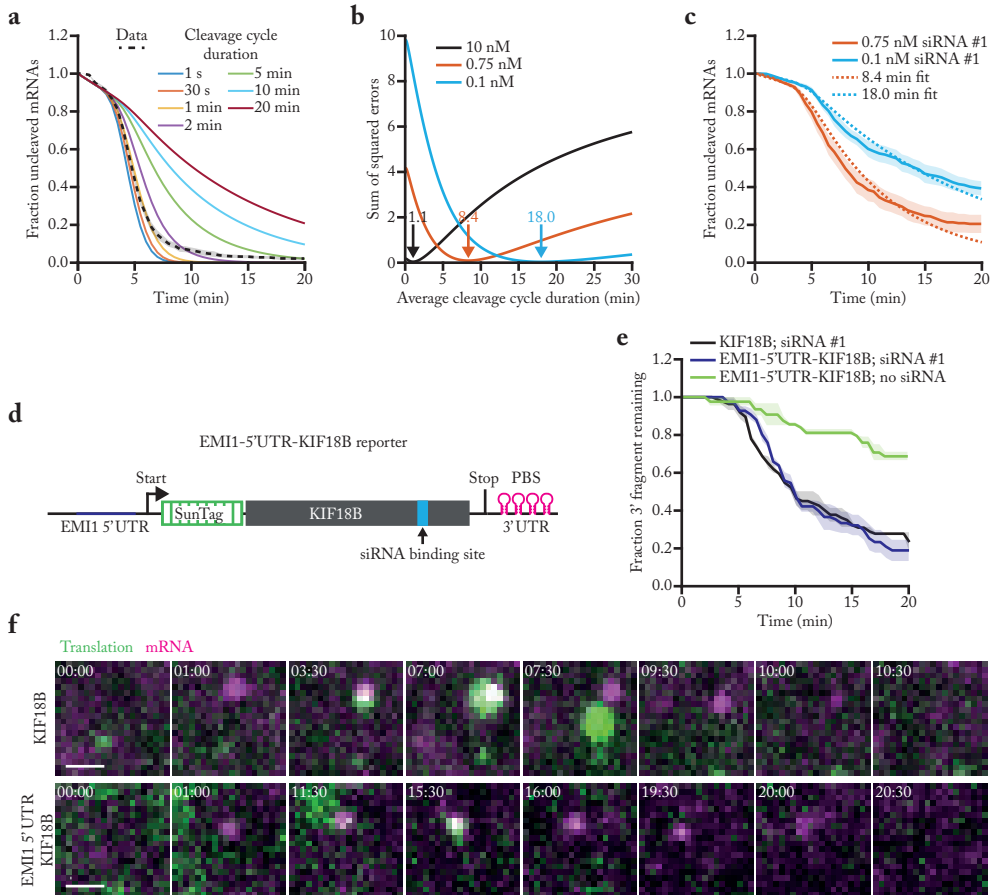
### 3.3.4 In vivo kinetics of the AGO2 cleavage cycle

While several studies have determined the kinetics of each step of the cleavage cycle of AGO2 in vitro [12–16,26,42], very little is known about the cleavage kinetics in vivo. To estimate the duration of the entire cleavage cycle in vivo (from binding site availability to fragment release), we computed cleavage curves using different theoretical AGO2 cleavage cycle durations (Fig. 3a, colored lines show example curves, see methods 3.5.31). We compared the computed cleavage curves with the experimental cleavage curve (Fig. 3a), which revealed that a cleavage cycle duration of  $\leq 1$  min best fit the data (Fig. 3a,b, black line). An in vitro cleavage assay confirmed a cleavage cycle duration of 1-2 min (Supplemental Fig. 3a,b).

Next, we focused on the binding step in more detail. To assess the effective AGO-siRNA concentration in cells at different transfected siRNA concentrations, we decreased the concentration of siRNA from 10 nM to 1.0 nM, 0.75 nM or 0.1 nM to slow down the binding rate and analyzed cleavage rates (Fig. 3c, solid lines). This analysis revealed a linear correlation between cleavage rate and siRNA concentration between 0.1 and 1.0 nM siRNA, but a lower than expected cleavage rate at 10 nM, possibly due to saturation of siRNA association with AGO2 (Supplemental Fig. 3c). Comparison of the cleavage curves for 0.75 nM and 0.1 nM siRNA with simulated cleavage time distributions revealed a good fit with an average cleavage cycle duration of  $\sim 8$  min and  $\sim 18$  min, respectively (Fig. 3b, orange and blue line, Fig. 3c, dotted lines, and methods 3.5.31). Since the catalysis and release steps are unlikely affected by a decrease in the siRNA concentration, these results suggest that even at moderately high siRNA concentrations (i.e. at least between 0.75 nM and 10 nM), target binding is the rate-limiting step, whereas AGO2 structural rearrangements, catalysis and fragment release all occur relatively fast ( $< 1$  min).

It is possible that the estimated time for the release step described above reflects ribosome-stimulated release; our earlier results show that the first ribosome promotes AGO2-target binding by unmasking the target site (see Fig. 2k), but does not exclude the possibility that a following ribosome stimulates fragment release by colliding with AGO2 after catalysis has occurred. However, we observed a very similar cleavage cycle duration for an mRNA reporter translated by a single ribosome [37,43,44] (Fig. 3d-f and methods 3.5.21), further indicating that fragment release occurs rapidly after mRNA cleavage in vivo. Finally, these results also show that a

single ribosome translating the siRNA binding site is sufficient to stimulate binding site accessibility for AGO2/siRNA complexes.



**Fig. 3 | In vivo kinetics of the AGO2 cleavage cycle.**

**a**, Simulated cleavage curves for indicated (theoretical) durations of the cleavage cycle (solid colored lines) are compared to the data (KIF18B + 10 nM siRNA #1; black dotted line, replotted from Fig. 1e). Time represents time since GFP appearance. **b**, Sum of squared error values are shown for different average cleavage cycle durations for indicated siRNA concentrations. Arrows and values indicate average cleavage cycle duration of the optimal fit. **c**, Dotted lines indicate the best cleavage curve fit for the indicated siRNA concentrations. **c**, **e-f**, SunTag-PP7 cells expressing the KIF18B reporter (**c**) or the EMI1-5'UTR-KIF18B reporter (**e-f**) were transfected with KIF18B siRNA #1 at indicated concentrations (10 nM in **e** and **f**). **c**, **e**, The time from first detection of translation until separation of GFP and mCherry foci (**c**) (i.e. mRNA cleavage) or mCherry disappearance (**e**) (i.e. exonucleolytic decay of the 3' cleavage fragment) is shown. Solid lines and corresponding shaded regions represent mean  $\pm$  SEM. **d**, Schematic of the EMI1-5'UTR-KIF18B reporter. **f**, Representative images of a time-lapse movie are shown. Scale bar, 1  $\mu$ m. Time is shown in min:sec. Note that fluorescent intensities for the KIF18B reporter and the EMI1-5'UTR-KIF18B reporter images are scaled differently to allow visualization of the very dim GFP signal associated with translation by a single ribosome (bottom).

### 3.3.5 Interactions of the AGO2 target sequence with flanking mRNA sequences drive target site masking

Translating ribosomes can promote binding site accessibility either by displacing RBPs from the binding site, or by unfolding RNA structure(s) that mask the AGO2 binding site. To distinguish between these possibilities, we designed new reporters (referred to as ‘luciferase reporters’ due to the luciferase gene in the CDS) in which the siRNA binding site is positioned close to the 3’ end of the mRNA (immediately upstream of the PCP binding site array). Due to the position of the siRNA binding site, structures masking the AGO2 binding site will arise predominantly from interactions between the AGO2 binding site and upstream mRNA sequences. Stop codons were introduced either 27 nt or 110 nt upstream of the siRNA binding site (‘late stop’ reporters) (Fig. 4a, left). In these ‘late stop’ luciferase reporters ribosomes can disrupt interactions of the AGO2 binding site with upstream mRNA sequences, without displacing RBPs from the binding site (as the AOG2 binding itself is not translated). As controls, we generated reporters in which the stop codon is positioned downstream of the siRNA binding site (‘downstream stop’ reporter) for which ribosomes can remove both structures and RBPs, or reporters with a stop codon 1677 nt upstream of the binding site (‘early stop’ reporters), for which ribosomes remove neither flanking structures nor RBPs (Fig. 4a, left). For these experiments, we selected the target sites of KIF18B siRNAs #1 and #2, and GAPDH siRNA #3, as each of these target sites showed strong stimulation of cleavage by ribosomes (i.e. target site masking) in their original contexts (see Fig. 2d).

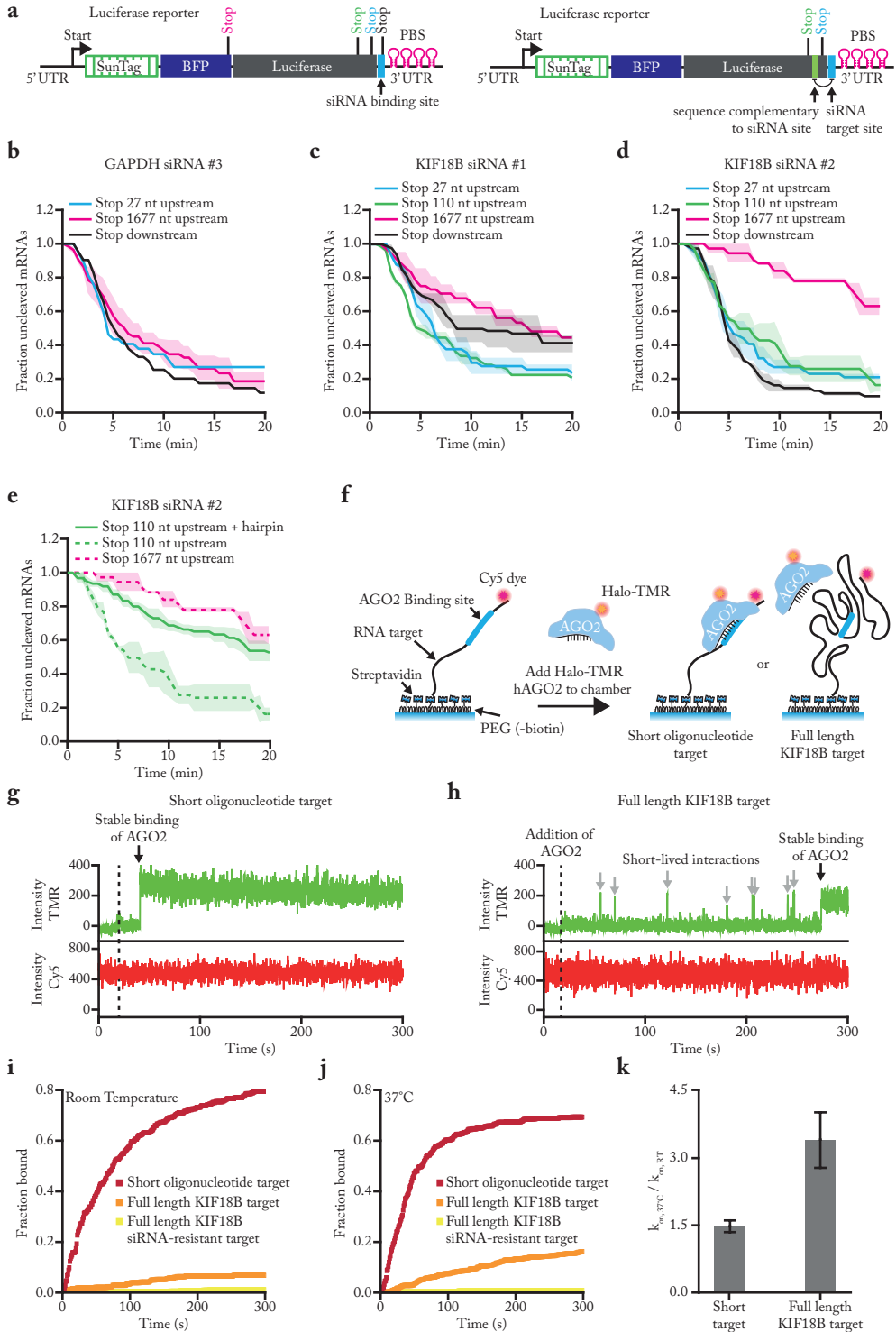
mRNAs containing the GAPDH siRNA #3 site showed fast cleavage even when the AGO2 binding site was positioned in the non-translated region (Fig. 4b), suggesting that the binding site is not masked in these new reporters. In contrast,

**Fig. 4 (opposite page) | Masking of mRNA target sites by RNA structures inhibits AGO2-target interactions.**

**a**, Schematic of the ‘luciferase’ reporter. Position of different stop codons and nt complementary to siRNA target site (right) are indicated. **b-e**, SunTag-PP7 cells expressing the indicated reporters were transfected with 10 nM of the indicated siRNA. The time from first detection of translation until separation of GFP and mCherry foci (i.e. mRNA cleavage) is shown. Solid lines and corresponding shaded regions represent mean  $\pm$  SEM. Dotted lines in (e) indicate that the data is replotted from Fig. 4d for comparison. **f**, Schematic of the in vitro single-molecule binding assay. **g-h**, Representative traces of AGO2-siRNA complex binding to (g) the short oligonucleotide target or (h) the full length KIF18B target. Green line represents Halo-TMR AGO2 (top) and red line represents Cy5 RNA signal (bottom). Grey arrows indicate short binding events by AGO2 and black arrow indicates stable binding by the AGO2-siRNA complex. **i-j**, The cumulative fraction of target RNAs bound by Halo-TMR AGO2 is plotted as a function of time for the indicated reporters at room temperature (RT) (i) and 37°C (j). **k**, Ratio of  $k_{on}$  at 37°C and RT for the short oligonucleotide target and the full length KIF18B target. Data are plotted as mean  $\pm$  SEM (n=3 independent experiments).



# mRNA structural dynamics shape Argonaute–target interactions



for the reporters containing either the KIF18B siRNA #1 or #2 site, cleavage was substantially faster for the ‘downstream stop’ reporters compared to the ‘early stop’ reporters (Fig. 4c,d, compare black and pink lines), indicative of target site masking. Importantly, cleavage rates of the ‘late stop’ reporters were at least as fast as cleavage rates of the ‘downstream stop’ reporter (Fig. 4c,d, compare blue and green to black lines), indicating that ribosomes stimulate AGO2 mRNA binding and cleavage by unfolding mRNA secondary structure, rather than displacing RBPs from the binding site for both reporters. To further confirm the role of RNA structure in AGO2 target site masking, we placed two copies of a 7 nt sequence with complementarity to the siRNA binding site just downstream of the stop codon in the 110 nt-early-stop reporter, embedding the AGO2 binding site in a hairpin structure (‘hairpin reporter’) (Fig. 4a, right). The ‘hairpin reporter’ showed a severely reduced cleavage rate compared to its parent reporter (Fig. 4e, compare solid and dotted green lines), confirming that intramolecular RNA interactions inhibit AGO2–target binding, consistent with previous findings [27]. Surprisingly, for the KIF18B siRNA #1, the rate of cleavage of the ‘early stop’ reporters was even faster than that of the ‘downstream stop’ reporter (Fig. 4c, compare blue and green lines to black line). A possible explanation for this result is that ribosomes passing over the AGO2 binding site impair mRNA cleavage by displacing AGO2 from the mRNA upon collision before cleavage has occurred.

Interestingly, while cleavage by GAPDH siRNA #3 showed strong stimulation by ribosomes when the siRNA binding site was in its native context, the same binding site was no longer ribosome-stimulated in the sequence context of the luciferase reporter (Fig. 4b), suggesting that the interactions between the AGO2 binding site and flanking mRNA sequences determines the degree of binding site masking. Indeed, when AGO2 binding sites were inserted in different mRNAs and at different positions in an mRNA, the magnitude of target site masking (i.e. the ribosome-dependent cleavage stimulation) varied (Supplemental Fig. 4a–g).

To directly test the role of flanking sequences in AGO2 target site masking, we established an *in vitro* assay to visualize AGO2 binding to either a short RNA oligonucleotide or the full length KIF18B mRNA (Fig. 4f–h). As a control, we mutated the siRNA binding site. AGO2 binding to the oligonucleotide target occurred rapidly ( $t_{1/2} = 73 \pm 8$  s, mean  $\pm$  SD), while binding to the full length mRNA target was much slower ( $t_{1/2} = 4.1 \pm 0.7 \times 10^3$  s, mean  $\pm$  SD) (Fig. 4i), indicating that RNA structures formed in the full length transcript inhibit binding of AGO2 to the target site. We did not observe many binding events to the mRNA with a mutated AGO2 target site, suggesting that AGO2 does not stably interact with other sequences in the mRNA (Fig. 4i, yellow line). Since RNA folding is strongly dependent on temperature (with higher temperature resulting in reduced RNA

folding), we repeated the binding assay at 37°C instead of room temperature and found that AGO2 bound to the full length KIF18B target 3.4-fold faster at 37°C, while binding to the oligonucleotide target was only 1.5-fold faster at 37°C (Fig. 4j, k), suggesting that structural remodeling of the mRNA driven by thermal fluctuations affects AGO2 binding site availability.

### 3.3.6 Combinations of multiple weak intramolecular mRNA interactions result in potent AGO2 target site masking

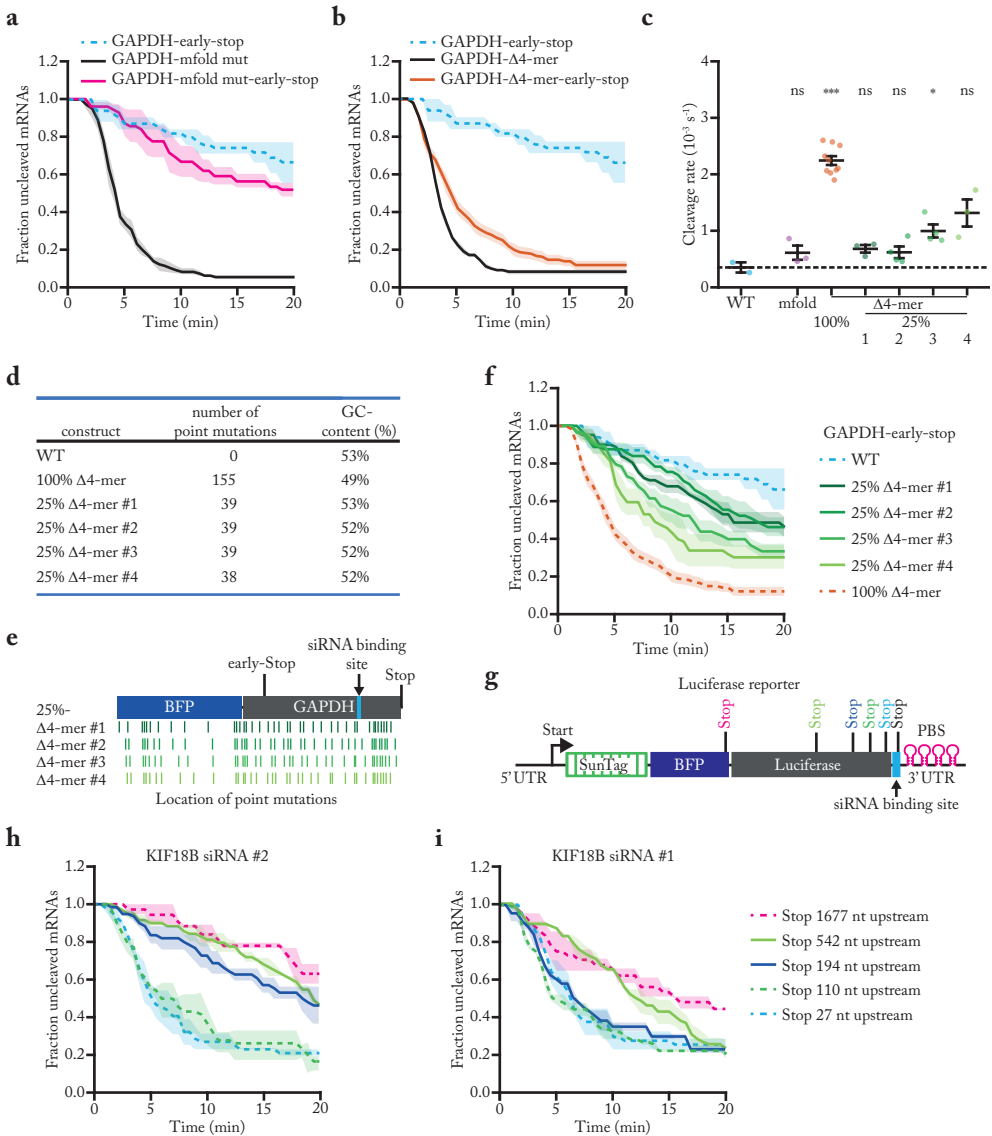
To determine the nature of the structures that mask the AGO2 target sites, we performed structure prediction using mfold [45]. We generated a new reporter ('mfold reporter', based on the GAPDH reporter) that contained 19 substitutions in the mRNA sequence flanking the target site, which disrupted all the strongest predicted RNA structures involving the AGO2 binding site (see methods 3.5.2). Surprisingly however, mRNA cleavage was still strongly stimulated by ribosomes (Fig. 5a), indicating that the AGO2 binding site remained masked by RNA structures in the 'mfold reporter' mRNA.

Possibly, AGO2 binding site masking arises from numerous weak interactions between the AGO2 binding site and short complementary nucleotide sequences in the target mRNA. To test this hypothesis, we mutated 4-mer sequences in the GAPDH mRNA reporter that showed complementarity to the AGO2 target sequence (i.e. disrupting intramolecular RNA-RNA interactions), referred to as the 'Δ4-mer' reporter (see methods 3.5.2). Removal of 4-mers substantially increased the cleavage rate in the absence of ribosomes translating the AGO2 binding site (~6 fold) (Fig. 5b,c, compare blue and orange dots), indicating that AGO2 binding site masking was largely disrupted in the Δ4-mer mRNA. Removal of complementary 4-mers for KIF18B siRNA #1 in the KIF18B reporter also substantially reduced ribosome-dependent stimulation of mRNA cleavage, although residual cleavage stimulation could still be observed (Supplemental Fig. 5a), possibly due to other short sequences with complementarity (e.g. 3-mers or 6 or 7-mers with single mismatches). Together, these results suggest that multiple nucleotide sequences, each with weak affinity for the AGO2 binding site, together can drive strong target site masking.

If many different sequences within the mRNA contribute to AGO2 target site masking, it is likely that a high degree of structural heterogeneity can exist. To test this, we generated four new 'Δ4-mer' reporters; in each of these reporters a non-overlapping set of 25% of the single nucleotide substitutions were introduced that disrupt the complementary 4-mers (Fig. 5d,e, and methods 3.5.2). All four '25% Δ4-mer' reporters showed a partial effect on the cleavage rate (Fig. 5f, c, compare

blue and green dots, and Supplemental Fig. 5b–e), further suggesting that multiple low affinity interactions cooperatively cause AGO2 target site masking, and that structural heterogeneity underlies robust target site masking.

Finally, we varied the distance between the stop codon and the AGO2 target site to map the distances over which flanking sequences can act to mask the AGO2 target site. This revealed that structures spanning several hundred nucleotides can contribute to AGO2 target site masking (Fig. 5g–i and Supplemental Fig. 5f), consistent with



other studies showing that base-pairing interactions can occur over large distances [46,47].

### 3.3.7 mRNA folding kinetics and the translation rate control AGO–target interactions

While several methods are available to capture ‘snapshots’ of RNA structure [46,48–56], very little is known about the structural dynamics of mRNAs *in vivo*. Such dynamics of mRNA (un-)folding are likely important, as structural unmasking of binding sites is a key driver of AGO2–target interactions.

Inhibiting ribosome translocation by addition of CHX results in a decreased cleavage rate (see Fig. 2c and Supplemental Fig. 2c–i), suggesting that mRNAs refold after ribosome-dependent unfolding. We reasoned that upon inhibition of ribosome translocation, the cleavage rate will decrease over time as re-folding of the target site occurs. Indeed, for 3 out of 4 reporter–siRNA combinations, we observed a fast initial cleavage rate after CHX addition, followed by a slower cleavage rate at later time points (Fig. 6a and Supplemental Fig. 6a–c, red lines, see methods 3.5.22). Fitting these cleavage curves with a double exponential decay distribution and correcting for the delay in ribosome stalling upon addition of CHX (Supplemental Fig. 6d, methods 3.5.27) revealed that open AGO2 target sites are masked within ~30–90 s of ribosome-dependent unfolding (Fig. 6a, Supplemental Fig. 6a,b, dotted lines, and Supplemental Fig. 6e). For the fourth reporter–siRNA condition (GAPDH siRNA #3) the cleavage rate was faster than expected (compare Supplemental Fig. 6c, red line, and Supplemental Fig. 2g, pink line), suggesting that the target site remains in a (partially) unmasked state in the presence of stalled ribosomes for this reporter. Possibly, stalled ribosomes near the target site inhibit mRNA re-folding for this

**Fig. 5 (opposite page) | Multiple weak intramolecular mRNA interactions cooperatively mask AGO2–target sites.**

**a, b, f, h–i**, SunTag–PP7 cells expressing the indicated reporters were transfected with (a, b, f) 10 nM GAPDH siRNA #3 or (h–i) indicated siRNAs. The time from first detection of translation until separation of GFP and mCherry foci (i.e. mRNA cleavage) is shown. Solid lines and corresponding shaded regions represent mean  $\pm$  SEM. Dotted lines indicate that the data is replotted from an earlier figure panel for comparison. (a, b) replotted from Supplemental Fig. 1g; (f) replotted from Fig. 5b and Supplemental Fig. 1g; (h) replotted from Fig. 4d; (i) replotted from Fig. 4c. **c**, Calculated cleavage rates in the absence of ribosomes translating the siRNA target site are shown for indicated reporters treated with GAPDH siRNA #3. Each dot represents a single experiment and lines with error bars indicate the mean  $\pm$  SEM. P-values are based on a two-tailed Student’s t-test. P-values are indicated as \* ( $p < 0.05$ ), \*\* ( $p < 0.01$ ), \*\*\* ( $p < 0.001$ ), ns = not significant. **d**, Characteristics of the different  $\Delta$ 4-mer reporters. **e**, Schematic overview of the location of the single nucleotide substitutions in the 25%  $\Delta$ 4-mer reporters. **g**, Schematic overview of the luciferase reporters used in (h–i) containing stop codons at variable distances from the siRNA binding site.

region of the mRNA.

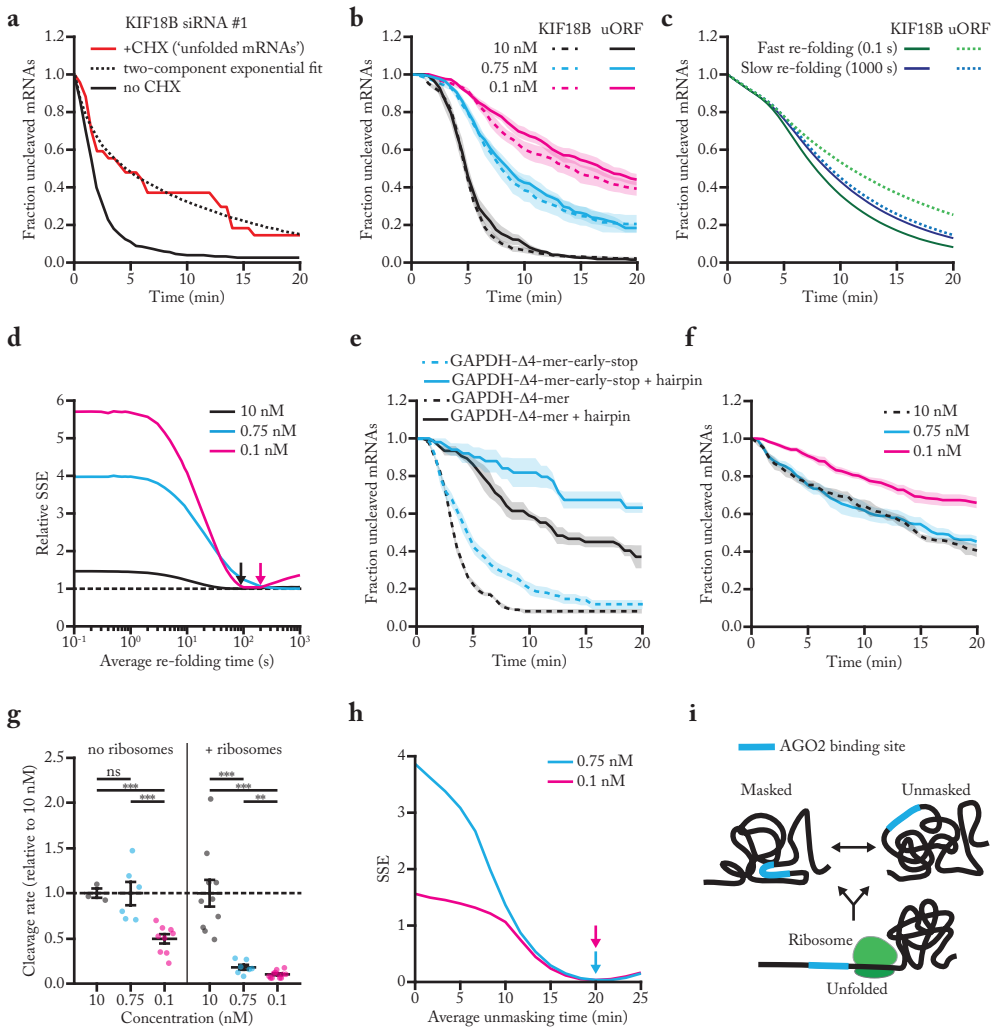
To confirm our measurements of mRNA re-folding kinetics after ribosome-induced unfolding, we examined the relationship between the translation initiation rate, the mRNA re-folding rate, and the mRNA cleavage rate. We reasoned that the cleavage rate depends on the fraction of time that the target site is unmasked. If mRNA re-folding is fast, the target site will be mostly masked and increasing the frequency of ribosome-induced mRNA unfolding (dependent on the translation initiation rate) will increase the cleavage rate. In contrast, if the mRNA re-folds slowly, the target site will still be unmasked when the next ribosome arrives, and increasing the translation initiation rate won't increase the cleavage rate. Thus, by measuring the cleavage rate at different translation initiation rates the mRNA folding rate can be assessed.

We introduced an upstream open reading frame (uORF) into the KIF18B reporter ('uORF-KIF18B' reporter), which reduced the translation initiation rate by 3.3 fold (Supplemental Fig. 6f, and methods 3.5.24). Measuring mRNA cleavage rates for both the KIF18B and 'uORF-KIF18B' reporters at three different siRNA concentrations (10 nM, 0.75 nM, and 0.1 nM) revealed very similar cleavage rates for both reporters at each concentration of siRNA (Fig. 6b), suggestive of a relatively

**Fig. 6 (opposite page) | Kinetics of mRNA folding shape AGO2-mRNA interactions.**

**a**, SunTag-PP7 cells expressing the KIF18B reporter were transfected with 10 nM siRNA KIF18B #1 and treated with CHX, where indicated. Only mRNAs for which translation initiated between 3.5-6.5 min before CHX addition were included (see methods 3.5.22). The time since CHX addition is shown for the '+CHX' cleavage curve. Dotted line represents optimal fit with a two-component exponential decay distribution. The no CHX cleavage curve is re-normalized and plotted from 3.5 min after the start of translation. **b,e-f**, SunTag-PP7 cells expressing indicated reporters were transfected with (b,f) KIF18B siRNA #1 as indicated or (e) 10nM GAPDH siRNA 3 and treated (b, e) without or (f) with CHX. The time from (b, e) first detection of translation or (f) CHX addition until separation of GFP and mCherry foci (i.e. mRNA cleavage) is shown. Solid lines and corresponding shaded regions represent mean  $\pm$  SEM. Dotted lines indicate that the data is replotted from an earlier figure panel for comparison. (b) replotted from Fig. 1e and 3c; (e) replotted from Fig. 4b; (f) replotted from Fig. 2c. **c**, Simulated cleavage curves for the KIF18B reporter and uORF reporter simulated with a fast or slow mRNA re-folding time. Note that the fast and slow re-folding curves use different AGO2 cleavage rates to generate an optimal fit to the data (see methods 3.5.32). **d**, Goodness-of-fit score (relative SSE) at different simulated mRNA re-folding times for the data shown in (b) at indicated siRNA concentrations (see methods 3.5.32). Arrows indicate re-folding time of best fit. **g**, Calculated cleavage rates for different siRNA concentrations (relative to 10 nM siRNA) for the data shown in (b, f). Each dot represents a single experiment and lines with error bars indicate the mean  $\pm$  SEM. P-values are based on a two-tailed Student's *t*-test. P-values are indicated as \* ( $p < 0.05$ ), \*\* ( $p < 0.01$ ), \*\*\* ( $p < 0.001$ ), ns = not significant. **h**, Goodness-of-fit score (SSE) at different simulated unmasking times (see methods 3.5.33). Arrows indicate unmasking time of the best fit. **i**, Schematic model of AGO2 target site masking and unmasking and the role of ribosomes in target site unfolding.

slow mRNA refolding rate. To quantitatively assess the mRNA re-folding rate, we developed a computational framework to simulate mRNA cleavage curves at different (theoretical) AGO2 cleavage rates, mRNA folding rates and translation initiation rates (see methods 3.5.32). As expected, we found that at fast re-folding rates the simulations predict a relatively large difference between the cleavage rate of the KIF18B and uORF-KIF18B reporter, while a small difference in cleavage rate is predicted at slow mRNA re-folding rates (Fig. 6c). To compare the simulated and experimental cleavage curves at different theoretical mRNA folding rates, we computed a goodness-of-fit score (sum of squared errors (SSE)) (Fig. 6d, and methods 3.5.32). For both the 10 nM and 0.1 nM, the optimal fit was achieved when simulating an mRNA re-folding time of ~30-180 s, while a somewhat slower



re-folding time ( $>180$  s; a precise value could not be given due to the absence of a local minimum) was found for the 0.75 nM condition (Fig. 6d). Overall, these results are in reasonably good agreement with the measurements of mRNA re-folding upon CHX treatment (30-90 s).

Our results suggest that the (complex) mRNA structures that stably mask the AGO2 target site re-fold slowly after ribosome-dependent unfolding, allowing the target site to be unmasked continuously even when a ribosome passes the target site only 1-2 times per minute. In contrast for, a simple RNA structure, such as a hairpin, re-folding is expected to occur rapidly after ribosome-mediated unfolding, limiting the stimulatory effect of ribosomes on AGO2-target interactions. To test this, we introduced the siRNA target site into a hairpin structure (see methods 3.5.2). For this experiment, we selected different reporters in which the target site is unmasked in the absence of ribosomes (e.g. GAPDH  $\Delta$ 4-mer reporter, see Fig. 5b orange line, or the luciferase reporter, see Fig. 4d green line). As expected, introduction of the hairpin structure strongly reduced the cleavage rate in the absence of ribosomes (Fig. 6e and Supplemental Fig. 6g, compare blue lines). Interestingly, cleavage rates in the presence of translating ribosomes were also reduced in these hairpin reporters (Fig. 6e and Supplemental Fig. 6g, compare black lines), indicating that ribosomes indeed unmask target sites less efficiently when the target site is present in a fast-folding RNA structure (Supplemental Fig. 6h). These results suggest that ribosomes predominantly stimulate AGO2 target binding by unfolding slowly re-folding structures, which likely represent more complex secondary or tertiary structures.

### 3.3.8 Slow structural dynamics limit AGO2 binding in the 3'UTR

When positioned in non-translated regions of the mRNA (i.e. 3' UTR), AGO2 binding sites are not unfolded by ribosomes, yet cleavage still occurs (albeit at a slower rate). Therefore, structural unmasking of the target site must occur through alternative mechanisms. One possibility is that mRNAs switch between different structural conformations over time and that AGO2 target sites are only masked in a subset of all possible structural configurations. If structural rearrangements occur on a timescale that is much faster than AGO2-target binding ( $\sim$ 1-18 min for 0.1-10 nM siRNA, see Fig. 3b), AGO2-target binding will be rate-limiting for mRNA cleavage in the 3' UTR, and the mRNA cleavage rate will depend primarily on the AGO2-siRNA concentration. In contrast, if structural rearrangements occur at rates similar to or slower than AGO2-target binding, structural unmasking becomes an (additional) rate-limiting step, and the mRNA cleavage rate will become less sensitive to the siRNA concentration. Interestingly, the cleavage rate in the absence of ribosomes showed a weak dependency on siRNA concentration (Fig. 6f,g and



Supplemental Fig. 6i), demonstrating that, in the absence of translating ribosomes, structural unmasking of target sites, rather than AGO2–siRNA concentration in the cell is the main rate-limiting step for mRNA cleavage.

To quantitatively investigate the dynamics of ribosome-independent structural rearrangements, we simulated the effect of decreasing the siRNA concentration on the mRNA cleavage rate in the absence of ribosomes (see methods 3.5.33). For slow (simulated) structural dynamics, we found that the simulated mRNA cleavage rate is less sensitive to siRNA concentration (0.1–10 nM) than for fast dynamics (Supplemental Fig. 6j, compare the solid and dotted pink lines to the black line), consistent with the experimental data (Fig. 6f, compare the pink line to the black line). To extract quantitative information about the ribosome-independent unmasking time of our reporter mRNA through simulations, we compared the experimental cleavage curves at different siRNA concentrations (0.75 nM and 0.1 nM) to multiple simulated cleavage curves (each with different unmasking times) using a goodness-of-fit score (SSE). Further analysis revealed that the best fits were obtained with unmasking times of >10 min (Fig. 6h), indicating that target site unmasking becomes a rate-limiting step in mRNA cleavage, especially at higher concentrations of siRNA (between 0.75 nM and 10 nM). Furthermore, these simulations indicate that target site unmasking in the 3' UTR (i.e. in the absence of translating ribosomes) is much slower (>10 min) than the unmasking rate in the CDS (where target sites are unfolded every ~25 s by a translating ribosome for our reporters), highlighting the importance of ribosome-mediated unmasking of target sites for efficient AGO–target interactions.

### 3.4 Discussion

In this study, we use a live-cell imaging approach to visualize translation and AGO2-mediated cleavage of individual mRNA molecules. This work provides *in vivo* measurements of AGO2 cleavage kinetics and reveals how mRNA structural dynamics and heterogeneity shape AGO–target interactions.

#### 3.4.1 Paradoxical roles of ribosomes in controlling AGO2–mRNA target interactions

Recent reports showed that ribosomes reduce the overall degree of structure in the CDS of the transcriptome [27,54,57,58] and that AGO target sites are less efficiently recognized if they are embedded within a strong structure [26,27,29–31]. Here, we show that ribosome-dependent unfolding of mRNA structures stimulates AGO–target interactions, thereby providing a direct, causal link between mRNA translation and AGO2–target binding. Interestingly, not all siRNA–target combinations were

stimulated to the same extent by ribosomes (see Fig. 2d), indicating that some siRNA target sites are always accessible, or, alternatively, are masked by mRNA structures that re-fold rapidly after ribosome-dependent unfolding.

The observation that siRNA-mediated mRNA cleavage is more efficient in an actively translated region appears to contrast previous reports that miRNAs repress their target more efficiently when bound to the 3' UTR [24,25]. It is possible that ribosomes also inhibit AGO-target interactions, for example by displacing AGO from the mRNA through physical collisions [24,59]. Thus, ribosomes may have two opposing activities that affect AGO-target interactions. The net effect of ribosomes on AGO-dependent target silencing may depend on a number of different factors, including the degree of target masking, the translation initiation rate and the time required for AGO-siRNA complexes to repress their target mRNA upon binding. Interestingly, previous analysis revealed that miRNA target sites positioned immediately downstream of the stop codon are highly active [24]. Consistent with this, we find that AGO binds very efficiently to target sites positioned immediately downstream of the stop codon, because ribosomes translating upstream mRNA sequences can stimulate unmasking of target sites immediately downstream of the stop codon (see Fig. 4c,d). Therefore, binding sites immediately downstream of the stop codon may be most potent, as they benefit from the stimulatory activity of ribosomes, while being protected from the inhibitory effect of ribosome-AGO collisions. A paradoxical role of ribosomes in both stimulating and inhibiting RBP-mRNA interactions may not be limited to AGO family proteins, but may broadly shape the interactions of RBPs with their target RNAs.

### **3.4.2 mRNA structural dynamics and heterogeneity**

For AGO target mRNA sites masked by RNA structure, mutation of all or subsets of short complementary 4-mer sequences in the target mRNA reduced target site masking (see Fig. 5f), indicating that multiple (or many) sequences in the target mRNA can contribute to target site masking and that different structural configurations exist that can mask the AGO2 target site. In addition, mRNA molecules may also form intermolecular RNA-RNA interactions, potentially further inhibiting AGO2-target interactions. Finally, RBPs may also inhibit AGO2 binding at specific target sites, although our results suggest that inhibition through structural masking is a more common mechanism (see Fig. 4c-d).

Individual 4-mer base-pair interactions generally have rapid binding and unbinding kinetics. Surprisingly though, our data suggest that target sites can remain masked for >10 min in the absence of ribosomes (see Fig. 6h). So how can we reconcile these

two apparently contradictory findings? One speculative model is that mRNAs form stable 3-dimensional structures in which multiple sequences with weak affinity for the AGO2 target site are positioned in close proximity to the target site, resulting in frequent interactions and robust target site masking (Fig. 6i). In this model, the key activity of ribosomes would be to unfold the stable 3-dimensional structure that facilitates target site masking, rather than directly disrupting target site interactions with complementary sequences. In the absence of translating ribosomes such structures could persist for long periods of time (>10 min), explaining the very slow cleavage kinetics for some reporters in which the AGO2 binding is located in the 3' UTR (e.g. see Fig. 2c and Supplemental Fig. 2g). Possibly, such 3-dimensional structures stochastically rearrange over time through thermal fluctuations of the mRNA, resulting in AGO2 target site unmasking (without complete unfolding of the mRNA), or structures are unfolded or refolded sporadically by cellular helicases such as EIF4A [60] to allow target cleavage. Together, these results provide a high temporal resolution analysis of the structural dynamics of an mRNA molecule in vivo and a framework for understanding the role of mRNA structural dynamics in shaping RBP–mRNA interactions.

## 3.5 Methods

### 3.5.1 Cell culture

Insect Sf9 cells (Expression Systems (Davis, CA), 94-001S) were grown in Insect Xpress medium (Lonza). Human U2OS cells (ATCC, HTB-96) and HEK293T cells (ATCC, CRL-3216) were grown in DMEM (4.5g/L glucose, Gibco) containing 5% fetal bovine serum (Sigma-Aldrich) and 1% penicillin and streptomycin (Gibco). Cells were grown at 37°C and with 5% CO<sub>2</sub>. Where indicated, Cycloheximide (CHX) (ThermoFisher) was used at a final concentration of 200 µg/ml, and Emetine (Eme) (Sigma-Aldrich) was used at a final concentration of 100 µg/ml. All human cell lines were tested for mycoplasma and found mycoplasma free.

Live-cell imaging experiments were performed using U2OS cells, stably expressing TetR, scFv-sfGFP and PCP-mCherry-CAAX (referred to as SunTag-PP7 cells) [37] as well as the reporter of interest. The smFISH imaging experiments were performed in a monoclonal cell line, stably expressing TetR, scFv-sfGFP, PCP-Halo-CAAX and the 24xGCN4-KIF18B-24xPP7 reporter. Northern blot experiments were performed using two monoclonal cell lines, both expressing TetR, scFv-sfGFP, PCP-Halo-CAAX and in addition either the 24xGCN4-KIF18B-24xPP7 reporter or the 24xGCN4-KIF18B-early-stop-PP7 reporter.

### 3.5.2 Plasmid design

For reporters where a stop codon was introduced upstream of the siRNA binding site, the distance between the stop codon and the siRNA binding site was as follows:

KIF18B siRNA #1	1050 nt	(nt 1029 of KIF18B)
KIF18B siRNA #2	826 nt	(nt 321 of KIF18B)
KIF18B siRNA #3	1355 nt	(nt 1029 of KIF18B)
GAPDH siRNA #1	235 nt	(nt 121 of GAPDH)
GAPDH siRNA #2	711 nt	(nt 121 of GAPDH)

GAPDH siRNA #3	518 nt	(nt 121 of GAPDH)
BFP-Luciferase 1677 nt stop	1677 nt	(immediate upstream of luciferase)
BFP-Luciferase 27 nt stop	27 nt	(immediate downstream of luciferase)
BFP-Luciferase 110 nt stop	110 nt	(nt 1567 of luciferase)
BFP-Luciferase 194 nt stop	194 nt	(nt 1482 of luciferase)
BFP-Luciferase 542 nt stop	542 nt	(nt 1135 of luciferase)

Mutant reporters, that were designed to reduce mRNA structures using the mfold web server (mfold mutants) [45] and  $\Delta 4$ -mer mutant reporters, were ordered from Genewiz. To design the mfold mutants we examined the 32 most stably folded structures that were generated by the mfold web server in BFP-GAPDH in which the GAPDH siRNA3 binding site was part of an RNA duplex. To remove possible structures caused by base pairing between the siRNA binding site and the surrounding sequences, we introduced single nucleotide substitutions in the sequences predicted to base pair with the siRNA binding site. A total of 19 point mutations were required to remove all possible interactions between the siRNA binding site and the surrounding sequences that were present in the 32 most stably folded structures predicted by mfold. When designing the mfold mutant reporter, the GC content was kept within similar range as the parental reporter. 10/19 mutations were made in the BFP sequence, and 9/19 mutations were made in the GAPDH sequence, 7 of which were upstream and 2 of which were downstream of the siRNA binding site. In 11/19 mutations, altering the nucleotide sequence changed the encoded amino acid. To assess cleavage rates in the absence of ribosomes translating the siRNA binding site, a stop codon was introduced 518 nt upstream of the siRNA binding site. 9 mutations (all in the GAPDH sequence) were downstream of this stop codon.

To design the  $\Delta 4$ -mer mutants in BFP-GAPDH and KIF18B, stretches of 4 consecutive nucleotides (4-mer) which had complete sequence complementarity to a 4-mer sequence in the siRNA binding site (GAPDH siRNA #3 and KIF18B siRNA #1, respectively) were identified. For these 4-mers, a single nucleotide substitution was introduced to disrupt complementarity with the siRNA binding site. In BFP-GAPDH 155 point mutations were introduced, 50 of which were in BFP and 105 in GAPDH. To assess cleavage rates in the absence of ribosomes translating the siRNA binding site, a stop codon was introduced 518 nt upstream of the siRNA binding site. 88 point mutations were downstream of this early-stop codon. If possible, the nucleotide sequence was altered without changing the amino acid sequence. However, in 3/155 mutations the altered nucleotide sequence resulted in a change in the encoded amino acid. Note that for GAPDH siRNA #3 sequence, nucleotides 6-9 and nucleotides 15-18 are fully complementary. This complementarity was not disrupted, as this would change the GAPDH siRNA #3 sequence. In addition, due to a design error 1/20 4-mers, which has complementarity to nt 7-10 of the siRNA binding site, was not disrupted in this reporter, resulting in the presence of seven positions in the construct where 4 consecutive nucleotides are complementary to the GAPDH siRNA #3 sequence. To design the four 25%  $\Delta 4$ -mer mutants in BFP-GAPDH, 25% of the point mutations were made, resulting in a total of 39 point mutations compared to the WT BFP-GAPDH reporter. In order to make four reporters with a non-overlapping set of point mutations, different point mutations were introduced in each reporter. In the 25%  $\Delta 4$ -mer reporter #1, point mutations 1, 5, 9 etc. were made, in reporter #2, point mutations 2, 6, 10 etc. were made, in reporter #3, point mutations 3, 7, 11 etc. were made, and in reporter #4, point mutations 4, 8, 12 etc. were made. In total 5 point mutations were shared between the mfold BFP-GAPDH reporter and the  $\Delta 4$ -mer BFP-GAPDH reporter. In KIF18B 211 point mutations were introduced. To assess cleavage rates in the absence of ribosomes translating the siRNA binding site, a stop codon was introduced 1050 nucleotides upstream of the siRNA binding site. Here, 118 point mutations were downstream of this early-stop codon. In 6/211 mutations, altering the nucleotide sequence resulted in a change in the encoded amino acid.

To insert the siRNA target sequence in the GAPDH  $\Delta 4$ -mer reporter in a hairpin, we added, both

upstream and downstream of the siRNA target site, two copies of a 7 nt sequence with complementarity to the siRNA binding site. To assess cleavage rates in the absence of ribosomes translating the siRNA binding site, a stop codon was introduced 518 nt upstream of the siRNA binding site. To insert the siRNA target sequence in the Luciferase reporter (with KIF18B siRNA #2) in a hairpin, we added two copies of a 7 nt sequence with complementarity to the siRNA binding site between the location of the 110 nt stop and the 27 nt stop. To assess cleavage rates when ribosomes are allowed to disrupt interactions of the AGO2 binding site with upstream mRNA sequences (one arm of the hairpin, formed by the 7 nt complementary sequence), we used a reporter with a stop codon at the 27 nt position. To assess cleavage rates in the absence of ribosomes translating the newly introduced arm of the hairpin of the siRNA binding site, we used a reporter with a stop codon at the 110 nt position.

To create the KIF18B reporter in which the binding site for KIF18B siRNA #1 was disrupted (siRNA-resistant reporter), we introduced 5 mutations in the siRNA binding site. The sequence of the siRNA #1 binding site in this reporter is GCgACtGTgATaAAAAGtC compared to GCCACAGTCATCAAAAAGCC in the KIF18B reporter.

### 3.5.3 Plasmid transfections for stable integration

Cells were plated one day prior to transfection in a 6 cm dish (Greiner Bio-one). A transfection mix, containing 100  $\mu$ l OptiMEM (Sigma-Aldrich), 2  $\mu$ l FUGENE 6 (Promega), and ~1  $\mu$ g of DNA, was added to the cells in a total volume of 1 ml cell culture medium per dish. Selection for stable integration was initiated 24h after transfection, using 0.4 mg/ml Zeocin (Invitrogen), and continued for at least 10 days. To generate monoclonal cell lines, single cells from the polyclonal cell line were sorted into 96-well plates (Greiner Bio-one) by FACS, and grown for 14 days. Individual clones were inspected by microscopy and clones in which a high percentage of cells expressed the transgene were selected for further use. For generating stable monoclonal cell lines expressing reporter mRNA, clones were additionally screened for the number of mRNAs expressed per cell. Clones expressing ~10-50 mRNAs per cell were selected.

### 3.5.4 siRNA transfections

The complete list and sequence of all siRNAs is:

KIF18B siRNA #1:	5' phosphate-GGCUUUUGAUGACUGUGGCtg 3'
KIF18B siRNA #2:	5' phosphate-UACACUUGGAGCUUCUCCUu 3'
KIF18B siRNA #3:	5' phosphate-UCGCAACGCGCUUCUUCUuu 3'
GAPDH siRNA #1:	5' phosphate-AAAGUUGUCAUGGAUGACCuu 3'
GAPDH siRNA #2:	5' phosphate-AGGUCCACCACCCUGUUGCuu 3'
GAPDH siRNA #3:	5' phosphate-UUCACCACCUUCUUGAUGUuu 3'
TPI siRNA #1:	5' phosphate-UUGGACUUCAGCCAUCCUCuu 3'
TPI siRNA #2:	5' phosphate-AGUCCUUCACGUUAUCUGCuu 3'

siRNAs were designed using the siDESIGN center (horizon) and ordered from Dharmacon, except KIF18B siRNA #1 (AM16708, 251223; ThermoFisher) and GAPDH siRNA #1 (4390849, ThermoFisher). siRNAs were reverse-transfected at a final concentration of 10 nM (unless stated otherwise) using RNAiMAX (Invitrogen) according to the manufacturer's guidelines. KIF18B siRNA #3 was transfected at a final concentration of 50 nM, as it showed weak target repression at 10 nM. For all microscopy experiments, cells were seeded at a confluency of ~40-50% in 96-well glass-bottom imaging plates (Matriplate, Brooks) in a final volume of 200  $\mu$ l and imaged 16-24 hr after transfection. For northern blot experiments, cells were seeded at a confluency of ~40-50% in a 6 cm plate (Greiner Bio-one) in a final volume of 3 ml and harvested 16-24 hr after transfection. For qPCR experiments, cells were seeded in a 24-well plate (Greiner Bio-one) and harvested 16-24 hr after transfection.

### 3.5.5 Lentivirus production and infection

For lentivirus production, HEK293T cells were plated in a 6-well plate (Greiner Bio-one) at 30% confluency, and transfected 24 hr after plating with a mixture of 50  $\mu$ l OptiMEM (Sigma-Aldrich), 10  $\mu$ l Polyethylenimine (PEI) (Polysciences Inc) (1mg/ml), 0.4  $\mu$ g pMD2.g, 0.6  $\mu$ g psPAX2, and 1  $\mu$ g of lentiviral vector. The medium was replaced with 2 ml fresh culture medium 24 hr after transfection, and 72 hr after transfection, viral supernatant was collected. For lentiviral infections, cells were seeded in a 6-well plate (Greiner Bio-one) at 70% confluency. Viral supernatant was added to the cells along with Polybrene (10  $\mu$ g/ml) (Santa Cruz Biotechnology Inc) and the cells were spun at 2000 rpm for 90 min at 37°C (spin infection). After the spin infection, the culture medium was replaced with fresh medium, and cells were incubated for at least 48 hr before further analysis.

### 3.5.6 CRISPRi-mediated knockdown of endogenous AGO2

To knockdown endogenous AGO2 we made use of CRISPRi, since we reasoned that siRNA-mediated approaches would not be very efficient as they rely on the presence of AGO2. For CRISPRi-mediated knockdown of AGO2 we expressed dCAS9-BFP-KRAB in cells stably expressing TetR, scFv-sfGFP, PCP-Halo-CAAX and 24xGCN4-KIF18B-24xPP7. The 30% highest BFP expressing cells were isolated by FACS sorting for further use. A sgRNA targeting AGO2 (sequence: GCGCGTCGGGTAAACCTGTT) was expressed in cells together with BFP through lentiviral transduction. The BFP signal associated with the sgRNA was much higher than the BFP associated with dCAS9-BFP-KRAB, and thus sgRNA positive cells could be identified in dCAS9-BFP-KRAB-expressing cells. qPCR and imaging were performed 4 to 5 days after infection with the sgRNA. In experiments where cleavage was measured after AGO2-knockdown in combination with expression of an exogenous AGO2 rescue construct (insensitive to the sgRNAs targeting endogenous AGO2), cells were infected with an AGO2 expression construct 10 to 11 days before imaging. As AGO2 rescue construct, we used pLJM1-FH-AGO2-WT, which was a gift from Joshua Mendell (Addgene plasmid #91978; <http://n2t.net/addgene:91978>) [61]. Cells expressing exogenous AGO2 were selected with puromycin (2  $\mu$ g/ml) (ThermoFisher). Infection with exogenous AGO2 was followed by a second infection 4 to 5 days before imaging with the sgRNA targeting AGO2 to knockdown endogenous AGO2.

### 3.5.7 smFISH

Single-molecule Fluorescence In Situ Hybridization (smFISH) was performed as described previously [5,62,63]. Five oligonucleotide probes against the PP7-array and 48 probes against the SunTag-array were designed, using the website [www.biosearchtech.com](http://www.biosearchtech.com). Probes were synthesized with a 3' amine modification. Probes were then coupled to either a Cy5 or an Alexa 594 fluorescent dye (Cy5 succinimidyl ester (GE Healthcare) or Alexa Fluor 594 Fluorocarboxylic acid succinimidyl ester (Molecular probes/Invitrogen), respectively) as described previously [63], and HPLC purified (ELLA Biotech GmbH). Purified probes were dissolved in 50  $\mu$ l TE and used at a final dilution of 1:2000. For hybridization, cells were plated in a 96-wells glass bottom dish (Matriplate, Brooks) 16-24 hr before fixation. Doxycycline (1  $\mu$ g/ml) (Sigma-Aldrich) was added 40 to 90 min before fixation (as indicated). Cells were fixed in PBS with 4% paraformaldehyde (Electron Microscopy Science) for 15 minutes at Room Temperature (RT), washed twice with PBS and incubated for 30 min in 100% ethanol at 4°C. After fixation, cells were washed twice in hybridization buffer with 10% formamide (ThermoFisher) at RT, followed by overnight incubation with the probes in hybridization buffer at 37°C. Following overnight incubation, samples were washed 3x for 1 hour in wash buffer at 37°C. DAPI (Sigma-Aldrich) was added to the final wash step in order to stain the nuclei. Shortly before imaging, samples were placed in anti-bleach buffer [62,63] to reduce fluorescence bleaching.

### 3.5.8 Expression and purification of TMR-HaloTag-AGO2-siRNA

His<sub>6</sub>-Flag-TEV-Halo-tagged human AGO2 protein was expressed in Sf9 cells using a baculovirus system (Invitrogen). 750 ml of Sf9 cells at  $1.7 \times 10^6$  cells/ml were infected for 60 hours at 27°C. Infected cells were harvested by centrifugation and resuspended in 30 ml of Lysis Buffer (50 mM NaH<sub>2</sub>PO<sub>4</sub>, pH 8, 300 mM NaCl, 5% glycerol, 0.5 mM TCEP). Resuspended cells were lysed by passing twice through an M-110P lab homogenizer (Microfluidics). The resulting total cell lysate was clarified by centrifugation (30,000 x g for 25 min) and the soluble fraction was applied to 1.5 ml (packed) Ni-NTA resin (Qiagen) and incubated at 4°C for 1.5 h in 50 ml conical tubes. Resin was pelleted by brief centrifugation and the supernatant solution was discarded. The resin was washed with 50 ml ice cold Nickel Wash Buffer (300 mM NaCl, 15 mM imidazole, 0.5 mM TCEP, 50 mM Tris, pH 8). Centrifugation and wash steps were repeated a total of three times. Co-purifying cellular RNAs were degraded by incubating with 100 units of micrococcal nuclease (Clontech) on-resin in ~15 ml of Nickel Wash Buffer supplemented with 5 mM CaCl<sub>2</sub> at room temperature for 45 minutes. The nuclease-treated resin was washed three times again with Nickel Wash Buffer and then eluted in four column volumes of Nickel Elution Buffer (300 mM NaCl, 300 mM imidazole, 0.5 mM TCEP, 50 mM Tris, pH 8). Eluted AGO2 was incubated with a synthetic siRNA and 150 µg of TEV protease during an overnight dialysis against 1–2 liters of Dialysis Buffer (300 mM NaCl, 0.5 mM TCEP, 50 mM Tris, pH 8) at 4°C. The sequence of the siRNA is: 5' phosphate-rUrGrCrUrUrUrUrGrArUrGrArCrUrGrUrGrGrCTT3'. Please note that the first nucleotide is a U instead of a G (as in the original KIF18B siRNA #1 sequence) to improve siRNA loading in AGO2, which does not affect AGO2-target binding [64]. AGO2 molecules loaded with the siRNA were isolated using an immobilized capture oligonucleotide with complementarity to the siRNA, and then eluted by adding competitor DNA with more extensive complementarity to the capture oligonucleotide via the Arpon method [65]. Loaded AGO2 proteins were further purified by size exclusion chromatography using a Superdex Increase 10/300 column (GE Healthcare Life Sciences) equilibrated in 1 M NaCl, 50 mM Tris pH 8, 0.5 mM TCEP. Purified Halo-AGO2-siRNA complex was incubated with Halo-TMR ligand (Promega) and dialyzed against 2 L of 1xPBS (137 mM NaCl, KCl 2.7 mM, 10 mM Na<sub>2</sub>HPO<sub>4</sub>, 1.8 mM KH<sub>2</sub>PO<sub>4</sub>), concentrated to ~2 mg/ml, aliquoted, flash frozen with liquid N<sub>2</sub>, and stored at -80°C.

### 3.5.9 In vitro target RNA synthesis and purification

A short RNA oligonucleotide (with complementarity to the siRNA sequence) was ordered from IBA Lifesciences, labeled with a Cy5 dye (Sigma-Aldrich), as described previously [66], and purified using ethanol precipitation. The labeled oligonucleotide was subsequently ligated to a U30-mer with biotin using T4 RNA ligase II (NEB) and a DNA splint.

Full length mRNA targets (KIF18B sequence or KIF18B sequence with a mutated siRNA target site) were in vitro transcribed using the HiScribe™ T7 High Yield RNA Synthesis Kit (NEB), and purified using phenol-chloroform extraction and ethanol precipitation. The full length mRNA targets were ligated to a 22 nt Cy5 labeled and biotinylated RNA oligonucleotide using a 40 nt DNA strand as a splint. After ligation with T4 RNA ligase II (NEB), the ligated constructs were purified from an agarose gel using a Zymo Gel RNA recovery kit (Basclear).

### 3.5.10 In vitro cleavage assay

Slicing reactions were carried out with loaded AGO2 proteins. Briefly, 10 nM of loaded AGO2 was added to 1 nM Cy5 labeled siRNA5 target in cleavage buffer (20 mM Tris, pH 8.0, 0.15 M Sodium Chloride, 2 mM Magnesium Chloride and 0.5 mM TCEP) at 37°C in a total reaction volume of 100 µl. At each time point, 10 µl was removed and added to 10 µl loading buffer (99.5% formamide and 10 mM EDTA) to quench the reaction. Finally, samples were resolved on a 12.5% denaturing polyacrylamide gel and visualized with an Amersham Typhoon Imaging System.

### 3.5.11 In vitro binding assay

Quartz slides were prepared as described previously [67]. Briefly, quartz slides and coverslips were treated with KOH after which slides were treated with piranha followed with (3-Aminopropyl) triethoxysilane (APTES) (Sigma-Aldrich). Slides were subsequently PEGylated with mPEG-SVA (MW 5.000) (Laysan) under a humid atmosphere overnight. Before experiments, an additional round of PEGylation took place with (MS)PEG-4 (ThermoFisher). Quartz slides and coverslips were assembled with double-sided scotch tape after which the chambers were sealed with epoxy glue. Next, slides were incubated with T50 and 1% Tween-20 for > 10 min to further passivate the chambers [68]. Chambers were subsequently rinsed with T50 and streptavidin (0.1 mg/ml) (ThermoFisher) was introduced inside the chamber for 1 min and rinsed out by T50. The RNA sample was then introduced inside the chamber at a concentration of 100 pM. After 1 minute incubation, unbound RNAs were flushed out with T50. Tubing was attached to the outlet of the microfluidic chambers through epoxy glue and an injection needle was attached to the other side of the tubing. Imaging buffer (150 mM NaCl, 2 mM MgCl<sub>2</sub>, 50 mM Tris (pH 8.0), 1 mM Trolox, 0.8% glucose, 0.1 mg/ml glucose oxidase (Sigma-Aldrich), and 17 µg/ml catalase (ThermoFisher)) was then introduced inside the chamber.

A short movie was taken with the 637 nm laser as a reference for the position of the RNAs of interest (referred to as reference movie), since the RNA molecules are labeled with Cy5. Subsequently, both 532 nm and 637 nm lasers were turned on, and movies were taken for 3500 frames at an exposure time of 0.1 s (referred to as measurement movies). After 200 frames, 1 nM Halo-TMR AGO2 complexes together with imaging buffer were introduced in the microfluidic chamber.

### 3.5.12 Microscopy

In vitro imaging experiments were performed on a custom built inverted microscope (IX73, Olympus) using prism-based total internal reflection. The Halo-TMR was excited with a 532nm diode laser (Compass 215M/50mW, Coherent), and Cy5 was excited with a 637 nm diode laser (OBIS 637 nm LX 140 mW). A 60x water immersion objective (UPLSAPO60XW, Olympus) was used for imaging. A 532 nm notch filter (NF03-532E-25, Semrock) and a 633 nm notch filter (NF03-633E-25, Semrock) were used to block the scattered light. A dichroic mirror (635 dcmr, Chroma) separates the fluorescence signal into separate channels, and the light is projected onto an EM-CCD camera (iXon Ultra, DU-897U-CS0-#BV, Andor Technology). The in vitro experiments were either performed at RT (20°C) or at 37°C through the use of a custom built heating elements and custom written Labview code. All in vivo imaging experiments were performed using a Nikon TI inverted microscope with perfect focus system equipped with a Yokagawa CSU-X1 spinning disc, a 100x 1.49 NA objective and an iXon Ultra 897 EM-CCD camera (Andor) using Micro-Manager software [69] or NIS elements software (Nikon). All live-cell imaging experiments were performed at 37°C, while smFISH experiments were imaged at RT.

For the live-cell imaging of mRNA cleavage experiments, cell culture medium was replaced with pre-warmed CO<sub>2</sub>-independent Leibovitz's-15 medium (Gibco) containing 5% fetal bovine serum (Sigma-Aldrich) and 1% penicillin and streptomycin (Gibco) 15 to 30 minutes before imaging. Transcription of the reporters was induced by addition of doxycycline (1 µg/ml) (Sigma-Aldrich) to the cell culture medium. During the experiments, cells were maintained at a constant temperature of 37°C. Cells were selected for imaging based on the levels of mature protein (an indication of reporter expression) and a small number of mRNAs at the start of imaging [70]. For CRISPRi experiments, cells were additionally selected based on the presence of BFP. Camera exposure times of 500 ms were used for both GFP (488 laser) and mCherry (561 laser), and images were acquired every 30 s for 45 minutes, unless stated otherwise. Since mRNAs are tethered to the plasma membrane, we focused the objective slightly above the plasma membrane to focus on both mRNAs and translation sites, and single Z-plane images were acquired. For the smFISH experiments, images for all 3 colors (DAPI, Cy5 and Alexa 594)



were acquired with a camera exposure time of 50 ms. Z-stacks were acquired for all 3 colors with an inter-slice distance of 0.5  $\mu\text{m}$  each.

### 3.5.13 Northern blot

Northern blots were performed using NorthernMax-Gly kit from ThermoFisher according to the protocol supplied by the manufacturer. In short, cells stably expressing the TetR, scFv-sfGFP, PCP-Halo-CAAX, and 24xGCN4-KIF18B-24xPP7 or 24xGCN4-KIF18B-early-stop-24xPP7 were incubated for 90 min with doxycycline (1  $\mu\text{g}/\text{ml}$ ) (Sigma-Aldrich) to induce expression of the mRNA reporter and RNA was extracted using TRIsure (Bioline). RNA mixed 1:1 with Glyoxal Load Dye was incubated for 30 min at 50°C to denature RNA before loading 10  $\mu\text{g}$  of RNA onto a 0.8% agarose gel. After running the gel, rRNA (18S and 28S) bands were visualized using UV and signal intensities were quantified to ensure that RNA samples were loaded equally and RNA was intact. RNA transfer from the agarose gel to a positively charged nylon membrane was performed for 2 hours at RT, followed by RNA crosslinking to the membrane using UV light (120mJ/cm<sup>2</sup> at 254 nm) for 1 min. After prehybridization at 68°C for 1 hour, the membrane was incubated with a DIG-labeled RNA probe targeting the BGH sequence present in the 3' UTR of the mRNA reporter, and hybridization was performed overnight at 68°C. The membrane was washed 3x, and incubated with the anti-DIG antibody-AP (Sigma-Aldrich) for 16 to 40 hours at 4°C. The membrane was washed 9x (6x in PBS-T and 3x in AP buffer) and incubated with a few drops of CDP-star for 5 minutes at RT. The film was exposed and developed for 2 to 10 minutes, using an Amersham Imager 600 (GE).

### 3.5.14 Quantitative RT-PCR (qPCR)

The complete list and sequence of primers for RT-PCR used in this study is:

GAPDH Forward Primer: 5' CACCGTCAAGGCTGAGAACGGG 3'  
 GAPDH Reverse Primer: 5' GGTGAAGACGCCAGTGGACTCC 3'  
 KIF18B Forward Primer: 5' CACCAAGGTGTCCTAACCTACCG 3'  
 KIF18B Reverse Primer: 5' TGAGTAGCAGGATGGATGTCTGG 3'  
 AGO2 Forward Primer: 5' GTTTCTCCTCTGATGAGCTGCAG 3'  
 AGO2 Reverse Primer: 5' GAGTGTCTTGGTGAACCTGGAC 3'

To determine the siRNA knockdown efficiency of endogenous KIF18B by qPCR, siRNA treated cells were harvested 24 hr after transfection and RNA was isolated. To measure knockdown levels of endogenous AGO2 by CRISPRi, cells expressing the dCAS9-BFP-KRAB were infected with sgRNAs targeting AGO2 and harvested 4-5 days later to isolate RNA. RNA was isolated using TRIsure (Bio-line), according to manufacturer's guidelines. Next, cDNA was generated using Bioscript reverse transcriptase (Bioline) and random hexamer primers. qPCRs were performed using SYBR-Green Supermix (Bio-Rad) on a Bio-Rad Real-time PCR machines (CFX Connect Real-Time PCR Detection System). RNA levels were normalized to the levels of GAPDH mRNA.

### 3.5.15 Quantification of smFISH experiments

To quantify the number of mRNAs based on smFISH, multiple Z slices were made (with an interslice distance of 0.5  $\mu\text{m}$  each) and maximum intensity Z-projections were created. Dependent on whether we wished to quantify the number of mRNAs in the nucleus and cytoplasm (Supplemental Fig. 1b-f and Supplemental Fig. 2j), or tethered to the membrane (Supplemental Fig. 2j), maximum projections containing different slices were created. For measurements in the nucleus and cytoplasm, maximum projections which included all slices in which the nucleus was present (based on DAPI) were used (to prevent calling cytoplasmic mRNAs nuclear). To quantify the number of mRNAs tethered to the membrane, maximum projections of the two slices containing the bottom membrane were made. Using TransTrack, the nucleus was identified based on DAPI and the number of mRNAs at each location

was quantified.

To quantify the percentage of co-localization between the SunTag and PP7 smFISH probes, mRNAs were identified using TransTrack based on the SunTag probe signal (Cy5). For each mRNA we manually determined whether the SunTag smFISH signal co-localized with the PP7 signal (Alexa 594).

To quantify the transcription site intensities, maximum intensity Z-projections were created (images were taken with an interslice distance of 0.5  $\mu\text{m}$  each). To ensure that all the fluorescence signal of the transcription site was captured, the maximum intensity projections included all slices in which the nucleus was present (based on DAPI signal). To quantify the fluorescence intensity of the transcription site, an ROI was manually drawn around the transcription site, and the integrated fluorescence intensity was measured. For each spot the background fluorescence intensity was measured in the cytoplasm using a second ROI with the same dimensions. The background fluorescence intensity was subtracted from the transcription site fluorescence intensity.

### 3.5.16 Quantification of northern blots

Northern blot images were analyzed using ImageQuant TL. The total intensity of each band was measured and background was subtracted using the manual baseline option (i.e. the background intensity was measured manually). To control for loading differences, the RNA gel was analyzed and both the 18S rRNA and 28S rRNA integrated band intensity were measured. An average normalization factor was calculated based on the 18S rRNA and 28S rRNA integrated intensity and the northern blot band intensities were normalized accordingly.

### 3.5.17 Quantification of in vitro AGO2 on rate

RNA molecules were first localized in the reference movie through custom code written in IDL. Non-specific interactions of AGO2-siRNA complexes with the chamber surface (i.e. interactions that do not show co-localization with the RNA molecules) are ignored. Next, intensity time traces were created in the measurement movie for each RNA molecule (based on the positions of the RNA molecules in the reference movie) and the resulting intensity time traces were further processed in MATLAB (Mathworks) using custom code. To determine the binding rate, we measured the time between introduction of AGO2-siRNA complexes in the sample chamber and the time when stable binding occurred (stable binding is defined by interactions of  $> 1$  s).

The binding time of AGO2 was calculated as the time between AGO2-siRNA complex introduction into the imaging chamber and the first stable binding event for each RNA molecule. For the short RNA oligonucleotide target, the majority of molecules was bound by an AGO2 molecule within our time window of 350 s. To calculate the on-rate, we fit the data with Equation 3.1.

$$F(t) = A \cdot (1 - e^{-k_{\text{on}} \cdot t}) \quad (\text{Equation 3.1})$$

where  $F$  is the fraction of bound molecules,  $A$  the maximum bound fraction,  $t$  the time and  $k_{\text{on}}$  the on rate.

For the full length mRNA targets, most molecules were not bound by an AGO2 molecule within our time window of 350 s. Therefore, it was not possible to fit the data with Equation 3.1 and we instead linearized the equation resulting in Equation 3.2.

$$F(t) = A \cdot (1 - e^{-k_{\text{on}} \cdot t}) = A \cdot [1 - (1 - k_{\text{on}} \cdot t + (k_{\text{on}} \cdot t^2)/2! - \dots)] \sim A \cdot k_{\text{on}} \cdot t \quad (\text{Equation 3.2})$$

The approximation is valid as long as the product  $k_{\text{on}} \cdot t$  is very small. Using Equation 3.2, we fit the datapoints from the first hundred seconds to determine the value of  $A \cdot k_{\text{on}}$ . Next, we calculated the on rate ( $k_{\text{on}}$ ) by dividing with  $A$ . We took the value of  $A$  that we had fit with the short oligonucleotide target.

### 3.5.18 Scoring mRNA cleavage

Two types of mRNA cleavage events could be observed in our live-cell imaging assay: (1) cleavage events where GFP and mCherry foci separated completely, which represents mRNAs in which all ribosomes were present on the 5' cleavage fragment (for example, because the first translating ribosome had not yet passed the cleavage site) and (2) cleavage events where one or more ribosomes were present on the 3' cleavage fragment upon cleavage, resulting in GFP foci that separated from dual-color GFP-mCherry foci. In the latter case, at least one ribosome must have already passed the cleavage site at the moment of cleavage.

For mRNAs that did not show cleavage, we tracked the total time that the translation signal associated with the mRNA could be observed. A track was ended either when: (1) translation of the mRNA could no longer be observed, (2) the end of the time lapse movie was reached, (3) when the mRNA moved out of the field of view, (4) when a mRNA would cross paths with a fluorescent lysosome or another mRNA, or (5) when the mRNA detached from the plasma membrane.

### 3.5.19 Determining the moment of mRNA cleavage

To quantify the precise moment of cleavage of reporter mRNA molecules relative to the start of translation, we determined the first time point at which we observed GFP signal for each mRNA, as well as the moment where we observed spatial separation of the GFP and mCherry foci (which was scored as cleavage). If no cleavage was observed for an mRNA during the course of the experiment, we tracked the total time that the translation signal on the mRNA could be observed (see section 'scoring mRNA cleavage'). To ensure that we had observed the first round of translation on each mRNA molecule, we excluded mRNAs that (1) were already present in the field of view at the start of the movie, (2) mRNAs that were already associated with GFP fluorescence when they first appeared in the field of view, or (3) mRNAs that were never associated with GFP fluorescence. We plotted the fraction of uncleaved mRNAs using a Kaplan-Meier curve (which we refer to as a cleavage curve), which takes into account both the track length of the cleaved and uncleaved mRNAs.

### 3.5.20 Distinguishing mRNA cleavage from translation termination

In our assay, spatial separation of GFP and mCherry foci was scored as mRNA cleavage. However, separation of GFP foci from mRNAs (mCherry foci) could also represent translation termination of a ribosome and release of the (GFP-positive) nascent polypeptide. A key difference between these two processes is that for translation termination the GFP foci that are separating from the mRNA represent a single ribosome, while for mRNA cleavage the GFP foci represent the 5' mRNA cleavage fragment, which generally contains multiple ribosomes and thus is associated with a higher GFP fluorescence intensity. To demonstrate that separation of GFP and mCherry foci represented mRNA cleavage rather than translation termination, we quantified the intensity of the GFP foci that separated from the mCherry foci. We compared the fluorescence intensity of the GFP foci to the intensity of a single mature protein (which is similar to the fluorescence intensity that is associated with a single ribosome at the moment of translation termination) to establish whether GFP foci after cleavage represented a single ribosome (as expected for translation termination) or multiple ribosomes (as expected for mRNA cleavage). To visualize single mature proteins, we imaged with 100% laser power and 20 ms exposure. Next, to determine the fluorescence intensity of GFP foci that separated from mCherry foci and single mature proteins, GFP intensities were measured in FIJI in an ROI with a width and height

of 9 pixels (1.09  $\mu\text{m}$ ). For each GFP spot (both GFP foci after separation from mCherry foci and single mature proteins), local background fluorescence intensity was measured using a second ROI at multiple locations directly next to the spot of interest, and the mean background fluorescence intensity was subtracted from the mean spot intensity. Mean intensities of single mature proteins were then compared to GFP foci that separated from mCherry foci.

### 3.5.21 Analysis of mRNA cleavage with the EMI1-KIF18B reporter

To analyze mRNA cleavage on mRNAs translated by a single ribosome, we introduced a 5' UTR sequence of the EMI1 gene in our reporter. This isoform of the 5' UTR of EMI1 has been shown to act highly repressive on translation initiation, frequently limiting the number of ribosomes per mRNA molecule to one ribosome [37,44]. To ensure that exclusively translation events with single ribosomes were analyzed, translation events were selected on two criteria: (1) the GFP intensity correlated with the intensity of a single ribosome and (2) the time between GFP appearance and disappearance was <7.5 minutes, as longer events are more likely to represent either translation events by multiple ribosomes or stalled ribosomes. Because translation termination of a single ribosome cannot be readily distinguished from mRNA cleavage, as they both result in the separation of a single ribosome (SunTag array) from the mRNA, we used mRNA disappearance (i.e. decay of the 3' cleavage fragment) as a read-out for mRNA cleavage [43]. Therefore, we determined the time from GFP appearance to mRNA disappearance for each mRNA molecule.

### 3.5.22 Analysis of 'late' cleavage events in CHX-treated cells

We used different datasets in which cleavage was scored after CHX treatment to assess the mRNA refolding rate (Fig. 6a, and Supplemental Fig. 6b–d). We found that CHX addition reduces the cleavage rate, likely due to refolding of the mRNA and masking of the target site. However, a delay is expected between CHX addition and the decrease in the cleavage rate, which represents the time needed for mRNA re-folding. Thus, the transition time from a fast to a slow cleavage rate reports on the mRNA refolding rate. To assess the mRNA refolding rate, we analyzed the cleavage rates for CHX treated cells in more detail. In our initial analyses all translating mRNAs were included. However, in the current analysis we wanted to only include mRNAs on which the first ribosome had recently translated (and unfolded) the siRNA target site to omit mRNAs on which the first ribosome had not reached the siRNA target site yet (as the target site had not been unmasked on these mRNAs yet), and mRNAs on which the first ribosome has passed the siRNA target site long ago (as these are likely enriched in the small subset of mRNAs that appear resistant to siRNA-mediated cleavage). To determine the time window during which the first ribosome recently translated the siRNA target site we used the first ribosome arrival distribution at the siRNA target site (see methods 3.5.25) and we selected for each reporter the time points during which on 85–90% of the mRNAs the first ribosome had arrived at the siRNA target site. For KIF18B siRNA #1 this included mRNAs on which translation had started between 3.5 and 6.5 minutes before CHX addition. For KIF18B siRNA #2 this included mRNAs on which translation had started between 2.5 and 5.0 minutes before CHX addition. For GAPDH siRNA #1 this included mRNAs on which translation had started between 2.0 and 4.5 minutes before CHX addition. For GAPDH siRNA #3 this included mRNAs on which translation had started between 2.0 and 5.0 minutes before CHX addition. Next, we fit the resulting cleavage curves with a double exponential function to assess the transition time from a fast to a slow cleavage rate. For 1/4 reporters (GAPDH reporter with GAPDH siRNA #3) the data did not fit well with a double exponential function and was omitted from further analysis.

We note that a complicating factor in these experiments is, that for a subset of mRNAs, the sequence surrounding the target site may re-fold rapidly into a structure in which the AGO2 binding site is continuously unmasked, and as such these mRNAs may also be cleaved during the fast phase of cleavage.

### 3.5.23 Quantification of mRNA cleavage rates

From the Kaplan–Meier cleavage curves (referred to as cleavage curves) we quantified the mRNA cleavage rates for different reporters in the presence or absence of ribosomes translating the siRNA binding site (Fig. 2d, Fig. 5c, Fig. 6g, Supplemental Fig. 3c, Supplemental Fig. 4g, Supplemental Fig. 5f, and Supplemental Fig. 6h). To determine the cleavage rate in the absence of ribosomes translating the siRNA binding site (e.g. experiments in which CHX was added to cells, or in which a stop codon was introduced upstream of the AGO2 binding site), we fit the cleavage curve (of each individual replicate experiment) with a single exponential function. A plateau was included in the fit, as we frequently observed that a small fraction of mRNAs was resistant to cleavage (e.g. see Fig. 5b and Equation 3.3).

$$\text{Fraction uncleaved (t)} = (1 - c) \cdot e^{-bt} + c \quad (\text{Equation 3.3})$$

Here,  $b$  is the cleavage rate,  $c$  is the plateau level, and  $t$  is the time in seconds after the first detection of GFP on an mRNA molecule (i.e. the start of translation). We fixed the plateau level for all siRNAs at the plateau level observed for GAPDH siRNA #3 in the GAPDH  $\Delta 4$ -mer reporter (~15%) (see Fig. 5b).

To determine the cleavage rate in the presence of ribosomes translating the siRNA target site, we could not fit the cleavage curve with a single exponential function, because the cleavage curve is more complex. Initially the cleavage curve decreases slowly, followed by a faster decrease from the moment the first ribosome arrives at the siRNA target site, and a final slower decrease of the last 10–20% of mRNAs. Therefore, instead of fitting with a single exponential function, we determined the cleavage rate at individual time points using Equation 3.4.

$$\text{Cleavage rate (t)} = \ln(\text{fraction uncleaved(t)}/(\text{fraction uncleaved (t-1)})) \quad (\text{Equation 3.4})$$

Using Equation 3.4, the cleavage rate can be computed at each time point and we averaged the cleavage rate over a specific time window (see below) to determine the cleavage rate in the presence of ribosomes translating the siRNA target site. The start of this time window was set at the time point when on 50% of the mRNAs the first ribosome has reached the siRNA target site. This time point was determined for each individual siRNA and reporter based on the ribosome arrival distribution (see methods 3.5.25). The end of the abovementioned time window was set at the time point when the fraction of uncleaved mRNAs reached 20% (or at the end of the cleavage curve in case the 20% is never reached), because we observed that in most experiments the last 10–20% of mRNAs is cleaved at a lower rate, for reasons that are currently unclear. Statistical significance of differences between cleavage rates were tested using a two-tailed Student's  $t$ -test.

### 3.5.24 Number of ribosomes per mRNA molecule

We determined the number of ribosomes per mRNA molecule (ribosome occupancy) for the uORF-KIF18B and KIF18B reporters. First, we quantified the ribosome occupancy on the uORF-KIF18B reporter. Next, we determined the fold difference in ribosome occupancy between the uORF-KIF18B and KIF18B reporter and used the ribosome occupancy on the uORF-KIF18B reporter to determine the ribosome occupancy on the KIF18B reporter. To determine the ribosome occupancy on a uORF-KIF18B reporter mRNA, we divided the GFP intensity of a translation site by the average intensity of a single mature protein (as the intensity of a single mature protein is equal to the fluorescence intensity associated with a single ribosome that translated the entire SunTag). Next, we multiplied the normalized GFP intensity of a translation site with a correction factor of 1.32 [37] to account for those ribosomes that have not completely translated the SunTag yet and will thus have a lower intensity compared to a single mature protein. To determine the GFP fluorescence intensity of single translation sites and single mature proteins, cells containing the uORF-KIF18B reporter were imaged using very short (20 ms) exposure time and maximal laser power (instead of 500 ms exposure time with low laser power, which

was used for the majority of experiments), since at long exposure times single mature proteins cannot be visualized due to motion blurring. Next, the GFP fluorescence intensity of single mature proteins and translation sites was measured in FIJI in a circular region of interest (ROI) with a width and height of 6 pixels (0.81  $\mu\text{m}$ ). For each measurement, local background fluorescence intensity was measured using a second ROI with the same dimensions at a location directly next to the spot of interest. Finally, the mean background fluorescence intensity was subtracted from the mean spot intensity to obtain the fluorescence intensity of single mature proteins and translation sites.

To obtain the fold difference in ribosome occupancy between the uORF-KIF18B and KIF18B reporters, we divided the average GFP fluorescence intensity of translation sites of the KIF18B reporter by the average GFP fluorescence intensity of translation sites of the uORF-KIF18B reporter. To determine the GFP fluorescence intensity of single translation sites, cells containing either the uORF-KIF18B reporter or the KIF18B reporter were imaged using 500 ms exposure time and low laser power (as we did not image single mature proteins here, the standard imaging conditions could be used). After this, the GFP fluorescence intensity of single translation sites was measured in FIJI in a circular ROI with a width and height of 10 pixels (1.35  $\mu\text{m}$ ). For each measurement, local background fluorescence intensity was measured using a second ROI with the same dimensions at a location directly next to the spot of interest. Next, the mean background fluorescence intensity was subtracted from the mean spot intensity to obtain the fluorescence intensity of single translation sites. Finally, the average GFP fluorescence intensity of translation sites of the KIF18B reporter was divided by the average GFP fluorescence intensity of translation sites of the uORF-KIF18B reporter and the previously determined ribosome occupancy for the uORF-KIF18B reporter was used to calculate the average ribosome occupancy for the KIF18B reporter.

### 3.5.25 Arrival time of the first ribosome at the siRNA binding site

For many reporter-siRNA pairs we observed that the probability that an mRNA is cleaved was not constant over time. Rather, we found a peak in the cleavage probability which occurred several minutes after the start of translation (represented by a strong decrease in the fraction of uncleaved mRNAs during this time window). This led us to hypothesize that the increased probability for mRNA cleavage may coincide with the time at which the first ribosome reaches the siRNA target site. To test this hypothesis, we wanted to precisely determine the time at which the first ribosome arrives at the siRNA target site for different reporter mRNAs. The time at which the first ribosome arrives at the siRNA target site could be variable between mRNAs as different ribosomes might move at different speeds [37]. Therefore, we wanted to determine not only the average time at which the first ribosome arrives at the siRNA target site but also the distribution of the arrival times of the first ribosome (this will be referred to as the first ribosome arrival time distribution). To determine the first ribosome arrival time distribution at the siRNA target site, we first determined the arrival time distribution at the stop codon and used these values to extrapolate the ribosome arrival time distribution at the siRNA target site (which is located upstream of the stop codon). We assumed that translation elongation speed is constant along the mRNA, and thus that the first ribosome arrival time distribution at the stop codon can be extrapolated to the siRNA target site (which is located upstream of the stop codon) based on the difference in length of the mRNA sequence that is translated.

We calculated the first ribosome arrival time distribution at the stop codon by measuring the fluorescence intensity build-up over time of newly translated mRNAs. The fluorescence intensity build-up on a single mRNA can be divided into two phases, which we will refer to as the increasing phase and the plateau phase, respectively. In the first phase, the fluorescence intensity increases, because an increasing number of ribosomes translates the mRNA and synthesizes SunTag peptides that are rapidly fluorescently labeled by the SunTag antibody. The fluorescence intensity of the translation site increases until the first ribosome reaches the stop codon, and remains constant afterwards (which is the start of

the plateau phase). The fluorescence intensity remains constant after the first ribosome reaches the stop codon, because at that point translation of the mRNA is in steady state: new ribosomes will initiate translation (which will cause an increase in fluorescence intensity), while at the same time ribosomes will terminate translation when they reach the stop codon (which will lead to a decrease in fluorescence intensity as the nascent polypeptide chain with the SunTag peptides is released from the ribosome and diffuses away from the mRNA). Therefore, on average, a constant number of ribosomes will translate the mRNA. Since the transition between the two phases occurs when the first ribosome reaches the stop codon, the time of the transition point reports on the arrival time of the first ribosome at the stop codon. Thus, to determine the distribution of arrival times of the first ribosome at the stop codon, we have to (1) measure fluorescence intensity build-up on single mRNAs and (2) determine for each mRNA the transition time point between the increasing and plateau phases.

To measure the fluorescence intensities over time on single mRNAs, we used a Matlab-based software package called ‘TransTrack’ that we have recently developed [71]. Using TransTrack, we tracked single mRNAs from the start of translation, measured the GFP fluorescence intensity of each translation site and performed local background subtraction. All traces were manually curated and incorrect traces were discarded. In order to determine the transition time point between the increasing phase and plateau phase, we first determined the slope of each fluorescence intensity trace at all time points by calculating the derivative of the intensity trace. The slope of the fluorescence intensity trace also contains two phases that represent the increasing and plateau phase: (1) a phase in which the derivative is positive (which represents the increasing phase); (2) a phase in which the derivative is around zero (which represents the plateau phase).

To find the most likely transition time point between the increasing phase and plateau phase, we calculated a probability score for each hypothetical transition time point and selected the transition time point with the highest probability. To calculate a probability score, we assessed the probability of observing the derivative intensity trace for each hypothetical transition time point: we expect to observe only positive derivative values before the hypothetical transition time point (as this is during the increasing phase), while we expect to observe both positive and negative derivative values after the hypothetical transition time point (as this is during the plateau phase). To calculate the probability score we first determined an expected distribution of the derivative during the increasing phase and an expected distribution of the derivative during the plateau phase. To obtain an expected increasing phase slope distribution, we only took the slopes from all traces at early time points (< 3.5 minutes after the start of translation) to ensure that only values were used that represented mRNAs for which the first ribosome had not yet reached the stop codon. Next, we obtained an expected plateau phase slope distribution by only taking the slopes from all traces at late time points, to ensure that on all mRNAs the first ribosome has already reached the stop codon. Next, we fit the increasing phase slope and the plateau phase slope distributions to a normal distribution with fitting parameters  $\mu$  and  $\sigma$ , which resulted in good fits.

Next, we used both the increasing and plateau phase slope distributions to calculate the probability score for the slopes observed before the hypothetical transition time point (using the expected distribution for the increasing phase) and the slopes observed after the hypothetical transition time point (using the expected distribution for the plateau phase). A total probability score was calculated by adding the probability scores before and after the hypothetical transition time point together. The total probability score was iteratively calculated for all hypothetical transition time points along the fluorescence intensity trace and the transition time point with the highest probability was selected. Finally, a threshold probability score was set to exclude traces that did not result in a good fit (10–20% of all traces). In this way we determined for all GFP fluorescence intensity traces the most likely transition point between the increasing phase and plateau phase, which marks the arrival of the first ribosome at the stop codon.

From the arrival time of the first ribosome at the stop codon on all mRNAs, we generated the ‘first ribosome arrival time’ distribution and we fit the distribution to a gamma distribution with fitting parameters  $k$  (shape) and  $\theta$  (scale). We used a gamma distribution as it fit the data well and only contains non-negative values (in contrast to a normal distribution).

The analysis described above was performed for three different reporters: KIF18B, KIF18B-early-stop, and GAPDH. This resulted in three gamma distributions for the arrival time of the first ribosome at the stop codon for the respective reporters. Next, the arrival distribution at the different siRNA target sites was determined by scaling the shape parameter of these distributions linearly to the effective mRNA length while keeping the scale parameter of these distributions the same. The effective mRNA length was calculated by determining the distance between the start codon and either the stop codon or the siRNA target site and subtracting a length of 753 nt, which corrects for the fact that the first ribosome is not detectable until approximately 8 SunTag peptides (which have a length of 753 nt) have been translated. For siRNAs targeting the KIF18B reporter (siRNA #1, #2, and #3) the average ribosome arrival time at the stop codon of both the KIF18B as well as the KIF18B-early-stop was used to extrapolate the ribosome arrival time at the siRNA target sites, an average of both reporters was used to increase the reliability. For siRNAs targeting the GAPDH reporter (siRNA #1, #2, and #3) the average ribosome arrival time at the stop codon of the GAPDH reporter was used to extrapolate to the siRNA target sites. For siRNAs targeting the TPI1 reporter (siRNA #1, and #2) the average ribosome arrival time at the stop codon of all three reporters (KIF18B, KIF18B-early-stop, and GAPDH) was used to extrapolate to the siRNA target sites, and the average of the three reporters was used to increase the reliability.

### 3.5.26 Effects of siRNAs on translation efficiency

To determine the effects of siRNAs on translation dynamics, we analyzed the GFP fluorescence intensity on the KIF18B reporter in the presence or absence of KIF18B siRNA #1 (Supplemental Fig. 1i,j). The GFP fluorescence intensity on newly transcribed mRNAs increases from the start of translation as multiple ribosomes sequentially initiate translation on an mRNA molecule and sequentially synthesize all 24 SunTag peptides that are rapidly fluorescently labeled with the SunTag antibody. The slope of the increasing fluorescence intensity of single translation sites depends both on the translation initiation and elongation rate. The higher the translation initiation rate, the more ribosomes will simultaneously synthesize SunTag peptides, and thus the faster the increase in the GFP fluorescence will be. Similarly, the higher the translation elongation rate, the faster ribosomes synthesize SunTag peptides. Therefore, by comparing the slope of the fluorescence intensity increase in the presence and absence of siRNA, we can assess whether the siRNA affects the translation initiation and/or elongation rate. To obtain the slope of the fluorescence intensity increase, we tracked single mRNAs in the presence and absence of siRNA, measured the GFP fluorescence intensity, and performed local background subtraction using TransTrack [71]. All traces were manually curated. Next, the slope of the fluorescence intensity increase was calculated by determining the derivative of the GFP fluorescence intensity traces.

To compare the slope of the fluorescence intensity in the presence and absence of siRNA two additional considerations have to be taken in account. First, the fluorescence intensity increase on a single mRNA is not constant over time, but rather contains three phases with distinct slopes. At the start of translation the slope of the fluorescence intensity increases, because at first only a single ribosome is synthesizing SunTag peptides. However, at later time points multiple ribosomes are simultaneously synthesizing SunTag peptides, resulting in a larger increase of the fluorescence intensity per time unit. The maximal slope of the fluorescence intensity is achieved when the first ribosome reaches the end of the SunTag array, because from this time point on the maximal number of ribosomes is simultaneously synthesizing SunTag peptides. Thus, the first phase is defined as the time from translation initiation by the first ribosome, until the time when the first ribosome reaches the end of the SunTag array. In the second



phase, the fluorescence intensity of the translation site will increase at a constant rate (i.e. the maximal rate) until the first ribosome reaches the stop codon. The third phase starts once the first ribosome has reached the stop codon, because from that time point on the increase in the fluorescence intensity by newly initiating ribosomes is offset by a decrease in the fluorescence intensity due to terminating ribosomes. This results in a net slope of zero. To compare the slope of the fluorescence intensities in the presence and absence of siRNA, we focused on the second phase where the fluorescence intensity increases linearly with a maximum slope. Since the slope during the second phase remains constant, this allowed us to average multiple time points to create a larger dataset. A time window between 1.5 and 4 minutes after the start of translation was selected based on visual inspection of the fluorescence intensity graph (as during this time window a linear increase was observed).

The second consideration to take into account when comparing the slopes of the fluorescence intensity traces in the presence and absence of siRNA, is that in the presence of siRNA it is unknown at which timepoint mRNAs have an AGO2–siRNA complex bound, as the AGO2–siRNA complex is not fluorescently labeled itself. We would not expect an effect on the translation dynamics for mRNAs that don't have an AGO2–siRNA complex bound. Therefore, we only examined translation efficiency immediately before cleavage was observed, since AGO2–siRNA complex must have been bound to the mRNA at the moment of cleavage. To determine the translation efficiency immediately before cleavage, we only calculated the slope of the GFP intensity curve between the two time points preceding the moment of cleavage, and grouped mRNAs together that were cleaved in the time window between 1.5 and 4 minutes after the start of translation (see methods 3.5.19 for a description on the quantification of the cleavage time) to average the slope of multiple mRNA molecules. The average slope of the fluorescence intensity increase in siRNA treated cells and control cells was compared using a paired two-tailed t-test (Supplemental Fig. 1k).

### 3.5.27 Time between CHX treatment and inhibition of ribosome elongation

To determine how long after addition of CHX ribosomes are completely immobilized, we analyzed GFP intensity build-up on mRNAs on which translation started 1 to 3 minutes before the addition of CHX. In the absence of CHX we expect a constant increase in GFP intensity during this phase (see methods 3.5.25). However, when CHX is active, translation elongation will be inhibited, so GFP intensity will no longer increase. To determine the GFP fluorescence intensity of translation sites, the mean GFP intensity of translation sites was measured in FIJI in a circular region of interest (ROI) with a width and height of 9 pixels (1.09  $\mu\text{m}$ ). For each spot, local background fluorescence intensity was measured using the same ROI at multiple locations directly next to the spot of interest, and the mean background fluorescence intensity was subtracted from the mean spot intensity. GFP intensity was measured from the first frame of GFP appearance until 3–5 minutes after CHX addition. Intensity traces were aligned at the moment of CHX addition and the moment at which GFP no longer increased over time was determined.

### 3.5.28 Detecting ribosomes on the 3' cleavage fragment

To distinguish between the release model (ribosomes collide into AGO2, stimulating release of cleavage fragments) and the binding model (ribosomes unmask the siRNA target site, stimulating the binding of AGO2 to the target site), we quantified the fraction of cleavage events where one or more ribosomes were present on the 3' cleavage fragment. The binding model predicts that one or more ribosomes should be present on most 3' cleavage fragments, while the release model predicts that ribosomes should be mostly absent from the 3' cleavage fragment. Since the fluorescence intensity of a single ribosome is weak and the signal is only present on the 3' cleavage fragment for a very short amount of time (as the length between the siRNA target site, where the mRNA is cleaved, and the stop codon is relatively short for most reporters), we modified our assay in two ways to allow efficient detection of ribosomes on the 3' cleavage fragment.

The first modification was made to the reporters. We made two new reporter constructs where a linker sequence was added between the siRNA binding site and the stop codon to increase the time a ribosome remains on the 3' cleavage fragment. The linker sequence facilitates the detection of the ribosome on the 3' cleavage fragment as it takes longer for the ribosome to reach the stop codon after passing the siRNA target site. For the KIF18B reporter, we inserted the sequence of the fluorescent protein mRuby (714 nt) as a linker; for the GAPDH reporter, we inserted the KIF18B sequence (2568 nt) as a linker (the KIF18B sequence could not be used as a linker in the KIF18B reporter as this would duplicate the siRNA target site). Introduction of these linker sequences in the KIF18B and GAPDH reporters did not substantially alter cleavage kinetics (Supplemental Fig. 2k).

The second modification to the assay was made to the image acquisition settings. We acquired images every 15 s instead of every 30 s to more efficiently detect GFP signals only present for a short amount of time on the 3' cleavage fragment, and we imaged with high laser power to detect weak signals associated with only a single ribosome. With these two modifications we quantified for each mRNA the time of cleavage relative to the start of translation (as described in methods 3.5.19) and scored for each cleavage event if there was a ribosome present on the 3' cleavage fragment (i.e. whether a GFP signal was associated with the 3' fragment after cleavage). Finally, we grouped mRNAs together that were cleaved at the same time point relative to the start of translation and calculated for each time point the fraction of cleavage events with a ribosome on the 3' cleavage fragment.

### **3.5.29 Normalization of the fraction of mRNAs with ribosomes on the 3' cleavage fragment**

To distinguish between the release and binding models, we performed two processing steps on the fraction of cleavage events with a ribosome on the 3' cleavage fragment.

First, we selected cleavage events that occurred around the time when the first ribosome reached the siRNA target site. Cleavage events that occurred before the first ribosome reached the siRNA target site (early cleavage events) or cleavage events that occurred when a second or subsequent ribosome reached the siRNA target site (late cleavage events) are not informative to distinguish between abovementioned models, because in both models none of the early cleavage events are predicted to have a ribosome present on the 3' cleavage fragment (as the first ribosome hasn't reached the siRNA target site yet), while all of the late cleavage events will have a ribosome present on the 3' cleavage fragment (as the first ribosome has already passed the siRNA target site). To select cleavage events that occurred around the time of the first ribosome arrival at the cleavage site, we used the first ribosome arrival distribution (see methods 3.5.25) to select the three time points when the first ribosome was most likely to reach the siRNA target site.

As reporter mRNAs can also be cleaved in the absence of ribosomes translating the siRNA target site (e.g. + CHX experiments), and ribosome-independent cleavage may even occur when a ribosome is close to the cleavage site, it is likely that not all cleavage events that occur in the time period when ribosomes arrive at the cleavage site are ribosome-dependent cleavage events (even in the absence of ribosomes translating the siRNA target site (CHX experiments) cleavage is observed during the same time window). Ribosome-independent cleavage events are not informative to distinguish between the models, because in all models the same fraction of ribosome-independent cleavage events will have a ribosome present on the 3' cleavage fragment. Therefore, during the second processing step we wanted to correct for ribosome-independent cleavage by calculating the fraction of ribosome-dependent cleavage events for each time point, and assessing which fraction of ribosome-dependent cleavage events contained a ribosome on the 3' cleavage fragment. To estimate the fraction of ribosome-dependent cleavage events with a ribosome on the 3' cleavage fragment at each time point relative to the start of translation, we used Equation 3.5.

$$S_{\text{Total}}(t) = F_{-}(t) \cdot S_{-}(t) + F_{+}(t) \cdot S_{+}(t) \quad (\text{Equation 3.5})$$

Here,  $t$  is the time relative to moment of GFP appearance (i.e. the start of translation),  $S_{\text{Total}}$  is the fraction of all cleavage events with a ribosome on the 3' cleavage fragment,  $F_{-}$  is the fraction of ribosome-independent cleavage events,  $F_{+}$  is the fraction of ribosome-dependent cleavage events, and  $S_{-}$  the fraction of ribosome-independent cleavage events with a ribosome on the 3' cleavage fragment.  $S_{+}$  is the fraction of ribosome-dependent cleavage events with a ribosome on the 3' cleavage fragment, and we calculated  $S_{+}$  for the three time points when the first ribosome was most likely to reach the siRNA target site. To calculate  $S_{+}$  with Equation 3.5, we first determined the other four parameters. As we have experimentally determined  $S_{\text{Total}}$  (the fraction of cleavage events with a ribosome on the 3' cleavage fragment), we only need to calculate  $F_{-}$ ,  $F_{+}$ , and  $S_{-}$ . To calculate the fraction of cleavage events that were either ribosome-independent ( $F_{-}$ ) or ribosome-dependent ( $F_{+}$ ) at each time point, we first calculated the rate of ribosome-independent cleavage using the data sets where cells were treated with CHX (+CHX datasets; as ribosomes are not translating the siRNA target site in the presence of CHX). We fit the +CHX datasets with a single exponential function, which was used as the rate of ribosome-independent cleavage. Next, to calculate the fraction of ribosome-dependent cleavage events, we subtracted the exponential function from the cleavage curve obtained in the absence of CHX. Finally, to calculate the fraction of ribosome-independent cleavage events with a ribosome on the 3' cleavage fragment ( $S_{-}$ ), we used the first ribosome arrival distribution at the siRNA target site. We expect that the fraction of ribosome-independent cleavage events with a ribosome on the 3' cleavage fragment for each time point is directly related to the ribosome arrival distribution (if at a given time point in 30% of the mRNAs the first ribosome has passed the siRNA target site, we would expect 30% of the ribosome-independent cleavage events to have a ribosome on the 3' cleavage fragment). Using the calculated values of  $F_{-}$ ,  $F_{+}$ , and  $S_{-}$ , and Equation 3.5, we calculated for the three time points (i.e. those with the highest probability of the first ribosome arriving at the cleavage site; see above) the expected fraction of ribosome-independent cleavage events with a ribosome on the 3' cleavage fragment ( $S_{-}$ ). Finally, we calculated a weighted average of  $S_{+}$  for the three time points by weighting each time point based on the total number of ribosome-dependent cleavage events at each time point.

### 3.5.30 Expectations from release and binding models

To determine whether the experimentally-observed fraction of mRNAs with a ribosome on the 3' cleavage fragment was most consistent with the release or binding model, we compared the fraction of ribosome-dependent cleavage events with a ribosome on the 3' cleavage fragment with the fraction that we would expect with the release and binding model. In the binding model we expect that a ribosome translates over the siRNA target site to make the target site available for AGO2, which will result in mRNA cleavage with a ribosome on the 3' cleavage fragment. Therefore, most ribosome-dependent cleavage events will lead to a cleavage with a ribosome on the 3' cleavage fragment (so we would expect close to 100% of the ribosome-dependent cleavage events to have a ribosome on the 3' cleavage fragment). Note that in some cases the siRNA binding site may interact with an upstream sequence in the mRNA to mask the target site. In these cases, translation of the upstream sequence by the ribosome may be sufficient to unmask the target site. If cleavage occurs very rapidly after unmasking, before the ribosome reaches the siRNA binding site, then the ribosome will not end up on the 3' cleavage fragment, even though it stimulated cleavage through the binding model. Thus, the binding model predicts that slightly less than 100% of cleavage events contain a ribosome on the 3' cleavage fragment.

In the release model we expect that a ribosome collides with AGO2 at the siRNA target site and stimulates AGO2 release from the 5' and 3' cleavage fragments. If the first ribosome translating an mRNA collides with AGO2 and stimulates release, no ribosomes will be present on the 3' cleavage fragment. However, it is also possible that the first (few) ribosome(s) translates the cleavage site, after which AGO2 binds and cleaves the mRNA, and that a subsequent ribosome collides with AGO2

to cause fragment release. In this case we do expect a ribosome on the 3' cleavage fragment. Thus, to determine the expected fraction of ribosome-dependent cleavage events with a ribosome on the 3' cleavage fragment for the collision model, we need to calculate the fraction of ribosome-dependent cleavage events that is caused by the first ribosome and by subsequent ribosomes.

To estimate the fraction of ribosome-dependent cleavage events caused by the first ribosome versus subsequent ribosomes, we need to calculate the fraction of mRNAs on which the first ribosome has passed the cleavage site at the moment of cleavage. For example, if on 30% of the mRNAs the first ribosome has passed the cleavage site, we would expect that 30% of the ribosome-dependent cleavage events are caused by a subsequent ribosome and will have a ribosome on the 3' cleavage fragment.

To calculate the fraction of mRNAs on which the first ribosome has passed the cleavage site, we made a correction to the first ribosome arrival distribution that was determined previously (see methods 3.5.25). From the first ribosome arrival distribution we know for each time point the mRNA fraction on which the first ribosome has passed the cleavage site. However, we also expect with the collision model that many mRNAs are cleaved when the first ribosome reaches the cleavage site. Thus, if we would expect, based on the ribosome arrival distribution, that on 30% of the remaining mRNAs the first ribosome has passed the cleavage site, the real fraction would be much lower as most of these mRNAs have been cleaved when the first ribosome reached the target site. To correct for mRNA cleavage when the first ribosome reaches the siRNA target site, we need to determine the fraction of mRNAs that is cleaved upon arrival of the first ribosome. To determine the fraction of mRNAs that is cleaved, we looked at the three time points when on most mRNAs the first ribosome reaches the cleavage site, and determined the amount of ribosome-dependent cleavage and divided this by the fraction of mRNAs where the first ribosome arrived during the three time points. Next, we used the fraction of mRNAs that is cleaved upon arrival of the first ribosome to correct the first ribosome arrival distribution; for example, if at a given time point on 30% of the mRNAs the first ribosome arrives at the cleavage site, but also 50% of these mRNAs are cleaved because of ribosome-AGO2 collisions, we expect that after this time point only 15% of these mRNAs are remaining. The first ribosome arrival distribution was corrected in this fashion, which resulted in a new distribution that reports for each time point on which fraction of mRNAs the first ribosome has already passed the cleavage site. As only ribosome-dependent cleavage events caused by subsequent ribosomes lead to a cleavage with a ribosome on the 3' cleavage fragment, we then calculated the expected fraction of ribosome-dependent cleavage events with a ribosome on the 3' cleavage fragment for the release model based on this corrected ribosome arrival distribution.

### 3.5.31 Calculating AGO2 cleavage cycle kinetics

Using the cleavage curves, we wished to estimate the average duration of the AGO2 cleavage cycle *in vivo* (see Fig. 3a-c). The AGO2 cleavage cycle consists of AGO2-siRNA target binding, mRNA slicing, and fragment release. In the cleavage curves only the moment of fragment release is quantified. To calculate the average duration of the AGO2 cleavage cycle, we have to determine the moment at which the target mRNA becomes accessible for AGO2-siRNA target binding, as target binding can only initiate when the AGO2 target site is unmasked. Target site unmasking can occur independently of ribosomes (e.g. through structural rearrangements) or through ribosomes translating and unfolding the AGO2 target site. For ribosome-dependent cleavage (i.e. unmasking through ribosomes unfolding the AGO2 target site), we can estimate the start of the cleavage cycle (defined here as the moment the target site becomes unmasked) based on the first ribosome arrival time distribution at the AGO2 target site. For ribosome-independent cleavage (i.e. unmasking through structural rearrangements) we cannot estimate the start of the cleavage cycle. Therefore, to estimate the average cleavage cycle duration we focused on ribosome-dependent cleavage events.

To estimate the duration of the AGO2 cleavage cycle, theoretical cleavage curves assuming different

durations of the AGO2 cleavage cycle were computed and compared to the experimentally-derived cleavage curves at different siRNA concentrations (10 nM, 0.75 nM, and 0.1 nM). The theoretical cleavage curve is composed of the sum of ribosome-independent cleavage events and ribosome-dependent cleavage events. To calculate the ribosome-independent cleavage curve, we fit the start of the experimentally-derived cleavage curves (first 3.5 minutes after the start of translation) with an exponential function. At these early time points the first ribosome has not reached the AGO2 target site yet, and therefore represent ribosome-independent cleavage. Next, we computed theoretical ribosome-dependent cleavage curves for different durations of the AGO2 cleavage cycle. As we plot the 'cleavage time' relative to GFP appearance (i.e. start of translation), the 'cleavage time' is the sum of (1) the time it takes for the first ribosome to reach the AGO2 target site (which will unmask the target site and start the cleavage cycle) and (2) the duration of the AGO2 cleavage cycle itself. For the arrival time of the first ribosome at the AGO2 target site we used the first ribosome arrival time distribution (see methods 3.5.25). For the duration of the AGO2 cleavage cycle, we used an exponential distribution with different mean cleavage cycle durations (ranging from 1 s to 30 min). Next, we convoluted the first ribosome arrival time distribution with the different cleavage cycle duration distributions to obtain ribosome-dependent cleavage curves. Finally, we added the ribosome-independent cleavage curve and ribosome-dependent cleavage curves together to obtain theoretical cleavage curves for different durations of the AGO2 cleavage cycle. To calculate the average cleavage cycle duration that best fit with the data (for 10 nM, 0.75 nM, and 0.1 nM), we calculated the sum of squared errors (SSE) between the theoretical cleavage curves (1 s to 30 min duration of the cleavage cycle) and the experimentally-derived cleavage curves (for 10 nM, 0.75 nM, and 0.1 nM).

### 3.5.32 Simulations of mRNA refolding rate

We found that ribosomes translate and unmask the siRNA target site by unfolding local mRNA structure, facilitating AGO2-siRNA complex binding and mRNA cleavage. After a ribosome unfolds the structure surrounding the siRNA site, the site is available for binding to AGO2-siRNA complexes until structures reform. Therefore, the more frequently an mRNA is translated by ribosomes, the more time the AGO2 target site will be in an unmasked state, which will result in a higher cleavage rate of the mRNA. However, the time that the AGO2 target site is in an unmasked state also depends on the mRNA refolding rate. In case of a very slow refolding rate, an siRNA target site that is unfolded by a ribosome, will not refold to its masked conformation before the next ribosome arrives at the siRNA target site. In this case, a higher translation rate will not increase the time the AGO2 target site will be in an unmasked state and thus does not lead to an increase in the cleavage rate. Therefore, the cleavage rate is dependent on both the translation rate (how often do ribosomes unfold the AGO2 target site) as well as the mRNA refolding rate (how fast do structures reform after a ribosome unfolds the AGO2 target site). To determine the refolding rate of our reporter mRNA, we therefore analyzed the difference in cleavage rates of two reporters with identical coding sequences and siRNA binding sites, but with different translation rates: the KIF18B reporter and the uORF-KIF18B reporter (which has a 3.3 fold lower translation rate; see methods 3.5.24). As we know the difference in translation rate between the two reporters, we can determine the expected difference in cleavage rate at different mRNA refolding rates and compare these values to the experimentally-observed difference in cleavage rate of the KIF18B and uORF-KIF18B reporters. To identify the mRNA refolding rate that matches the difference between the experimentally-derived cleavage rate of the KIF18B and uORF-KIF18B reporters best, we performed stochastic simulations to simulate cleavage curves of the KIF18B and uORF-KIF18B reporter at a range of different mRNA refolding rates. Next, we assessed at which mRNA refolding rate the difference between the simulated KIF18B and uORF-KIF18B cleavage curves matched the difference between the experimental cleavage curves by determining an SSE for the experimental cleavage curve with each simulated cleavage curve.

We simulated the cleavage curves for both the KIF18B reporter and the uORF-KIF18B reporter

at different siRNA concentrations (10 nM, 0.75 nM, and 0.1 nM). As the cleavage curve is based on the cleavage time of individual mRNAs, we simulated the cleavage time of a single mRNA, and repeated this for many mRNAs to obtain a simulated cleavage curve. To simulate the cleavage time of a single mRNA, we assumed that each mRNA could be cleaved either through; (1) a ribosome-dependent cleavage pathway: ribosome-facilitated AGO2 target binding through target site unmasking by ribosomes translating the siRNA target site; (2) a ribosome-independent cleavage pathway (this was included since mRNAs are also cleaved in the absence of ribosomes translating the siRNA target site). For each mRNA we therefore performed two parallel simulations: we simulated a cleavage time through the ribosome-independent cleavage pathway and through the ribosome-dependent cleavage pathway. The lowest of the two values was used as the cleavage time for that mRNA. To determine the cleavage time through the ribosome-independent pathway, a random number was selected from an exponential distribution that was generated by fitting the cleavage curve obtained in the presence of CHX.

To determine the cleavage time through the ribosome-dependent pathway, we simulated (1) when ribosomes arrive at the siRNA target site, and (2) for each ribosome if cleavage would occur during the time that the target site was unmasked by that ribosome. To simulate when ribosomes arrive at the siRNA target site we distinguished between the first and subsequent ribosomes. The arrival time of the first ribosome at the siRNA target site was simulated by selecting a random number from a gamma distribution, where the shape and scale parameter were taken from the first ribosome arrival distribution at the siRNA target site that we determined previously (see methods 3.5.25). In order to determine the arrival time of the next ribosome at the siRNA target site, we added together two values; (1) a fixed, minimum inter-ribosome arrival time, which was introduced, because each ribosome occupies multiple nucleotides on an mRNA and therefore a minimal nucleotide distance is always present between the centers of two ribosomes; (2) the time between two subsequent translation initiation events. In the simulations we assumed that the minimal distance between two ribosomes is 30 nucleotides (about the size of one ribosome footprint), and we converted the minimal inter-ribosome distance to a minimal inter-ribosome arrival time by dividing the minimal inter-ribosome distance with the average elongation speed of ribosomes on the KIF18B reporter (as was determined from the first ribosome arrival distribution). To simulate the time between two initiation events, we drew a number from an exponential distribution, which was based on the experimentally observed initiation rates of both the KIF18B and uORF-KIF18B reporters.

Next, we determined for each ribosome arriving at the siRNA target site if cleavage would occur during the time that the target site was unmasked by that ribosome; we assume that each ribosome unfolds the siRNA target site when the ribosome passes the site, which can lead to one of three possible outcomes: (1) the mRNA refolds (and will be opened again when the next ribosome arrives); (2) AGO2 binds and cleaves the mRNA; or (3) the mRNA remains open until the next ribosome appears at the site without getting cleaved.

We simulated the time for mRNA refolding and the time for AGO2-mediated mRNA cleavage (both of which are unknown) by drawing two random numbers. (1) The mRNA refolding time was drawn from an exponential distribution. The simulation is run multiple times with different values for the mean of the mRNA refolding time distribution (37 times with a range of 0.1 s to 1000 s). (2) The mRNA cleavage time was drawn from an exponential distribution. The mean of this distribution is the average mRNA cleavage time (which includes the time for AGO2 target binding, mRNA slicing, and fragment release). Simulations were performed using multiple distributions with distinct mean values and the simulation that resulted in the best fit with the data was selected and the mean of the distribution was recorded. The best fit was selected for each simulated mRNA refolding time. (3) The time until the next ribosome arrives at the siRNA binding site is known as we already determined the arrival time of all the ribosomes at the siRNA target site.

To simulate the outcome of a ribosome-dependent mRNA unfolding event (i.e. next ribosome arrives, mRNA refolds, or AGO2 binds and cleaves the mRNA), we selected the simulated process which was the fastest. Finally, the cleavage time through the ribosome-dependent pathway was determined by taking the arrival time of the ribosome after which AGO2-mediated mRNA cleavage occurred, and adding the AGO2 cleavage time. Using the simulations described above, we can simulate cleavage curves for the KIF18B and uORF-KIF18B reporters at different mRNA refolding rates. The values of the parameters used in the simulations can be found in the overview below.

Parameter	Value
Cleavage rate (ribosome-independent pathway)	10 nM control: $6.6396 \cdot 10^{-4} \text{ s}^{-1}$ , 10 nM uORF: $6.1039 \cdot 10^{-4} \text{ s}^{-1}$ 0.75 nM control: $7.4634 \cdot 10^{-4} \text{ s}^{-1}$ , 0.75 nM uORF: $6.9877 \cdot 10^{-4} \text{ s}^{-1}$ , 0.1 nM control: $3.8190 \cdot 10^{-4} \text{ s}^{-1}$ , and 0.1 nM uORF: $3.3234 \cdot 10^{-4} \text{ s}^{-1}$
First ribosome arrival time distribution at the siRNA target site (gamma distribution)	Shape: 15.34 and scale: 18.42.
Inter-ribosome distance	30 nt
Ribosome elongation speed	11.13 nt/second
Ribosome initiation rate	Control: $0.0469 \text{ s}^{-1}$ and uORF: $0.0135 \text{ s}^{-1}$

During the simulations we assumed that the mRNA refolding rate and the AGO2 cleavage rate are similar between the KIF18B and uORF-KIF18B reporter as both reporters have an identical coding sequence and siRNA target sequence. Next, we wanted to determine for each siRNA concentration (10 nM, 0.75 nM, and 0.1 nM) and each simulated mRNA refolding rate ( $k_{\text{refold}}$ ), the AGO2 cleavage rate ( $k_{\text{on}}$ ) that best fit the experimentally-observed difference between the cleavage curves of the KIF18B and uORF-KIF18B reporter. To find the AGO2 cleavage rate that best matched the data, we calculated a total sum of squared errors (total SSE) at each simulated mRNA refolding rate, which reports on how well the simulated cleavage curves matched the experimental cleavage curves for both the KIF18B and uORF-KIF18B reporters (see Equation 3.6).

$$\text{Total SSE}(k_{\text{refold}}, k_{\text{cleavage}}) = \sum_t [\text{Data}_{\text{Control}}(t) - \text{Simulations}_{\text{Control}}(t, k_{\text{refold}}, k_{\text{cleavage}})]^2 + [\text{Data}_{\text{uORF}}(t) - \text{Simulations}_{\text{uORF}}(t, k_{\text{refold}}, k_{\text{cleavage}})]^2 \quad (\text{Equation 3.6})$$

The total sum of squared errors (total SSE) is calculated by adding together the sum of squared errors of the KIF18B reporter and the sum of squared errors of the uORF-KIF18B reporter. The sum of squared errors of each reporter represents the difference between the experimentally-derived cleavage curve and the simulated cleavage curve at a specific mRNA refolding rate and AGO2 cleavage rate. Thus, the AGO2 cleavage rate that best fit with the experimental data at each simulated mRNA refolding rate can be found by minimizing the total SSE. To minimize the total SSE at each mRNA refolding rate, we simulated cleavage curves for a range of mRNA refolding rates (between 0.001 and  $10 \text{ s}^{-1}$ ), and at each mRNA refolding rate we searched for the AGO2 cleavage rate that minimized Equation 3.6. Each individual simulation consisted of 10.000 mRNAs, was performed in 10 replicates using a different random seed in each replicate, and all replicates were averaged. In this way, we obtained a total SSE value for each mRNA refolding rate.

To compare the total SSE values at different mRNA refolding rates, we first normalized the total SSE

value at each mRNA refolding rate with the lowest possible SSE value obtained when the experimental KIF18B and uORF-KIF18B cleavage curves were fit independently, at each mRNA refolding rate. To calculate this lowest possible SSE value we fit the KIF18B and uORF-KIF18B cleavage curves independently and calculated an independent SSE for both the KIF18B and uORF-KIF18B reporters using Equation 3.7. The lowest possible SSE value is the sum of both independent SSEs.

$$\text{Independent SSE}(k_{\text{refold}}, k_{\text{cleavage}}) = \sum_t [\text{Data}(t) - \text{Simulations}(t, k_{\text{refold}}, k_{\text{cleavage}})]^2 \quad (\text{Equation 3.7})$$

The AGO2 cleavage rate that best fit at each mRNA refolding rate with the experimental data for either the KIF18B or uORF-KIF18B cleavage curve can be found by minimizing the independent SSE. To minimize the independent SSE, we simulated cleavage curves for a range of mRNA refolding rates (between 0.001 and 10 s<sup>-1</sup>), and at each mRNA refolding rate we searched for the AGO2 cleavage rate that minimized Equation 3.7. After finding the independent SSE for both the KIF18B and uORF-KIF18B reporter at each mRNA refolding rate, we can normalize the total SSE using Equation 3.8.

$$\text{Normalized SSE}(k_{\text{refold}}) = \text{Total SSE}(k_{\text{refold}}) / [\text{SSE}_{\text{Control}}(k_{\text{refold}}) + \text{SSE}_{\text{uORF}}(k_{\text{refold}})] \quad (\text{Equation 3.8})$$

For each simulated mRNA refolding rate, we can normalize the total SSE by the sum of the KIF18B and uORF-KIF18B independent SSEs that were determined with Equation 3.7. As the sum of the KIF18B and uORF-KIF18B independent SSE represents the minimal total SSE possible, we obtain a normalized SSE.

### 3.5.33 Simulations of RNA structural dynamics

Our results showed that in the absence of ribosomes translating the siRNA target site, the cleavage rate is very similar at different siRNA concentrations. In contrast, lowering the siRNA concentration in the presence of ribosomes translating the siRNA target site strongly decreased the cleavage rate. These findings together suggest that in the presence of ribosomes translating the siRNA target site the AGO2-siRNA complex concentration is rate-limiting, while in the absence of ribosomes translating the siRNA target site, the availability of the siRNA target site can become another rate-limiting step (i.e. the target site is often in a masked state, unavailable for AGO2-siRNA complex binding). We can use the sensitivity of the cleavage rate to different siRNA concentrations in the absence of ribosomes translating the siRNA target site to probe the structural dynamics of the reporter mRNA; in case mRNAs change between different structural conformations on fast time scales (i.e. fast structural dynamics), it is expected that target site masking and unmasking also occur fast, as the siRNA target site is expected to be masked in only a subset of structural conformations. In this scenario, target site unmasking is not a rate-limiting step and the cleavage rate is expected to be solely determined and thus very sensitive to the siRNA concentration. In contrast, when mRNAs change between different structural conformations on slow time scales (i.e. similar or slower time scales than the AGO2-target binding), target site unmasking becomes rate-limiting as well. In this case the cleavage rate is expected to be less sensitive to the siRNA concentration as the cleavage rate is only partially determined by the siRNA concentration. Thus, to investigate the structural dynamics of our reporter mRNA (i.e. how fast does the siRNA target site change between a masked and unmasked state), we decided to simulate cleavage curves at different siRNA concentrations, and with different unmasking and masking rates. Specifically, we wanted to determine at which unmasking and masking rate, the effect of lowering siRNA concentration on the simulated cleavage curves best matched the experimentally-observed effect of lowering the siRNA concentration.

To determine the effect of lowering the siRNA concentration on the cleavage curve in the absence of ribosomes translating the siRNA target site, we simulated cleavage curves at different siRNA concentrations, and different unmasking and masking rates. To obtain a simulated cleavage curve, we



simulated the cleavage time of a single mRNA and repeated this many times. To simulate the cleavage time of a single mRNA, we set-up a simulation where an mRNA can be in three different states; in the first two states the mRNA is intact and either (1) in a masked state, unavailable for AGO2-siRNA complex binding and mRNA cleavage, or (2) in an unmasked state, available for AGO2-siRNA complex binding and mRNA cleavage. In the third state the mRNA is in (3) a cleaved state. As an mRNA is only available for AGO2-siRNA complex binding in the unmasked state, transitions can only occur from the unmasked to cleaved state, and transitions can occur between the unmasked and masked states. In the simulations, each mRNA starts either in the masked or unmasked state and transitions between the different states are simulated until the mRNA reaches the cleaved state. The time it takes in the simulations to reach the cleaved state is the cleavage time for that mRNA. To simulate the transitions and the transition times (how long does it take to transition from one state to the other), random numbers are selected from distributions, as described below.

**Masked to unmasked transition.** The transition time from the masked to unmasked state is referred to as the unmasking time. The mean unmasking time is unknown and the simulation is run multiple times with different values for the mean unmasking time (16 times with a range of 1 s to 1500 s for the mean unmasking time). The unmasking time is simulated by selecting a random number from an exponential distribution where the mean is the mean unmasking time.

**Transition from the unmasked to either the masked or cleaved state.** When an mRNA is in the unmasked state, it can transition either to the masked state or to the cleaved state. The transition time from the unmasked to the masked state is referred to as the masking time and the transition time from the unmasked to cleaved state is referred to as the cleavage time. To determine if an mRNA transitions to the masked or cleaved state, we simulate in parallel two transition times. (1) A masking time was simulated. The mean masking time is unknown and simulations were performed using multiple exponential distributions with distinct mean values. The simulation that resulted in the best fit with the data was selected, and the mean of the distribution was recorded. The masking time is simulated by selecting a random number from an exponential distribution where the mean is the mean masking time. (2) The AGO2 cleavage time, which is simulated by selecting a random number from an exponential distribution where the mean is the experimentally-derived average AGO2 cleavage time (which includes AGO2 target binding, mRNA slicing, and fragment release). The average AGO2 cleavage time (i.e. the AGO2 cleavage cycle duration) was already determined before at each siRNA concentration (10 nM, 0.75 nM, and 0.1 nM) (see Fig. 3b) and was used in the simulations. The shortest of the two simulated transition times determines if an mRNA transitions to the unmasked or cleaved state.

Finally, we simulate the starting state of each single mRNA, which can be in the masked or unmasked state. To simulate the starting state, we assume that at the start of our experiment the ratio between the masked and unmasked state has reached an equilibrium that depends on the average masking, unmasking and cleavage time. For example, if the average masking time is very long while the average unmasking time is very short, it is expected that most mRNAs are in the unmasked state. To obtain the ratio between the masked and unmasked state we run an initial simulation before starting the full simulation, where all mRNAs start in the masked state. Next, we evaluate the initial simulation until the ratio of mRNAs in the masked and unmasked state reaches an equilibrium (which will depend on the masking, unmasking and cleavage times used in this simulation). This ratio is used in the full simulation to determine if an mRNA starts in the masked or unmasked state.

Thus, we simulate for single mRNAs in which state they start the simulation and how long it takes to transition to the cleaved state. By repeating this procedure for many mRNAs, we obtain a simulated cleavage curve. Next, we performed two simulations where we assessed the effect of lowering siRNA concentration, from 10 nM to either 0.75 nM or 0.1 nM, on the cleavage curves in the absence of

translating ribosomes. In both simulations we simulated the effect of lowering the siRNA concentration at a range of mean unmasking and masking rates, and compared the simulated cleavage curves to experimentally-derived cleavage curves. For each mean unmasking time (we took a range of 1 s to 1500 s), we searched for the average masking time that resulted in the best fit with the data by calculating a goodness-of-fit score (total SSE) using Equation 3.9.

$$\text{Total SSE}(k_{\text{unmask}}, k_{\text{mask}}) = \sum_t [\text{Data}_{10 \text{ nM}}(t) - \text{Simulations}_{10 \text{ nM}}(t, k_{\text{unmask}}, k_{\text{mask}})]^2 + \sum_t [\text{Data}_{0.75/0.1 \text{ nM}}(t) - \text{Simulations}_{0.75/0.1 \text{ nM}}(t, k_{\text{unmask}}, k_{\text{mask}})]^2 \quad (\text{Equation 3.9})$$

The total SSE is calculated by adding together the SSE for the 10 nM cleavage curve in the absence of translating ribosomes (we used the experiments where we added CHX) and the SSE of either the cleavage curve with 0.75 nM siRNA, or 0.1 nM siRNA concentration in the absence of translating ribosomes (we used the experiments where we added CHX). The total SSE represents the difference between the experimentally-derived cleavage curve and the simulated cleavage curve at a specific unmasking and masking rate. For a range of mean unmasking times (1 s to 1500 s), we simulated cleavage curves with different mean masking times and selected the masking time that minimized the total SSE (Equation 3.9). Each individual simulation consisted of 10,000 mRNAs, and was done in 5 replicates using a different random seed in each replicate. The average of the 5 replicates was reported. In this way, we obtained a total SSE value for each unmasking rate.

#### 3.5.34 Statistics

Statistical comparisons were made using a two-tailed Student's t-test except for data shown in Supplemental Fig. 1k, which was based on a paired two-tailed Student's t-test.

#### 3.5.35 Code availability

Custom code used in this study is available on Mendeley data (<http://dx.doi.org/10.17632/h2r32zhgwn.1>).

#### 3.5.36 Data availability

A selection of the raw imaging data (related to figures 1-6) used in this study is available on Mendeley data (<http://dx.doi.org/10.17632/h2r32zhgwn.1>). Source data are available with the paper online.

### 3.6 Acknowledgements

We thank Martin Depken for helpful discussions with the computational modelling. We thank Loes Steller, Iris Bally, and Rupa Banerjee for help with experiments. We would also like to thank the Tanenbaum lab members for helpful discussions, and Tim Hoek and Deepak Khuperkar for critical reading of the manuscript. This work was financially supported by the European Research Council (ERC) through an ERC starting grant (ERCSTG 677936-RNAREG) to M.E.T., a VENI grant from the Netherlands Organization for Scientific Research (NWO) (NWO 016.VENI.171.050) to S.R., an ERC consolidator grant (819299) and a VIDI grant from NWO (864.14.002) to C.J., and the National Institute of General Medical Sciences (R35 GM127090) to I.J.M.; M.E.T., S.R., S.S., D.d.S and I.L. are supported by the Oncode Institute that is partly funded by the Dutch Cancer Society (KWF).

### 3.7 Author Contributions

SR, SS, and MT conceived the project; SR, SS, IL, and DS performed the *in vivo* experiments and analyzed the data; SS performed the computational modeling; TC performed the *in vitro* experiments and analyzed the data under supervision of CJ; YX purified the hAGO2 complex under supervision of IM; SR, SS, and TC prepared the figures; SR, SS and MT wrote the manuscript and TC, YX, IM, and CJ provided input.

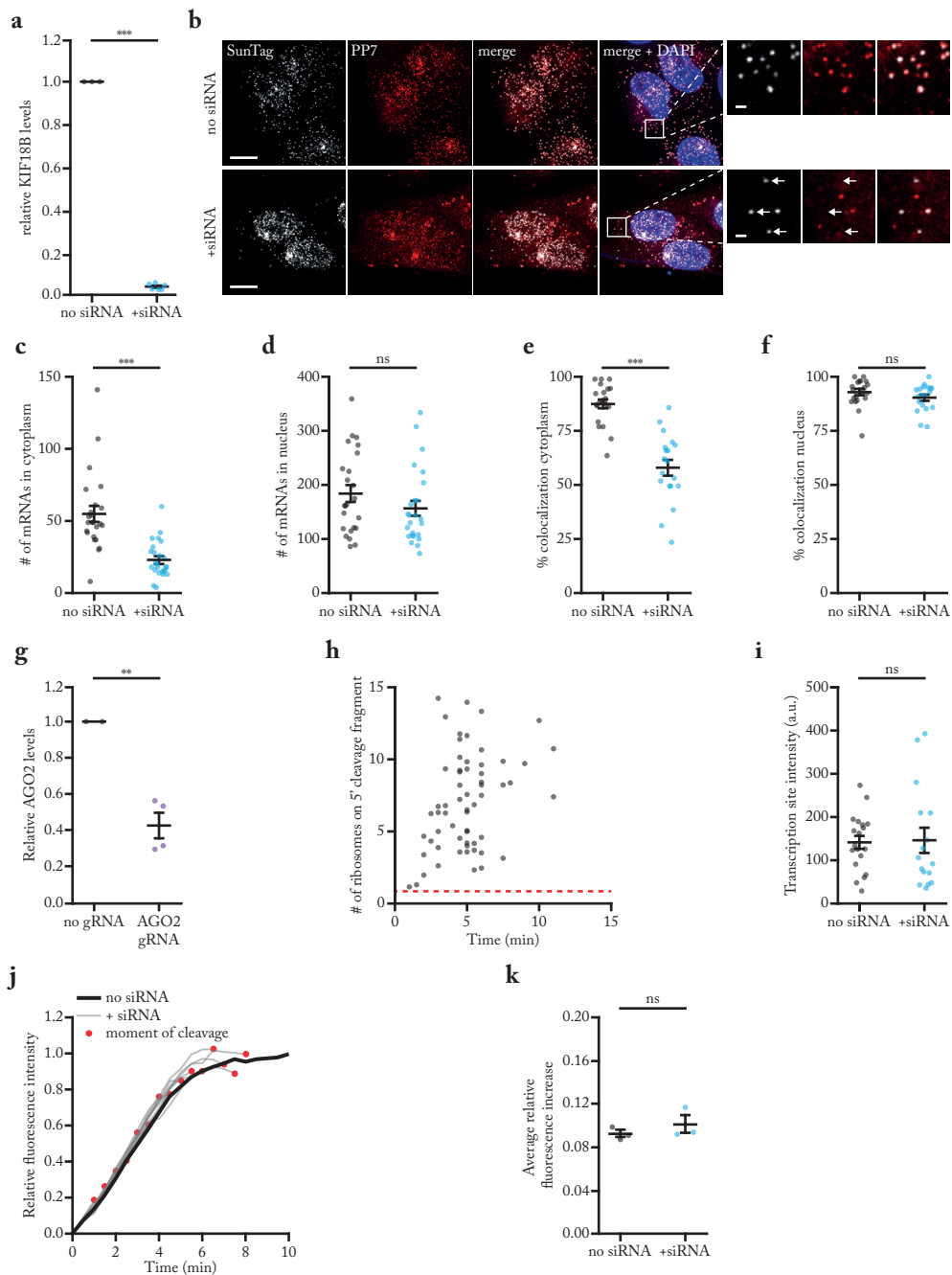
### 3.8 References

- 1 Bartel, D.P. (2018) Metazoan microRNAs. *Cell* 173, 20–51
- 2 Gebert, L.F.R. and MacRae, I.J. (2019) Regulation of microRNA function in animals. *Nat. Rev. Mol. Cell Biol.* 20, 21–37
- 3 Ghildiyal, M. and Zamore, P.D. (2009) Small silencing RNAs: an expanding universe. *Nat. Rev. Genet.* 10, 94–108
- 4 Malone, C.D. and Hannon, G.J. (2009) Small RNAs as guardians of the genome. *Cell* 136, 656–668
- 5 Ozata, D.M. et al. (2019) PIWI-interacting RNAs: small RNAs with big functions. *Nat. Rev. Genet.* 20, 89–108
- 6 Jonas, S. and Izaurralde, E. (2015) Towards a molecular understanding of microRNA-mediated gene silencing. *Nat. Rev. Genet.* 16, 421–433
- 7 Iwasaki, Y.W. et al. (2015) PIWI-interacting RNA: its biogenesis and functions. *Annu. Rev. Biochem.* 84, 405–433
- 8 Liu, J. et al. (2004) Argonaute2 is the catalytic engine of mammalian RNAi. *Science* 305, 1437–1441
- 9 Meister, G. (2013) Argonaute proteins: functional insights and emerging roles. *Nat. Rev. Genet.* 14, 447–459
- 10 Schirle, N.T. et al. (2014) Structural basis for microRNA targeting. *Science* 346, 608–613
- 11 Song, J.J. et al. (2004) Crystal structure of argonaute and its implications for RISC slicer activity. *Science* 305, 1434–1437
- 12 Chandradoss, S.D. et al. (2015) A dynamic search process underlies microRNA targeting. *Cell* 162, 96–107
- 13 Jo, M.H. et al. (2015) Human Argonaute2 has diverse reaction pathways on target RNAs. *Mol. Cell* 59, 117–124
- 14 Salomon, W.E. et al. (2015) Single-molecule imaging reveals that Argonaute reshapes the binding properties of its nucleic acid guides. *Cell* 162, 84–95
- 15 Wee, L.M. et al. (2012) Argonaute divides its RNA guide into domains with distinct functions and RNA-binding properties. *Cell* 151, 1055–1067
- 16 Yao, C. et al. (2015) Single-molecule analysis of the target cleavage reaction by the drosophila RNAi enzyme complex. *Mol. Cell* 59, 125–132
- 17 Jung, S.R. et al. (2013) Dynamic anchoring of the 3'-end of the guide strand controls the target dissociation of argonaute–guide complex. *J. Am. Chem. Soc.* 135, 16865–16871
- 18 Meister, G. et al. (2004) Human Argonaute2 mediates RNA cleavage targeted by miRNAs and siRNAs. *Mol. Cell* 15, 185–197
- 19 Sheu-Gruttadauria, J. and MacRae, I.J. (2017) Structural foundations of RNA silencing by Argonaute. *J. Mol. Biol.* 429, 2619–2639
- 20 Hentze, M.W. et al. (2018) A brave new world of RNA-binding proteins. *Nat. Rev. Mol. Cell Biol.* 19, 327–341
- 21 Jankowsky, E. and Harris, M.E. (2015)

- Specificity and nonspecificity in RNA-protein interactions. *Nat. Rev. Mol. Cell Biol.* 16, 533–544
- 22 Bhattacharyya, S.N. et al. (2006) Relief of microRNA-mediated translational repression in human cells subjected to stress. *Cell* 125, 1111–1124
- 23 Kedde, M. et al. (2007) RNA-binding protein Dnd1 inhibits microRNA access to target mRNA. *Cell* 131, 1273–1286
- 24 Grimson, A. et al. (2007) MicroRNA targeting specificity in mammals: determinants beyond seed pairing. *Mol. Cell* 27, 91–105
- 25 Gu, S. et al. (2009) Biological basis for restriction of microRNA targets to the 3' untranslated region in mammalian mRNAs. *Nat. Struct. Mol. Biol.* 16, 144–150
- 26 Ameres, S.L. et al. (2007) Molecular basis for target RNA recognition and cleavage by human RISC. *Cell* 130, 101–112
- 27 Beaudoin, J.-D. et al. (2018) Analyses of mRNA structure dynamics identify embryonic gene regulatory programs. *Nat. Struct. Mol. Biol.* 25, 677–686
- 28 Becker, W.R. et al. (2019) High-throughput analysis reveals rules for target RNA binding and cleavage by AGO2. *Mol. Cell* 75, 741–755
- 29 Brown, K.M. et al. (2005) Target accessibility dictates the potency of human RISC. *Nat. Struct. Mol. Biol.* 12, 469–470
- 30 Kertesz, M. et al. (2007) The role of site accessibility in microRNA target recognition. *Nat. Genet.* 39, 1278–1284
- 31 Tafer, H. et al. (2008) The impact of target site accessibility on the design of effective siRNAs. *Nat. Biotechnol.* 26, 578–583
- 32 Chen, S.J. (2008) RNA folding: conformational statistics, folding kinetics, and ion electrostatics. *Annu. Rev. Biophys.* 37, 197–214
- 33 Ganser, L.R. et al. (2019) The roles of structural dynamics in the cellular functions of RNAs. *Nat. Rev. Mol. Cell Biol.* 20, 474–489
- 34 Solomatina, S. V. et al. (2010) Multiple native states reveal persistent ruggedness of an RNA folding landscape. *Nature* 463, 681–684
- 35 Ditzler, M.A. et al. (2008) A rugged free energy landscape separates multiple functional RNA folds throughout denaturation. *Nucleic Acids Res.* 36, 7088–7099
- 36 Pichon, X. et al. (2016) Visualization of single polysomes reveals translation dynamics in living human cells. *J. Cell Biol.* 214, 769–781
- 37 Yan, X. et al. (2016) Dynamics of translation of single mRNA molecules in vivo. *Cell* 165, 976–989
- 38 Morisaki, T. et al. (2016) Real-time quantification of single RNA translation dynamics in living cells. *Science* 352, 1425–1429
- 39 Wu, B. et al. (2016) Translation dynamics of single mRNAs in live cells and neurons. *Science* 352, 337–342
- 40 Wang, C. et al. (2016) Real-time imaging of translation on single mRNA transcripts in live cells. *Cell* 165, 990–1001
- 41 Horvathova, I. et al. (2017) The dynamics of mRNA turnover revealed by single-molecule imaging in single cells. *Mol. Cell* 68, 615–625
- 42 Lam, J.K.W. et al. (2015) siRNA versus miRNA as therapeutics for gene silencing. *Mol. Ther. - Nucleic Acids* 4, e252
- 43 Hoek, T.A. et al. (2019) Single-molecule imaging uncovers rules governing nonsense-mediated mRNA decay. *Mol. Cell* 75, 324–339
- 44 Tanenbaum, M.E. et al. (2015) Regulation of mRNA translation during mitosis. *Elife* 4, e07957
- 45 Zuker, M. (2003) Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic Acids Res.* 31, 3406–3415
- 46 Lu, Z. et al. (2016) RNA duplex map in living cells reveals higher-order transcriptome structure. *Cell* 165, 1267–1279
- 47 Metkar, M. et al. (2018) Higher-order organization principles of pre-translational mRNPs. *Mol. Cell* 72, 715–726
- 48 Strobel, E.J. et al. (2018) High-

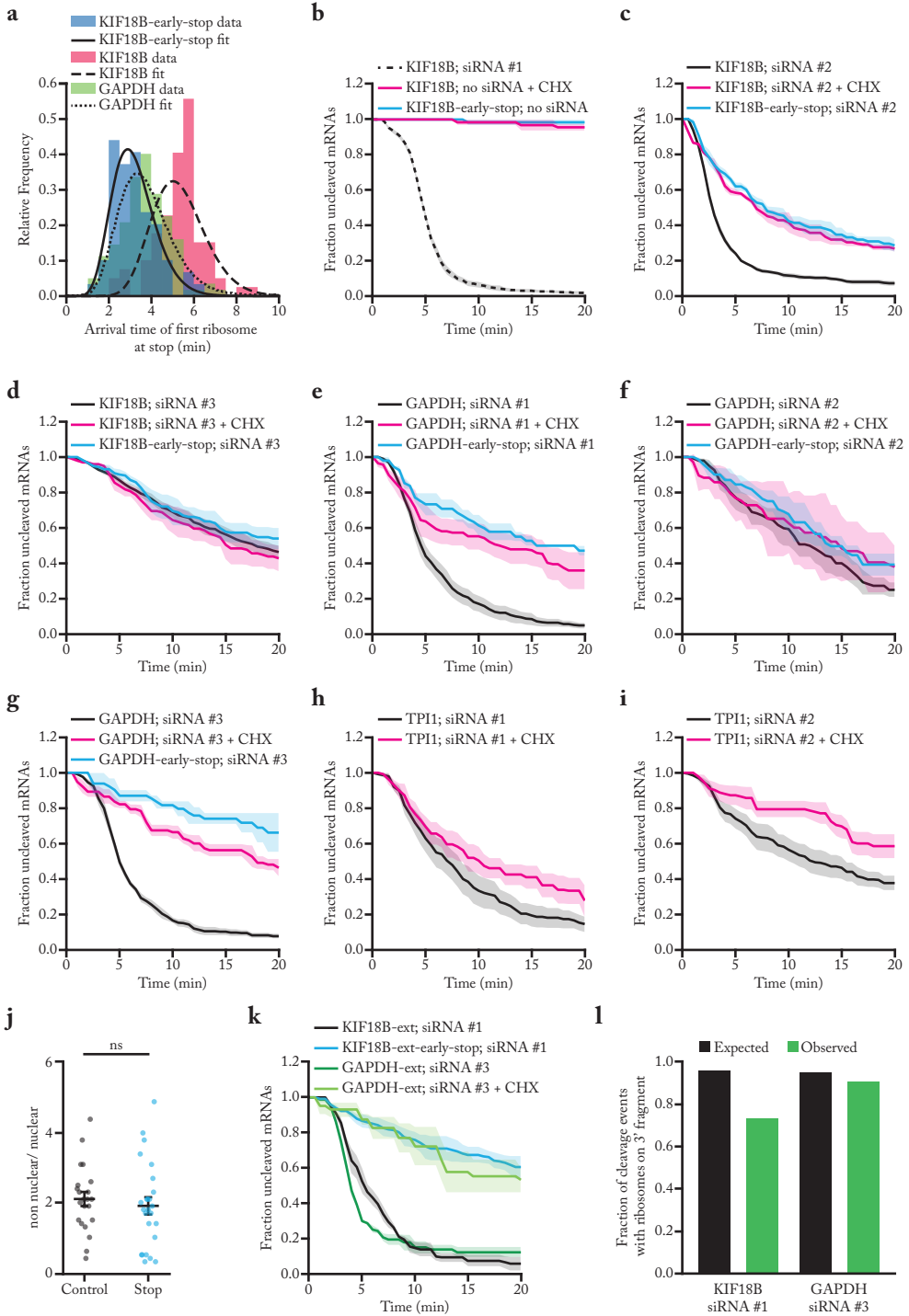
- throughput determination of RNA structures. *Nat. Rev. Genet.* 19, 615–634
- 49 Gong, J. et al. (2018) RISE: a database of RNA interactome from sequencing experiments. *Nucleic Acids Res.* 46, D194–D201
- 50 Aw, J.G.A. et al. (2016) In vivo mapping of eukaryotic RNA interactomes reveals principles of higher-order organization and regulation. *Mol. Cell* 62, 603–617
- 51 Rouskin, S. et al. (2014) Genome-wide probing of RNA structure reveals active unfolding of mRNA structures in vivo. *Nature* 505, 701–705
- 52 Wan, Y. et al. (2014) Landscape and variation of RNA secondary structure across the human transcriptome. *Nature* 505, 706–709
- 53 Zubradt, M. et al. (2016) DMS–MaPseq for genome-wide or targeted RNA structure probing in vivo. *Nat. Methods* 14, 75–82
- 54 Mustoe, A.M. et al. (2018) Pervasive regulatory functions of mRNA structure revealed by high-resolution SHAPE probing. *Cell* 173, 181–195
- 55 Bevilacqua, P.C. et al. (2016) Genome-wide analysis of RNA secondary structure. *Annu. Rev. Genet.* 50, 235–266
- 56 Ding, Y. et al. (2014) In vivo genome-wide profiling of RNA secondary structure reveals novel regulatory features. *Nature* 505, 696–700
- 57 Adivarahan, S. et al. (2018) Spatial organization of single mRNPs at different stages of the gene expression pathway. *Mol. Cell* 72, 727–738
- 58 Mizrahi, O. et al. (2018) Virus-induced changes in mRNA secondary structure uncover cis-regulatory elements that directly control gene expression. *Mol. Cell* 72, 862–874
- 59 Bartel, D.P. (2009) MicroRNAs: target recognition and regulatory functions. *Cell* 136, 215–233
- 60 Tauber, D. et al. (2020) Modulation of RNA condensation by the DEAD-box protein eIF4A. *Cell* 180, 411–426
- 61 Golden, R.J. et al. (2017) An Argonaute phosphorylation cycle promotes microRNA-mediated silencing. *Nature* 542, 197–202
- 62 Raj, A. et al. (2008) Imaging individual mRNA molecules using multiple singly labeled probes. *Nat. Methods* 5, 877–879
- 63 Lyubimova, A. et al. (2013) Single-molecule mRNA detection and counting in mammalian tissue. *Nat. Protoc.* 8, 1743–1758
- 64 Frank, F. et al. (2010) Structural basis for 5'-nucleotide base-specific recognition of guide RNA by human AGO2. *Nature* 465, 818–822
- 65 Flores-Jasso, C.F. et al. (2013) Rapid and specific purification of Argonaute–small RNA complexes from crude cell lysates. *RNA* 19, 271–279
- 66 Joo, C. and Ha, T. (2012) Single-molecule FRET with total internal reflection microscopy. *Cold Spring Harb. Protoc.* 7, 1223–1237
- 67 Chandradoss, S.D. et al. (2014) Surface passivation for single-molecule protein studies. *J. Vis. Exp.* 86, e50549
- 68 Pan, H. et al. (2015) A simple procedure to improve the surface passivation for single molecule fluorescence studies. *Phys. Biol.* 12, 045006
- 69 Edelstein, A. et al. (2010) Computer control of microscopes using manager. *Curr. Protoc. Mol. Biol.* 92, 14.20.1–14.20.17
- 70 Ruijtenberg, S. et al. (2018) Imaging translation dynamics of single mRNA molecules in live cells. *Methods Mol. Biol.* 1649, 385–404
- 71 Boersma, S. et al. (2019) Multi-color single-molecule imaging uncovers extensive heterogeneity in mRNA decoding. *Cell* 178, 458–472

## Supplemental figures



**Supplemental Fig. 1 (opposite page) | Effects of AGO2–siRNA complexes on mRNA transcription and translation.**

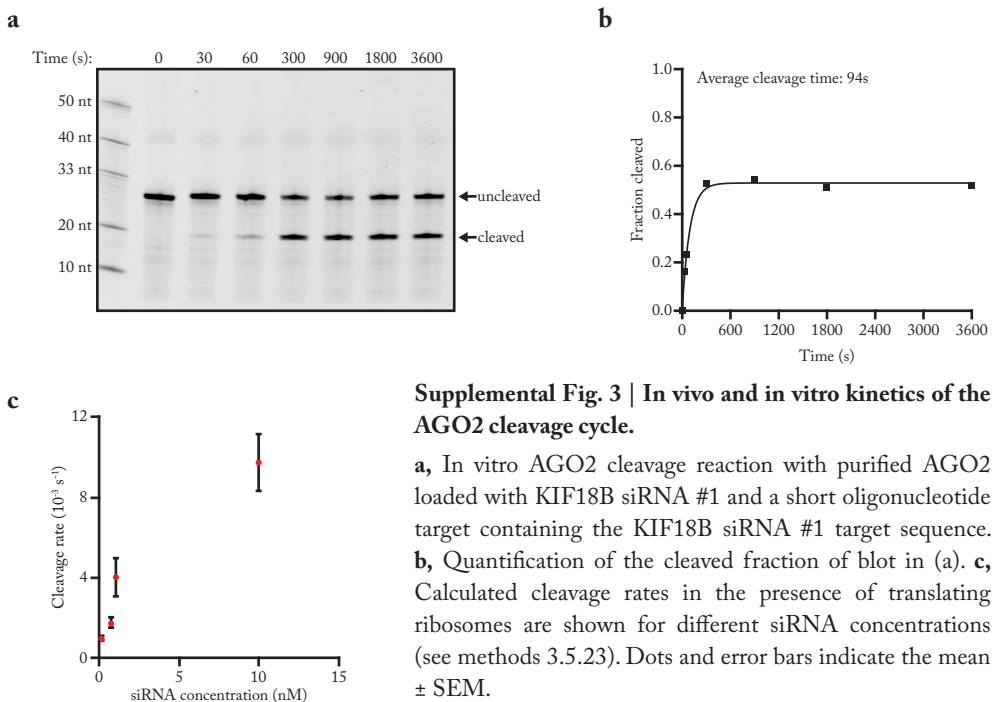
**a**, Relative mRNA levels of endogenous KIF18B based on qPCR in non-transfected cells (no siRNA) and cells transfected with KIF18B siRNA #1 (+ siRNA). Each dot represents an independent experiment and lines with error bars indicate the mean  $\pm$  SEM. **b–f, i**, Cells expressing the KIF18B reporter without siRNA (no siRNA) or transfected with 10 nM KIF18B siRNA #1 (+ siRNA) were fixed and incubated with smFISH probes to visualize reporter mRNAs. **b**, Representative images of cells incubated with smFISH probes targeting the KIF18B reporter (SunTag–Cy5 and PP7–Alexa594) in no siRNA cells (upper panel) and + siRNA cells (lower panel). Arrows in insets indicate mRNA molecules for which the 5' end (SunTag–Cy5 probe) and 3' end (PP7–Alexa594 probe) do not co-localize. Scale bar, 10  $\mu$ m in large images and 1  $\mu$ m in insets. **c–d**, Number of mRNAs in no siRNA and + siRNA cells in (c) the cytoplasm and (d) the nucleus determined based on smFISH using probes targeting the SunTag sequence. Each dot represents a single cell and lines with error bars indicate the mean  $\pm$  SEM. **e–f**, Percentage of mRNAs for which the 5' end (labeled with SunTag probes) and 3' end (labeled with PP7 probes) co-localized in no siRNA and + siRNA cells, either in (e) the cytoplasm or (f) the nucleus. Each dot represents a single cell and lines with error bars indicate the mean  $\pm$  SEM. **g**, Relative AGO2 mRNA levels based on qPCR in control cells (no gRNA) and cells treated with a CRISPRi guide targeting endogenous AGO2 (AGO2 gRNA). Each dot represents an independent experiment and lines with error bars indicate the mean  $\pm$  SEM. **h**, SunTag–PP7 cells expressing the KIF18B reporter were transfected with KIF18B siRNA #1. The number of ribosomes present on the 5' cleavage fragment was determined one frame after the moment of cleavage (see methods 3.5.28). Dotted red line indicates the intensity of a single SunTag array (i.e. the intensity associated with a single ribosome). **i**, Cells were treated for 40 min with dox and the integrated intensity of transcription sites was determined with smFISH probes targeting the SunTag sequence. Each dot represents a single transcription site and lines with error bars indicate the mean  $\pm$  SEM. **j–k**, SunTag–PP7 cells expressing the KIF18B reporter were untransfected (no siRNA) or transfected with KIF18B siRNA #1 (+ siRNA). **j**, GFP intensity over time associated with individual mRNAs is shown for no siRNA cells (black line) and + siRNA cells (grey lines). Black line indicates average of all mRNAs in no siRNA cells, while each grey line represents the average GFP intensity of all mRNAs cleaved at the same moment relative to the start of translation (see methods 3.5.26). The red dot indicates the moment of cleavage. **k**, Average increase in GFP fluorescence intensity either between 1.5–4 min after the start of translation (no siRNA) or at the moment preceding mRNA cleavage (+ siRNA) is shown (see methods 3.5.26). Each dot represents the average of an independent experiment and lines with error bars indicate the mean  $\pm$  SEM. **a, c–f, g, i**, P-values are based on a two-tailed Student's t-test. **k**, P-value is based on a paired two-tailed t-test. P-values are indicated as \* ( $p < 0.05$ ), \*\* ( $p < 0.01$ ), \*\*\* ( $p < 0.001$ ), ns = not significant.

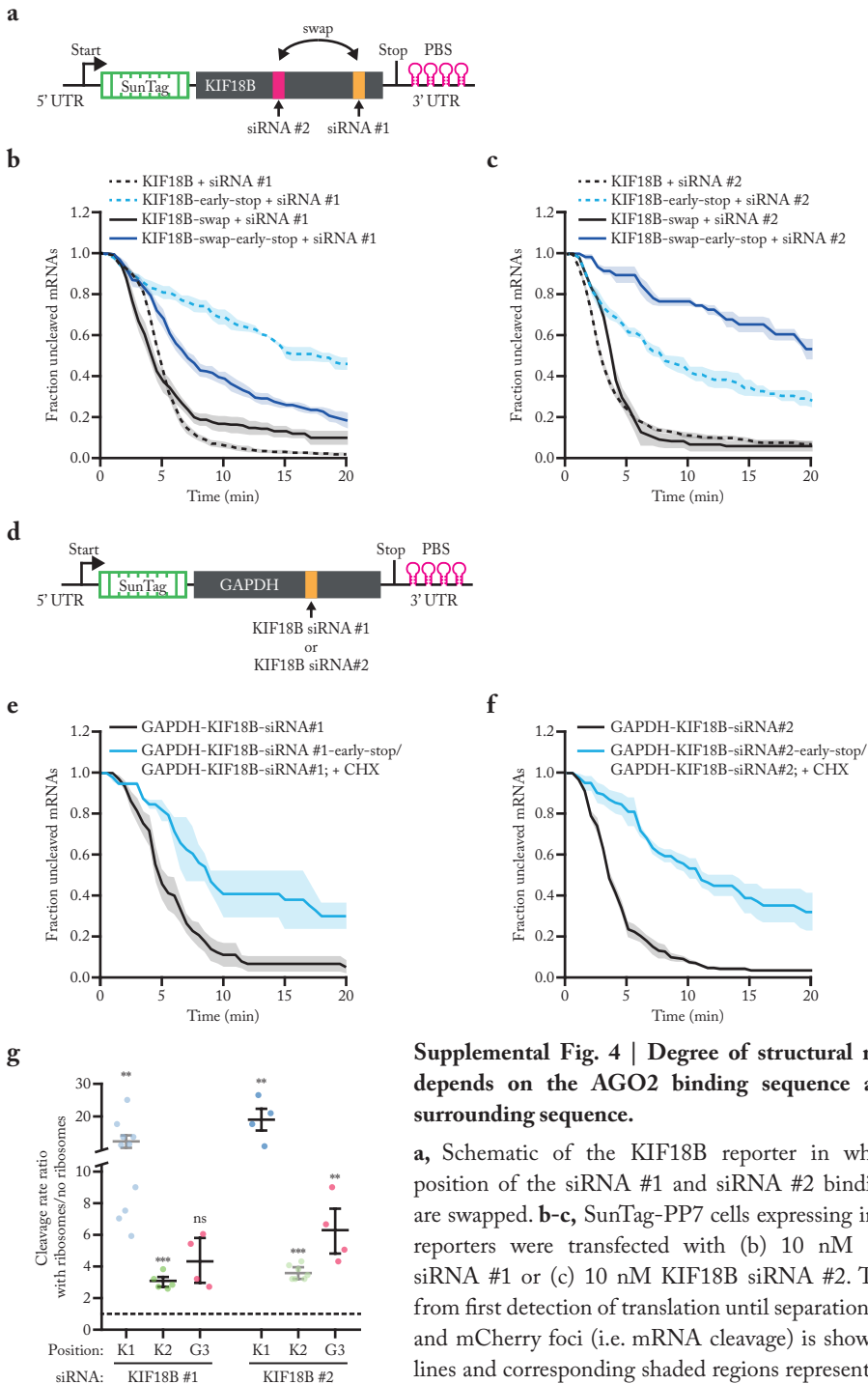




**Supplemental Fig. 2 (opposite page) | Ribosomes stimulate AGO2-dependent mRNA cleavage.**

**a**, The moment at which the first ribosome arrived at the stop codon was calculated for indicated reporters. The experimental data (colored bars) was fit with a gamma distribution (black lines) (see methods 3.5.25). **b-i**, SunTag-PP7 cells expressing the indicated reporters were transfected with 50 nM (KIF18B siRNA #3) or 10 nM (all others) siRNA and treated with CHX, where indicated. The time from first detection of translation or from CHX addition until separation of GFP and mCherry foci (i.e. mRNA cleavage) is shown. Solid lines and corresponding shaded regions represent mean  $\pm$  SEM. Dotted line indicates that the data is replotted from an earlier figure panel for comparison. **j**, Ratio of non-nuclear and nuclear mRNAs 90 min after addition of dox in cells expressing the KIF18B reporter (control) or KIF18B-early-stop reporter (Stop) as determined by smFISH using SunTag probes. Note that mRNA localization is similar for the two cell lines used for northern blot analysis (see Fig. 2e). Each dot represents one cell and lines with error bars indicate the mean  $\pm$  SEM. P-value is based on a two-tailed Student's *t*-test. **k**, SunTag-PP7 cells expressing the indicated reporters were transfected with 10 nM siRNA and treated with CHX, where indicated. The time from first detection of translation or from CHX addition (+ CHX) until separation of GFP and mCherry foci (i.e. mRNA cleavage) is shown. Solid lines and corresponding shaded regions represent mean  $\pm$  SEM. **l**, The fraction of mRNAs that contains a ribosome on the 3' cleavage fragment is shown for mRNAs on which translation started at least 7.5 minutes (KIF18B) or 6 minutes (GAPDH) before the moment of cleavage. On these mRNAs it is expected that the first ribosome has passed the AGO2 target site in ~95% of mRNAs (indicated by black bars) based on the experimentally-derived ribosome elongation rate. The expected fraction (black bars) and observed fraction (green bars) of mRNAs that contains a ribosome on the 3' cleavage fragment is shown.

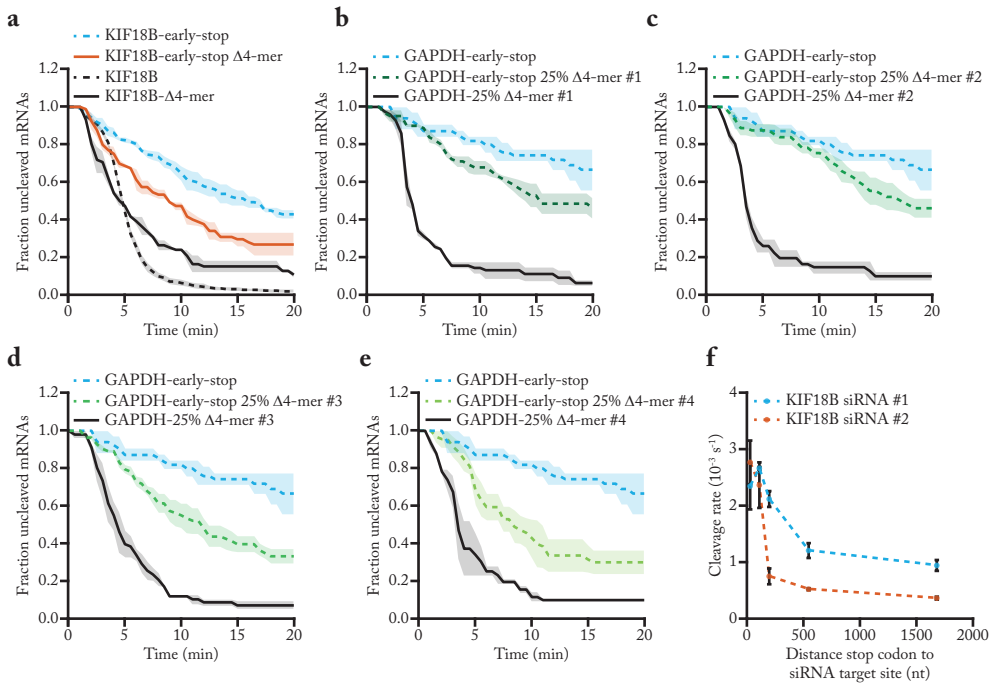




**Supplemental Fig. 4 | Degree of structural masking depends on the AGO2 binding sequence and the surrounding sequence.**

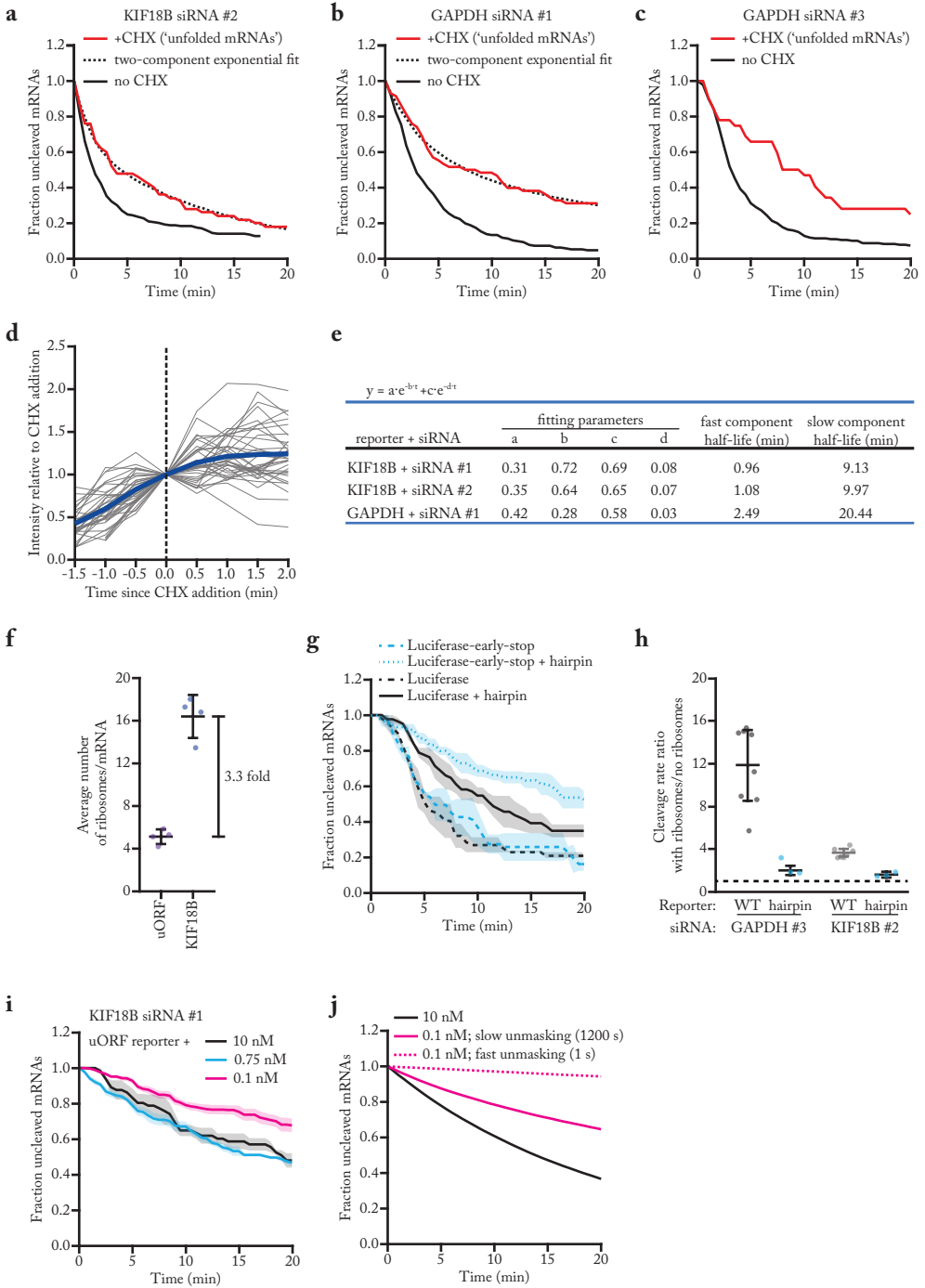
**a**, Schematic of the KIF18B reporter in which the position of the siRNA #1 and siRNA #2 binding sites are swapped. **b-c**, SunTag-PP7 cells expressing indicated reporters were transfected with (b) 10 nM KIF18B siRNA #1 or (c) 10 nM KIF18B siRNA #2. The time from first detection of translation until separation of GFP and mCherry foci (i.e. mRNA cleavage) is shown. Solid lines and corresponding shaded regions represent mean  $\pm$  SEM. Dotted lines indicate that the data is replotted from

an earlier figure panel for comparison. **d**, Schematic of the GAPDH reporter in which the KIF18B siRNA #1 or KIF18B siRNA #2 binding site is placed at the position of GAPDH siRNA #3. **e-f**, SunTag-PP7 cells expressing the indicated reporters were transfected with (e) 10 nM KIF18B siRNA #1 or (f) 10 nM KIF18B siRNA #2. The time from first detection of translation or CHX addition until mRNA cleavage is shown. Note that data of the KIF18B-early-stop reporter and KIF18B reporter treated with CHX are combined to generate the cleavage curve for cleavage in the absence of ribosomes. Solid lines and corresponding shaded regions represent mean  $\pm$  SEM. **g**, Ratio of cleavage rate in the presence and absence of ribosomes is shown for the indicated siRNAs and reporters. Each dot represents a single experiment and lines with error bars indicate the mean  $\pm$  SEM. P-values are based on a two-tailed Student's t-test. P-values are indicated as \* ( $p < 0.05$ ), \*\* ( $p < 0.01$ ), \*\*\* ( $p < 0.001$ ). K1, K2 and G3 indicate the position of the indicated siRNA. K1 refers to the position of KIF18B siRNA #1, K2 to KIF18B siRNA #2 and G3 to GAPDH siRNA #3. Light blue and light green data points are replotted from an earlier experiment.



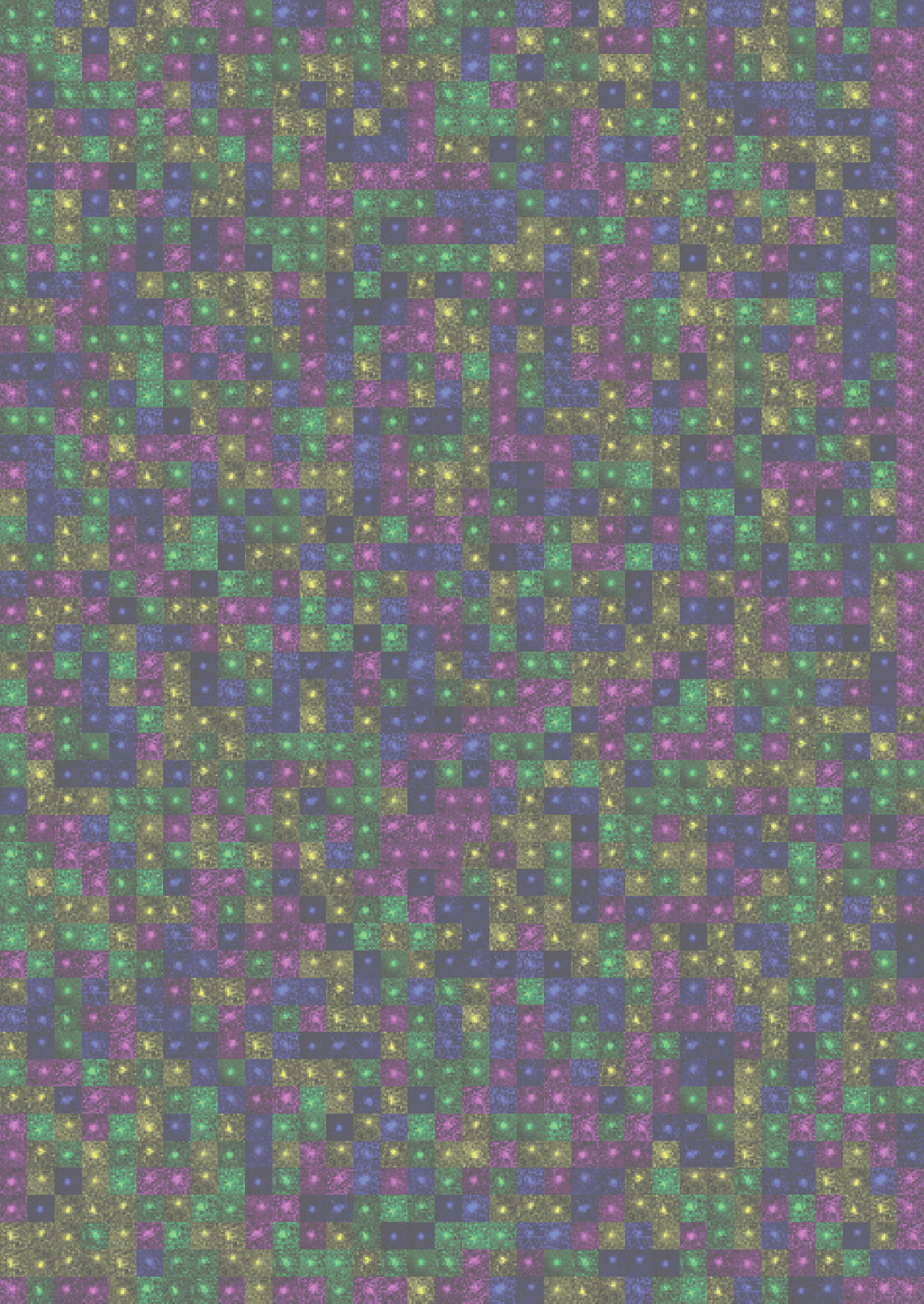
**Supplemental Fig. 5 | Multiple weak intramolecular mRNA interactions together result in potent AGO2 target site masking.**

**a-e**, SunTag-PP7 cells expressing the indicated reporters were transfected with (a) 10 nM KIF18B siRNA #1 or (b-e) 10 nM GAPDH siRNA #3. The time from first detection of translation until separation of GFP and mCherry foci (i.e. mRNA cleavage) is shown. Solid lines and corresponding shaded regions represent mean  $\pm$  SEM. Dotted lines indicate that the data is replotted from an earlier figure panel for comparison. **f**, Cleavage rates for the ‘luciferase’ reporters with indicated siRNA target sites and with different distances between the stop codon and the siRNA target site are shown. Each dot and error bar indicate the mean  $\pm$  SEM. Dotted lines are only for visualization.



**Supplemental Fig. 6 (opposite page) | Structural dynamics of RNA folding.**

**a-c**, SunTag-PP7 cells expressing the indicated reporters were transfected with 10 nM of the indicated siRNA and treated with CHX, where indicated. The CHX cleavage curves (red lines) only include mRNAs for which translation started between (a) 2.5-5.0 min, (b) 2.0-4.5 min, or (c) 2.0-5.0 min before CHX addition (see methods 3.5.22). Dotted lines represent optimal fit with a two-component exponential decay distribution. The no CHX cleavage curve is re-normalized and plotted from (a) 2.5 min or (b-c) 2.0 min after the start of translation. **d**, Relative GFP fluorescence intensities were measured before and after the addition of CHX in SunTag-PP7 cells expressing the KIF18B reporter. Intensity-time traces were aligned at the moment of CHX addition. GFP fluorescence intensities were normalized to the GFP fluorescence intensities at the moment of CHX addition. The thick blue line represents the average intensity of all traces, thin grey lines represent intensity traces of multiple single mRNAs. **e**, Fitting parameters and corresponding half-lives of the two-component exponential fits from Fig. 6a and Supplemental Fig. 6a-b. **f**, Average number of ribosomes per mRNA molecule for the KIF18B-uORF and KIF18B reporters. Each dot represents an independent experiment and lines with error bars indicate the mean  $\pm$  SEM. **g, i**, SunTag-PP7 cells expressing the indicated reporter were transfected with the indicated siRNA and (i) treated with CHX. **g**, The time from first detection of translation until separation of GFP and mCherry foci (i.e. mRNA cleavage) is shown or **i**, the time from CHX addition until mRNA cleavage is shown. Solid lines and corresponding shaded regions represent mean  $\pm$  SEM. Dotted lines indicate that the data is replotted from an earlier figure panel for comparison. **h**, Ratio of the cleavage rates in the presence and absence of translating ribosomes is shown for the indicated siRNAs and reporters. Each dot represents a single experiment and lines with error bars indicate the mean  $\pm$  SEM. Light black data points are replotted from an earlier experiment. **j**, Simulated cleavage curves for 10 and 0.1 nM siRNA concentration using fast or slow unmasking rates (average unmasking time of 1 s and 1200 s, respectively).



# 4

## **Time-resolved single-cell sequencing identifies multiple waves of mRNA decay during mitotic exit**

Lenno Krenning<sup>1,2,3</sup>, Stijn Sonneveld<sup>1,3</sup>, Marvin E. Tanenbaum<sup>1\*</sup>

<sup>1</sup> Oncode Institute, Hubrecht Institute–KNAW and University Medical Center Utrecht, Utrecht, The Netherlands

<sup>2</sup> Present address: Division of Cell Biology, Oncode Institute, Netherlands Cancer Institute, Amsterdam, the Netherlands

<sup>3</sup> Equal contribution

\* Corresponding author and lead contact. Email address for correspondence: [m.tanenbaum@hubrecht.eu](mailto:m.tanenbaum@hubrecht.eu).

## 4.1 Abstract

Accurate control of the cell cycle is critical for development and tissue homeostasis and requires precisely-timed expression of many genes. Cell cycle gene expression is regulated through transcriptional and translational control, as well as through regulated protein degradation. Here, we show that widespread and temporally-controlled mRNA decay acts as an additional mechanism for gene expression regulation during the cell cycle. We find that two waves of mRNA decay occur sequentially during the mitosis-to-G1 phase transition, and identify the deadenylase CNOT1 as a factor that contributes to mRNA decay during this cell cycle transition. Collectively, our data show that, akin to protein degradation, scheduled mRNA decay helps to reshape cell cycle gene expression as cells move from mitosis into G1 phase.

## 4.2 Introduction

Cell division is essential for the development and homeostasis of multicellular organisms. Precise control over cell division is paramount, as errors may contribute to carcinogenesis [1,2]. In order to divide, cells pass through a number of different phases collectively referred to as the cell cycle. The cell cycle in somatic cells consists of four phases: (1) in G1 phase a cell grows and prepares for DNA replication; (2) in S phase the DNA is replicated; (3) in G2 phase a cell prepares for segregation of the replicated genome; (4) in M phase (or mitosis) the cell divides and then enters into G1 phase of the next cell cycle. Progression through the cell cycle is accompanied by the periodic expression of many genes (referred to as cell cycle genes), whose protein products are required in a particular cell cycle phase [3–7]. Deregulated expression of cell cycle genes can decrease the fidelity of cell division. For instance, reduced expression of G2 and M phase cell cycle genes impedes mitotic entry and affects the fidelity of chromosome segregation [8]. Conversely, a failure to suppress expression of G2 and M phase genes as cells enter G1 phase results in a shortened G1 phase and causes DNA replication errors [9–11], and can even contribute to carcinogenesis [12–16]. These examples highlight the importance of tightly controlled gene expression for proper execution of the cell cycle.

To restrict cell cycle gene expression to the correct cell cycle phase, cells need to activate, but also to repress the expression of cell cycle genes as they move from one phase to the next. Scheduled protein degradation plays an important role in repression of cell cycle gene expression by ensuring that protein expression is restricted to the appropriate cell cycle phase [17,18]. In addition, cells prevent de novo synthesis of proteins through inhibition of transcription to further restrict protein expression to the correct cell cycle phase [19,20]. While inhibition of transcription will eventually



lower mRNA levels and thus decrease protein synthesis rates, this process is relatively slow, as it requires turnover of the existing pool of mRNAs. To circumvent this, cells can shut down translation or degrade pre-existing transcripts when transitioning from one cell cycle phase to another. Indeed, control of mRNA translation also contributes to the regulation of gene expression during the cell cycle [21] and several hundreds of genes are subject to translational regulation at different phases of the cell cycle [22,23].

Regulation of mRNA stability during the cell cycle has been studied relatively little, but recent work suggests that this type of regulation also contributes to restriction of cell cycle gene expression. Dynamic changes in mRNA stability during the cell cycle were observed in yeast using fluorescent in situ hybridization (FISH). More specifically, cyclin B mRNA was shown to be destabilized upon completion of mitosis [24]. Globally, mRNA synthesis and decay rates during the cell cycle of yeast were derived through metabolic mRNA labeling in synchronized populations, resulting in the identification of several hundred genes that show periodic changes in mRNA synthesis and degradation rates [25]. Regulation of mRNA stability is also reported to occur during the human cell cycle. For instance, the transcription factor ERG was shown to control the degradation of a set of mRNAs during S phase [26]. Recently, global mRNA synthesis and degradation rates during the human cell cycle were determined [27]. Here, a newly-developed method that simultaneously quantifies metabolically labeled and preexisting unlabeled transcripts in individual cells was used to determine synthesis and degradation rates of individual transcripts along the cell cycle. Together, these studies demonstrate that the stability of many mRNAs change during the cell cycle. However, due to the relatively long measurement time required for the pulse-chase approach used by Battich et al. (up to 6 hours), accurate dynamics and rapid changes, especially around the transition points in the cell cycle, are difficult to determine.

To obtain a highly quantitative view of transcriptome dynamics during cell cycle phase transitions, we set up a method that combines single cell mRNA sequencing and live-cell imaging of cell cycle progression to map transcriptome-wide mRNA expression levels with high temporal resolution during the cell cycle. We focus specifically on the mitosis to G1 (M-G1) phase transition when cells divide and enter into a new cell cycle, as gene expression needs to be 'reset' after cell division. The widespread protein degradation that occurs during the M-G1 phase transition is thought to contribute to this reset [18,28–30]. We hypothesized that, analogous to scheduled protein degradation, mRNA decay might play an important role in resetting cell cycle gene expression by limiting the carry-over of pre-existing G2/M-specific transcripts from one cell cycle into the next. Using our method, we identified

two temporally-distinct waves of mRNA decay: the first wave is initiated during mitotic exit and the second wave is initiated within the first hours of G1 phase. For several of these genes, we show that mRNA decay requires CNOT1, a subunit of the CCR4-NOT mRNA deadenylase complex that shortens the poly(A) tail of mRNAs, generally resulting in their decay [31,32]. Together, our findings demonstrate that, analogous to protein degradation, mRNA degradation occurs at the M-G1 phase transition, and provides an important contribution to the reset of the transcriptome after cell division.

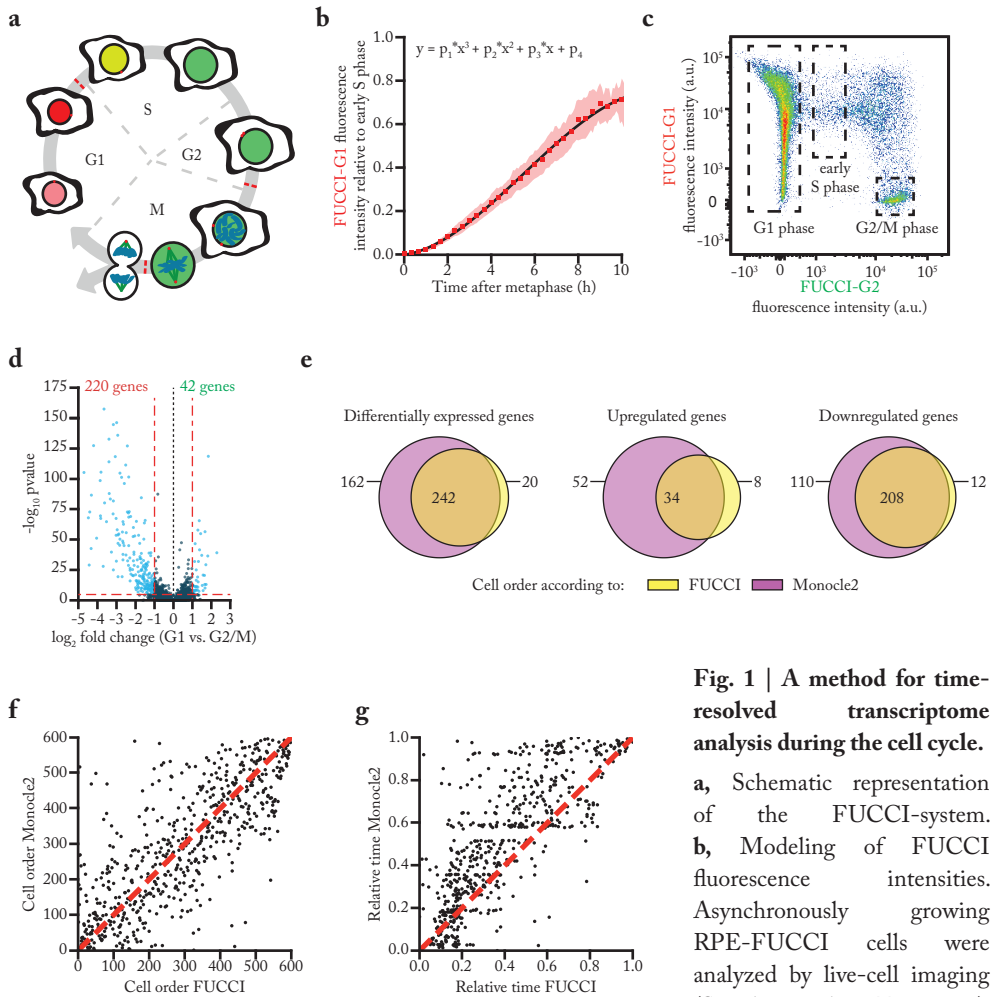
## 4.3 Results

### 4.3.1 Time-resolved transcriptome profiling during the cell cycle using the FUCCI system

To obtain a detailed view of mRNA levels as cells progress from M phase into G1 phase, we developed a method that connects live-cell microscopy with single cell RNA sequencing (scRNA-seq), through fluorescence activated cell sorting (FACS). This method allows us to assign an accurate, ‘absolute’ cell cycle time (i.e., the time in minutes since G1 phase entry) to individual sequenced cells, which we used to generate a high-resolution, time-resolved transcriptome profile of the M-G1 phase transition.

To assign an absolute cell cycle time for each cell, we expressed the fluorescent, ubiquitination-based cell cycle indicator (FUCCI) system in a human untransformed cell line, RPE-1 (RPE-FUCCI). In the FUCCI system an orange fluorescent protein (FUCCI-G1) is expressed in G1 and early S phase cells, while a green fluorescent protein (FUCCI-G2) is expressed in late S, G2 and early M phase (Fig. 1a) [33]. Importantly, the expression levels of both fluorescent markers change over time within each cell cycle phase, potentially allowing precise pinpointing of the cell cycle time of individual cells based on the fluorescence intensity of the FUCCI reporter. We used live-cell microscopy to measure FUCCI-G1 fluorescence intensity in cells as they progressed through G1 phase and observed a monotonic increase during the first 6–8 hours after G1 entry (Supplemental Fig. 1a). To allow accurate calculation of a cell cycle time based on FUCCI-G1 fluorescence intensity, we fit the average FUCCI-G1 fluorescence intensity to a polynomial equation (Fig. 1b). Using this equation, a cell cycle time can be calculated for each fluorescence intensity of the FUCCI-G1 marker as assayed by time-lapse microscopy.

However, in the scRNA-seq protocol the fluorescence intensity of each sequenced cell is measured by FACS analysis. To compare fluorescence intensities measured by FACS and imaging, we normalized FUCCI-G1 fluorescence. Since early S phase cells



**Fig. 1 | A method for time-resolved transcriptome analysis during the cell cycle.**

**a**, Schematic representation of the FUCCI-system. **b**, Modeling of FUCCI fluorescence intensities. Asynchronously growing RPE-FUCCI cells were analyzed by live-cell imaging (Supplemental Fig. 1a).

**c**, FACS analysis of asynchronously growing RPE-FUCCI cells, including the gating strategy used for the identification and isolation of G1, early S and G2/M phase cells. **d**, Differential gene expression analysis of RPE-FUCCI cells in G2/M- versus G1 phase. **e**, Venn diagram comparing differentially expressed genes (both up- and downregulated in G1 versus G2/M phase) identified after FUCCI- or Monocle2-based cell ordering. **f**, Comparison of FUCCI and Monocle2 based ordering of G1 phase cells. Dashed line indicates identical order of cells. **g**, Comparison of G1 phase timing based on either FUCCI-based ordering or pseudo-timing based on trajectory inference by Monocle2. Dashed line indicates identical timing of FUCCI and Monocle2.

can be identified in both live-cell imaging experiments and by FACS analysis (Fig. 1c and Supplemental Fig. 1b), the mean fluorescence intensity of the FUCCI-G1 marker in early S phase can be used as a normalization factor to directly compare the FUCCI-G1 fluorescence intensity values obtained by imaging and FACS (Fig. 1c, Supplemental Fig. 1b; see Methods). Using this normalization factor and the fluorescence intensity of the FUCCI-G1 marker as assayed by FACS, it is possible to map individual G1 cells assayed by FACS onto time-lapse microscopy data, allowing us to pinpoint the precise cell cycle time of each cell that is sorted by FACS.

To validate that our method of converting FACS fluorescence intensities into absolute cell cycle times is accurate, we performed a control experiment. We blocked cells in mitosis using the microtubule stabilizing drug Taxol for various durations, preventing entry of cells in G1 phase. For cells already in G1 phase the FUCCI-G1 fluorescent signal continues to increase. As no new cells enter G1 phase, a gradual loss of cells with low FUCCI-G1 fluorescence is observed by FACS (Supplemental Fig. 1c). By mapping the population of cells that is lost after different times of Taxol treatment we could calculate the FUCCI-G1 fluorescence intensity associated with cells that had spent various times in G1 phase. Comparison of this method to the values obtained with the polynomial equation revealed very similar results (Supplemental Fig. 1d). Thus, we conclude that we can accurately determine the time a cell has spent in G1 phase based on its FUCCI-G1 fluorescence as measured by FACS.

To identify changes to the transcriptome throughout the M-G1 phase transition, we FACS-isolated single G2, M and G1 phase cells based on their FUCCI-G1 and FUCCI-G2 fluorescence (Fig. 1c), and subjected them to scRNA-seq. In total, 1152 cells were sequenced in three replicate experiments, of which 841 cells passed quality checks (see Methods) and were used to generate a high-resolution temporal transcriptome profile of the M-G1 phase transition. Since the FUCCI system does not discriminate between cells in G2 and M phase, and as there are few transcriptome changes between these two phases [23], we averaged the transcript levels of all cells in G2 and M phase (referred to as G2/M). The average G2/M expression levels of individual genes displayed a high correlation between different replicate experiments (Supplemental Fig. 1e-g), allowing us to pool the data from the different experiments. The final dataset consisted of 86 G2/M phase cells and 755 cells from various time points in G1 phase (up to 9 hr after the M-G1 phase transition) (Supplemental Fig. 1h). After initial data processing (see Methods), we performed differential transcriptome analysis comparing G2/M phase to G1 phase cells. This analysis identified 220 genes that were downregulated and 42 genes that were upregulated when cells progressed from G2/M phase into G1 phase (using a cutoff of >2-fold expression change and a p-value of  $<10^{-5}$ ) (Fig. 1d). Gene ontology

analysis revealed that these differentially expressed genes were strongly enriched for cell cycle functions, as expected (Supplemental Fig. i).

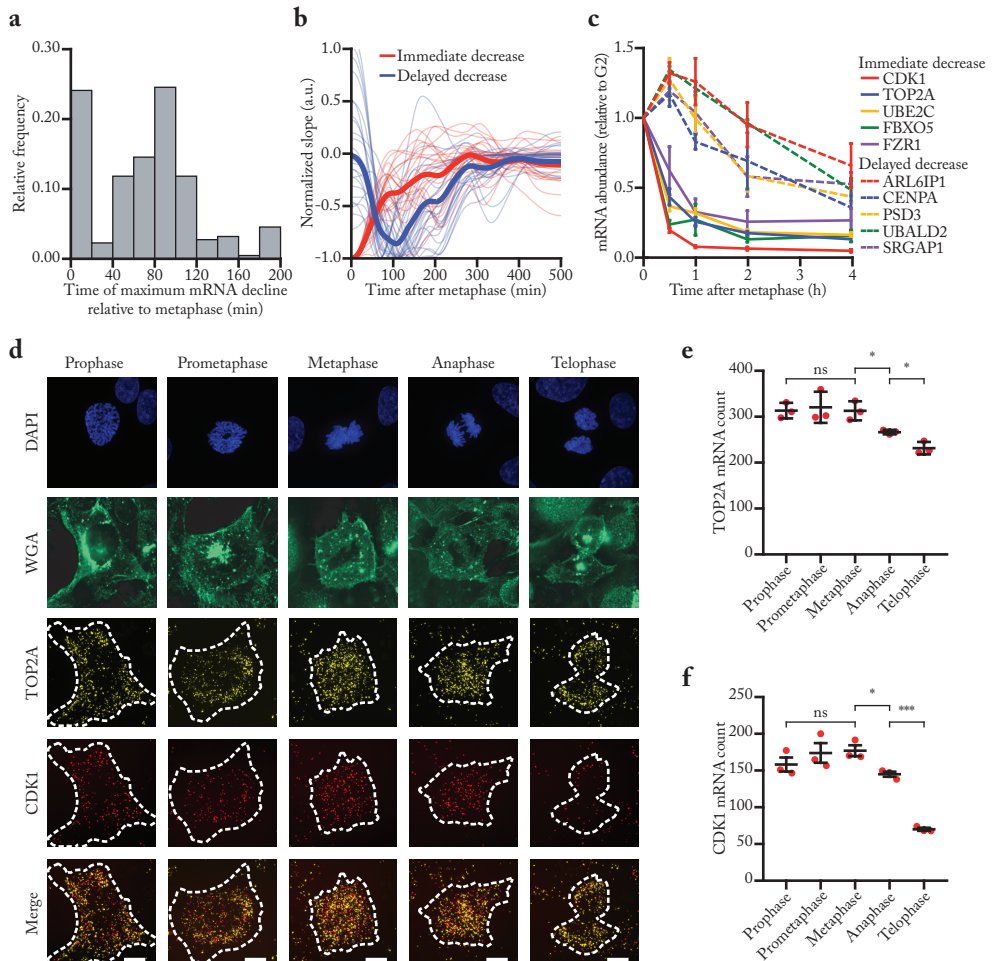
To compare our method of cell cycle time determination with previous computational methods, we used Monocle2, an *in silico* trajectory inference method that orders cells based on their transcriptomes [34–36]. We aligned cells using trajectory inference (Supplemental Fig. 1j and see Methods), and subsequently performed differential transcriptome analysis, which identified 318 downregulated genes and 86 upregulated genes in G1 phase compared to G2/M phase (Supplemental Fig. 1k). There was a large overlap between the differentially expressed genes identified by Monocle2 and our FUCCI-based method (Fig. 1e), and we found a good overall correlation between FUCCI-based ordering and Monocle2-based ordering of G1 phase cells (Fig. 1f and Supplemental Fig. 1j). Monocle2 cannot assign absolute cell cycle times, instead it can compute a ‘pseudo time’ for each G1 phase cell assuming that transcriptome changes occur smoothly over time. Comparing the pseudo time assigned by Monocle2 with the cell cycle time assigned by our FUCCI-based method revealed differences between both methods. In general, Monocle2 computed larger time intervals between cells early in G1 phase compared to our FUCCI-based method (Fig. 1g). As Monocle2 computes the time intervals between cells based on the magnitude of transcriptome changes, a possible explanation for this observation is that transcriptome changes are larger in early G1 phase than at the end of G1 phase, and Monocle2 thus positions cells in early G1 phase too far apart in (pseudo) time. In conclusion, by using the FUCCI-based single-cell sequencing approach we could generate a high-resolution, time-resolved transcriptome profile of cells spanning the transition from M phase into G1 phase.

#### 4.3.2 mRNA levels decline in multiple waves after cell division

As discussed above, we found a large group of genes (220) for which mRNA levels decline at the M-G1 phase transition. To determine the precise moment when mRNA levels started to decline for each gene, we fit the data for individual mRNAs to a smoothing spline and determined the moment of maximum negative slope of the spline, which is the moment when the mRNA level declined most rapidly (referred to as spline analysis; see Methods). Strikingly, the decline in mRNA levels of various genes initiated at two distinct times in the cell cycle: the first occurred around the time of mitotic exit and the second at ~80 minutes into G1 phase (Fig. 2a). To examine these two ‘waves’ of mRNA decline in more detail, we divided the 220 mRNAs into two groups: one for which the maximum negative slope occurred during mitotic exit (immediate decrease) and one for which the maximum negative slope occurred during G1 phase (delayed decrease) (see Methods). Plotting the average slope over

time for both groups (Fig. 2b) revealed that the mRNAs in the immediate decrease group declined most rapidly during the M-G1 phase transition and continued to decline during the first 2–3 hours of G1 phase, whereas the mRNAs in the delayed decrease group mostly declined between 1–4 hours after the start of G1 phase. For both groups, the slopes of individual mRNAs were mostly centered around zero at later times (>8 hours) in G1 phase, suggesting that most mRNAs reached a new steady-state level at later time-points in G1 phase.

To confirm that mRNA levels decline in two distinct temporal waves, we used RT-qPCR to measure mRNA levels for five genes in the immediate decrease group (CDK1, TOP2A, UBE2C, FBXO5, and FZR1) and five genes in the delayed decrease group (ARL6IP1, CENPA, PSD3, UBALD2, and SRGAP1) in G2/M phase and at various time-points in G1 phase (Supplemental Fig. 2). Consistent



with the RNA sequencing data, we observed two distinct waves of mRNA decline by RT-qPCR (Fig. 2c). The minor increase in mRNA levels seen for the 1 hr time-point in the delayed decrease group is likely an artifact caused by comparing a highly synchronized population of early G1 phase cells (which have not yet initiated the decline of delayed genes and thus express the highest possible levels of these transcripts) to a somewhat more heterogeneous population of G2/M phase cells. Collectively, these data demonstrate that there are two distinct waves of mRNA decline during the M-G1 phase transition.

To determine the moment of mRNA decline more precisely for the immediate decrease group, we assessed mRNA levels by single molecule FISH (smFISH) and fluorescence microscopy during different mitotic stages. We fixed asynchronous cultures of cells and stained them specifically for two mRNAs from the immediate decrease group (TOP2A and CDK1). We focused on TOP2A and CDK1 as they showed a strong mRNA decline after metaphase of mitosis (Fig. 2c). To determine the mitotic stages and the outline of the individual cells, we stained the DNA with DAPI, and the membranes with fluorescent wheat germ agglutinin (Fig. 2d). Quantification of TOP2A and CDK1 mRNA levels at various stages of mitosis revealed a significant decrease in mRNA levels as early as anaphase, and a further decrease in telophase for both genes (Fig. 2e-f). We conclude that the first wave of mRNA decline initiates at the start of anaphase whereas the second wave initiates during early G1 phase.

**Fig. 2 (opposite page) | Reduction in mRNA levels occurs in multiple waves, during and after cell division.**

**a**, Time, relative to metaphase, of the highest rate of mRNA decrease for 220 downregulated genes (see Methods). **b**, Average slope of mRNA levels over time for genes that display immediate (thick red line) or delayed decrease (thick blue line). Thin red and blue lines show a random selection of 25 individual genes belonging to the immediate or delayed decrease group, respectively. **c**, Validation of different waves of mRNA reduction. RPE-FUCCI cells at different stages of the cell cycle were isolated by FACS based on FUCCI fluorescence (see Supplemental Fig. 2 for gating strategy). mRNA expression levels of indicated genes was measured by RT-qPCR. Five genes from the immediate decrease group and five genes from the delayed decrease group were selected. Note that the moment of decrease as measured by RT-qPCR closely mirrors the values obtained by modeling of our single cell sequencing data. Lines with error bars represent average  $\pm$  SEM of 3 experiments. **d**, Example images of TOP2A and CDK1 mRNA levels during the different stages of mitosis. Asynchronously growing RPE-1 cells were fixed and stained for DNA (DAPI), membranes (WGA) and TOP2A and CDK1 mRNA (using bDNA-FISH). Scale bar, 10  $\mu$ m. **e-f**, Quantification of TOP2A (**e**) and CDK1 (**f**) transcripts (shown in **d**) using ImageJ. Each dot represents a single experiment and lines with error bars represent average  $\pm$  SEM of 3 experiments (at least 14 cells per experiment per mitotic phase). P-values are based on a one-tailed unpaired Student's t-test, and are indicated as \* ( $p < 0.05$ ), \*\* ( $p < 0.01$ ), \*\*\* ( $p < 0.001$ ), ns = not significant.

### 4.3.3 mRNA decay drives transcriptomic changes during the M-G1 phase transition

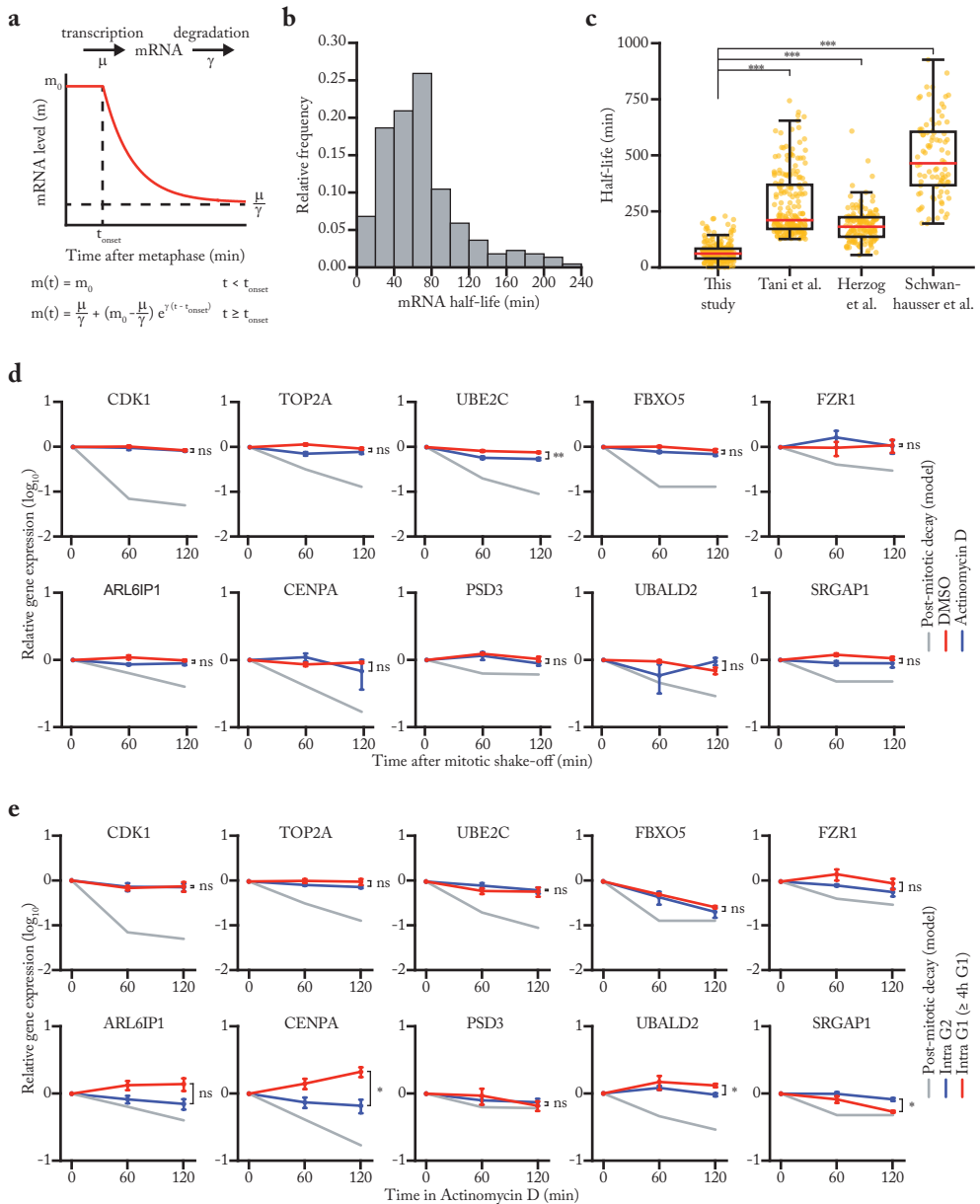
The decline in mRNA levels during early G1 phase may be caused by changes in the rate of mRNA synthesis (transcription) and/or degradation (mRNA stability). To investigate whether mRNA stability is altered during the M-G1 phase transition, we calculated the degradation rate of individual mRNAs using a simple mathematical model (Fig. 3a, see Methods). Briefly, our model describes two phases in the mRNA levels over time: in the first phase mRNA levels remain constant (at an initial level of  $m_0$ ), in the second phase mRNA levels decline to a new steady state level. The onset of decline is described by  $t_{\text{onset}}$ . The rate of decline is dependent on the mRNA degradation rate ( $\gamma$ ), while the new steady-state mRNA level is dependent on the mRNA synthesis rate ( $\mu$ ) and on the mRNA degradation rate ( $\gamma$ ). Using a quality of fit analysis (see Methods), we identified the parameters ( $m_0$ ,  $t_{\text{onset}}$ ,  $\mu$ , and  $\gamma$ ) that resulted in the optimal fit with the data for each of the 220 downregulated genes. Visual inspection showed that the fits described the data well (Supplemental Fig. 3a-f). Using this approach, we confirmed that the onset of decay for different genes occurred most strongly at two distinct times during the M-G1 phase transition; either during the M-G1 phase transition or during early G1 phase (Supplemental Fig. 3g), confirming the results from the spline analysis (Fig. 2a).

**Fig. 3 (opposite page) | mRNA decay occurs during a brief window of time as cells exit mitosis and enter G1 phase.**

**a**, Schematic of the mathematical model that was used to fit the decrease in mRNA levels as cells progress from M into G1 phase. **b**, Histogram of mRNA half-lives of the 220 genes that are downregulated during the M-G1 phase transition. **c**, Boxplot of mRNA half-lives of the 220 genes that were found to be downregulated in our data set (This study). mRNA half-lives of the same genes that were measured in previous studies using bulk assays and asynchronous cell populations are also shown; Tani et al. (HeLa cells) [38], Herzog et al. (mouse embryonic stem cells) [39], and Schwanhauser et al. (mouse fibroblasts) [37]. **d**, Relative mRNA levels in mitosis after different times of transcription inhibition, as measured by RT-qPCR. Mitotic cells were collected by mitotic shake-off, and cultured for an additional 2 hours in the presence or absence of the transcription inhibitor Actinomycin D (blue and red lines, respectively). For comparison, mRNA levels during the M-G1 phase transition are shown (grey line). Note that mRNA of indicated genes is stable in mitosis, indicating that mRNA is degraded specifically during the M-G1 phase transition. Lines with error bars indicate average  $\pm$  SEM of 3 experiments. **e**, Relative mRNA levels in G2 and late G1 phase after different times of transcription inhibition, as measured by RT-qPCR. Asynchronously growing RPE-FUCCI cells were treated with Actinomycin D for indicated times. Cells were then FACS-sorted and G2 phase cells and late G1 phase cells (> 4 hr into G1 phase) were isolated based on FUCCI reporter fluorescence. The mRNA level of indicated genes was then measured by RT-qPCR. As in (d), mRNA levels during the M-G1 phase transition are shown for comparison (gray lines). Note that mRNA levels are substantially less stable in cells during the M-G1 phase transition compared to G2 or late G1 phase cells. Lines with error bars indicate average  $\pm$  SEM of 3 experiments. P-values are based on a one-tailed unpaired or paired Student's t-test (d-e or c, respectively), and are indicated as \* ( $p < 0.05$ ), \*\* ( $p < 0.01$ ), \*\*\* ( $p < 0.001$ ), ns = not significant.



We used mRNA degradation rates extracted from the model to compute the half-lives of the 220 mRNAs that we found to be downregulated in G1 phase. This revealed a median half-life of 61.5 minutes once mRNA levels start to decline during the M-G1 phase transition (Fig. 3b). In most cases the half-lives we computed are substantially shorter than the half-lives of the same mRNAs in asynchronously growing cells as reported previously (Fig. 3c and Supplemental Fig. 3h-j) [37–39].



The comparatively short mRNA half-lives we find during the M-G1 phase transition indicate that these transcripts are subject to scheduled degradation. We observed no significant differences between the half-lives of mRNAs belonging to the immediate decrease group versus the delayed decrease group (Supplemental Fig. 3k), suggesting that for both groups, mRNA degradation plays an important role in the decline of mRNA levels.

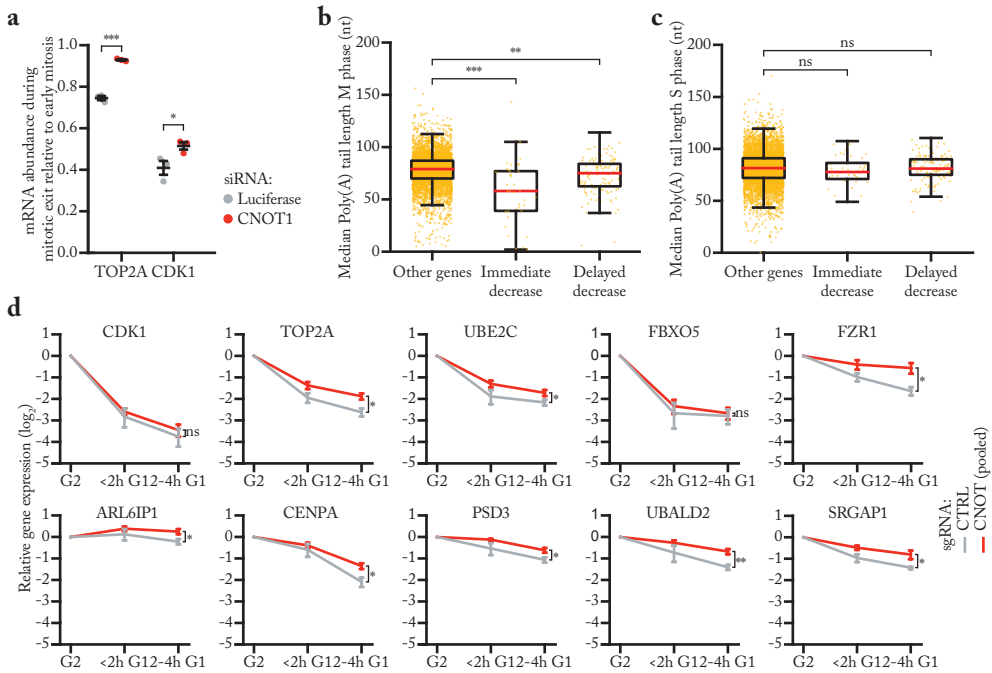
To confirm that mRNAs are subjected to scheduled degradation specifically during the M-G1 phase transition, we examined their stability during mitosis, G2 phase and late G1 phase using an alternative method. To measure mRNA stability in mitosis, we synchronized and arrested RPE-1 cells in prometaphase of mitosis (see Methods), followed by inhibition of transcription for 1 or 2 hours using Actinomycin D. Actinomycin D completely blocked de novo transcription (Supplemental Fig. 3l) and did not influence the arrest of cells in mitosis (Supplemental Fig. 3m). None of the ten mRNAs tested (belonging to both the immediate and delayed decrease groups) showed an appreciable decrease in mRNA levels during the two-hour time window of Actinomycin D treatment, indicating that these mRNAs are much more stable in mitosis than they are during the M-G1 phase transition (Fig. 3d, compare red or blue lines to gray line).

To measure mRNA stabilities in G2 and late G1 phases, we inhibited transcription with Actinomycin D for 1 or 2 hours in asynchronously growing RPE-FUCCI cells. Subsequently, we FACS-sorted populations of G2 cells and late G1 cells (cells that had spent at least 4 hours in G1 phase) and determined mRNA levels of immediate and delayed decay genes with or without Actinomycin D treatment. For all genes tested, mRNA stabilities in either G2 or late G1 phase substantially exceeded the mRNA stability calculated during the M-G1 phase transition (Fig. 3e). Collectively, these data demonstrate that for all genes tested, mRNAs are substantially more stable during G2 phase, mitosis (pre-anaphase), and late G1 phase compared to during the M-G1 phase transition and early G1 phase. Thus, these results uncover an active mRNA decay mechanism that specifically takes place during mitotic exit and early G1 phase.

#### **4.3.4 CNOT1 stimulates mRNA decay during the M-G1 phase transition**

Cytoplasmic mRNA degradation is often initiated by shortening of the poly(A) tail [40], followed by degradation from either end of the mRNA [31]. Shortening of the poly(A) tail is generally mediated by the CCR4-NOT complex [32]. To test whether the CCR4-NOT complex is required for mRNA decay during the M-G1 phase transition, we depleted CNOT1, the scaffold subunit of the CCR4-NOT

complex, using siRNA-mediated depletion in RPE-1 cells (Supplemental Fig. 4a).



**Fig. 4 | CNOT1 promotes decay of mRNAs during mitotic exit and early G1 phase.**

**a**, Cells were transfected with indicated siRNAs. Two days after transfection, cells were fixed and TOP2A and CDK1 mRNAs were stained using smFISH. The mRNA levels were quantified using ImageJ. To calculate the relative abundance of mRNAs during mitotic exit, we divided the number of mRNAs present in telophase by the average number of mRNAs present in prophase, prometaphase and metaphase (mRNA abundance is similar during these phases of mitosis (Fig. 2d-f)). Note that relative abundance was used instead of absolute copy number, as the absolute number of identifiable foci varied between experiments due to variations in labeling intensity of smFISH probes. Each dot represents a single experiment and lines with error bars indicate mean  $\pm$  SEM of three independent experiments. Per experiment, at least 10 cells during mitotic exit and 10 early mitotic cells were quantified. P-values are based on a one-tailed Student's t-test. **b-c**, Boxplot of poly(A)-tail lengths in mitosis (b) and S phase (c) (poly(A) lengths data-set is from [44]) for genes subject to decay during mitotic exit (Immediate decrease), during early G1 phase (Delayed decrease), or genes that are not subject to decay (Other genes). P-values are based on a one-tailed Student's t-test. **d**, CNOT1 contributes to mRNA decay at the M-G1 phase transition for many genes. RPE-FUCCI CRISPRi cells, infected with control- or CNOT1-targeting sgRNAs, were sorted into populations of G2/M phase and G1 phase cells at 5 days post sgRNA infection. The mRNA levels of indicated genes were measured by RT-qPCR. Per experiment, mRNA expression was measured in three CNOT1-depleted samples (using 3 independent sgRNAs targeting CNOT1) and one control sample. mRNA levels of the 3 independent samples were averaged for each biological replicate. Lines and error bars indicate average  $\pm$  SEM of 3 individual experiments. P-values are based on a one-tailed Welch's t-test. P-values are indicated as \* ( $p < 0.05$ ), \*\* ( $p < 0.01$ ), \*\*\* ( $p < 0.001$ ), ns = not significant.

For initial experiments, we focused on TOP2A and CDK1 mRNAs, as both were rapidly and robustly degraded at the M-G1 phase transition. In line with the fact that CNOT1 is essential for proliferation [41–43], fewer mitotic cells were present in the cultures following CNOT1 depletion. Nonetheless, in the mitotic cells that could be identified, depletion of CNOT1 caused a 20–25% increase in the relative abundance of both TOP2A and CDK1 mRNAs in telophase (when decay of these mRNAs has normally occurred) compared to control cells (Fig. 4a). These results suggest that CNOT1-dependent mRNA deadenylation is important for mRNA decay at the M-G1 phase transition.

A previous study found that the mRNAs of many cell cycle genes contain significantly shorter poly(A) tails in M phase compared to S phase [44]. Interestingly, re-analysis of their data revealed that the poly(A)-tails of transcripts in the immediate decrease group were shorter than those in the control group (i.e. genes that did not show mRNA decay at the M-G1 phase transition) (median value 58 versus 79, respectively) (Fig. 4b). The delayed decrease group mRNAs also have shorter poly(A) tails than those in the control group during mitosis, although the effect was relatively minor (median value 75 versus 79, respectively). Importantly, the shortened poly(A) tails were specific to mitosis, as poly(A) tail lengths in S phase of immediate and delayed decrease groups were similar to those of control genes (Fig. 4c). Collectively, these data show that CNOT1 is important for the decay of TOP2A and CDK1 mRNAs during the M-G1 phase transition and suggest that CNOT1-dependent deadenylation in mitosis may contribute to the decay of many mRNAs at the M-G1 phase transition.

To determine whether CNOT1 is also involved in the second wave of mRNA decay, we depleted RPE-FUCCI cells of CNOT1 using CRISPR interference (CRISPRi) [45,46]. Using CRISPRi, we could knock down CNOT1 in a large population of cells, allowing subsequent FACS-based isolation of sufficient numbers of early G1 phase cells. We used three independent single guide (sg)RNAs to target CNOT1 by CRISPRi, which resulted in a modest (~50%) reduction of CNOT1 mRNA levels (Supplemental Fig. 4b). Nonetheless, the modest reduction of CNOT1 mRNA levels causes a clear cell cycle arrest (Fig. 4c), confirming the essential function of CNOT1 for normal proliferation. Importantly, CNOT1 depletion did not affect the synthesis rates of the FUCCI-G1 fluorescent reporter (i.e. accumulation of fluorescence over time) during G1 phase (Supplemental Fig. 4d), allowing us to isolate control and CNOT1-depleted cells at similar times in G1 phase based on FUCCI fluorescence through FACS sorting. G2/M phase and early G1 phase cell populations were isolated by FACS and mRNAs of the immediate decrease and delayed decrease groups were measured by RT-qPCR. Even though depletion of CNOT1 was modest in these

experiments, a small but reproducible decrease in decay was observed for eight of the ten genes tested (Fig. 4d). Taken together, these data demonstrate that both waves of post-mitotic mRNA decay are stimulated by CNOT1.

## 4.4 Discussion

### 4.4.1 Assigning a precise cell cycle time to individual, sequenced cells

Understanding of the regulation and heterogeneity of gene expression has flourished due to the development of single cell sequencing techniques. To investigate transcriptome changes over time, trajectory inference methods have been developed that allow in silico ordering of cells, based on (dis)similarities in transcriptomes [47]. This creates single cell trajectories of a biological process of interest, such as differentiation or the cell cycle (Supplemental Fig. 1j) [36,48], and is useful to study dynamics in gene expression. However, due to clustering based on transcriptome (dis)similarities, pseudo time may under/overestimate true cellular state durations [49]. In addition, trajectories lack real temporal information and are therefore not ideal to determine absolute mRNA synthesis and degradation rates.

To circumvent these issues, we have developed a method that combines live-cell microscopy and FACS-analysis of the FUCCI system with scRNA-seq, to generate a high-resolution, time-resolved transcriptome profile of the M-G1 phase transition in human cells. Even though the FUCCI system has previously been used to order single cell transcriptomes along the cell cycle [27,50,51], a unique feature of our method is that it uses precisely calibrated FUCCI reporter fluorescence intensities for accurate assignment of cell cycle times of individual, sequenced cells. We use these calibrated fluorescence intensities to align cells along the cell cycle according to their cell cycle 'age'. We identify hundreds of mRNAs that show sharp transitions in their expression levels as cells progress from mitosis to G1 phase. The availability of temporal information allowed us to quantitatively determine mRNA degradation rates for these transcripts, which is not possible using trajectory inference methods. Similar approaches are likely possible for other biological events for which fluorescence reporters are available, making this approach a broadly applicable method.

### 4.4.2 mRNA decay at the M-G1 phase transition 'resets' the transcriptome

It is evident that the expression of G2/M-specific genes is reduced following the completion of cell division through scheduled protein degradation and transcriptional inactivation [3–7,17,18,28,30,52]. Here, we identify widespread scheduled mRNA decay during mitotic exit and early G1 phase as an additional mechanism acting to reset gene expression following cell division. Why would mRNA transcripts be

actively degraded when the clearance of transcripts will eventually be achieved by transcription shut-down alone? All the mRNAs we tested are stable during mitosis and late G1 phase (Fig. 3d-e). Therefore, transcription inhibition by itself would lead to significant carry-over of transcripts into the next cell cycle. Considering the half-life of these mRNAs in G2 and M phase, they would persist in G1 phase for many hours. Therefore, decay-mediated clearance of mRNAs as cells exit mitosis will aid to limit expression of the encoded proteins in G1 phase, especially since the majority of degraded mRNAs is efficiently translated during G1 phase [23]. As these genes include many genes that encode for proteins with important functions in cell cycle control (Supplemental Fig. 1i), their continued expression in G1 phase may perturb normal cell cycle progression, and could potentially even contribute to cellular transformation [9–11]. Thus, scheduled mRNA decay during the cell cycle may be important to restrict gene expression of many cell cycle genes to their appropriate cell cycle phases.

#### **4.4.3 CNOT1 promotes two waves of mRNA decay during the M-G1 phase transition**

We have identified two consecutive waves of mRNA decay as cells progress through mitosis and into G1 phase (Fig. 2a,c and Supplemental Fig. 3g). The fact that mRNA decay occurs in two waves may indicate the existence of two distinct mechanisms that act consecutively to degrade transcripts. Interestingly, these two waves of mRNA degradation during the M-G1 phase transition are highly reminiscent of the two consecutive waves of protein degradation that occur at the same time [53,54].

The regulation of mRNA decay often occurs through (sequence) specific interactions between mRNAs and RNA binding proteins (RBPs). Through direct interactions with mRNAs, RBPs can recruit components of the RNA decay machinery, such as the CCR4-NOT complex, to the mRNA. Recruitment of CCR4-NOT, a key regulator of gene expression, will then induce deadenylation of the target transcript, generally followed by degradation [31]. Interestingly, a previous report found that the poly(A)-tail lengths of the genes we identified as immediate decay are already shortened in early mitosis, before these mRNAs are degraded (Fig. 4b-c) [44]. The observation that poly(A) tails of transcripts decayed during the M-G1 phase transition are shorter in mitosis could suggest that CCR4-NOT-dependent deadenylation during early mitosis marks these transcripts for subsequent decay during mitotic exit and early G1 phase. Indeed, we identified CNOT1, an essential member of the CCR4-NOT complex, as a regulator of post-mitotic mRNA decay (Fig. 4a,d). We note that the effects of CCR-NOT depletion on mRNA decay are modest in our experiments, but the magnitude of the effect is likely caused, at least in part, by the inability to

effectively deplete CNOT1, while maintaining cells in a cycling state. Perhaps rapid degradation of CNOT1 protein using inducible protein degradation systems could solve this issue in the future [55–57].

Our data shows that both waves of mRNA decay during the M-G1 phase transition are stimulated by CNOT1 (Fig. 4d). Nonetheless, these waves of mRNA decay may be regulated independently, involving distinct RBPs. Binding of distinct RBPs could ensure the timely decay of specific sets of mRNA during either mitotic exit or early G1 phase. It will be interesting to investigate which RBPs are involved in recognizing different subsets of mRNAs that need to be degraded during particular times in the cell cycle. Identification of such RBPs will allow a better understanding of the function and mechanisms of scheduled mRNA degradation during the cell cycle.

## 4.5 Methods

### 4.5.1 Transfections and lentivirus production

Lentivirus was produced by transfecting HEK293T cells with packaging plasmids (pMD2.G and psPAX2; addgene #12259 and #12260, respectively) and lentiviral plasmids carrying the transgene of interest. Plasmids were transfected using FuGENE HD (Promega) according to the manufacturer's protocol. Two days post transfection, virus was harvested by collecting the culture medium, pelleting cell debris by centrifugation, and collecting the supernatant.

### 4.5.2 Cell culture and generation of cell lines

HEK293T cells were maintained in Dulbecco's Modified Eagle Medium (DMEM) supplemented with 5% fetal bovine serum (FBS, Sigma-Aldrich) and 1% penicillin/streptomycin (Gibco). RPE-1 cells and derivatives were maintained in DMEM/Nutrient Mixture F-12 (DMEM/F12, Gibco) supplemented with 10% FBS and 1% penicillin/streptomycin. To generate RPE-FUCCI cells, RPE-1 cells were transduced with lentivirus expressing mkO2-hCdt1(30/120) (FUCCI-G1) and lentivirus expressing mAG-hGem(1/110) (FUCCI-G2) [33]. Single clones were isolated by fluorescent activated cell sorting (FACS) on a BD FACSFUSION system. One clone was selected that showed cyclic expression of both reporter constructs. To generate RPE-FUCCI CRISPRi cells, RPE-FUCCI cells were transduced with lentivirus carrying dCas9-BFP-KRAB [58], and the 15% highest BFP-positive cells were isolated by FACS.

### 4.5.3 Synchronization of cells in mitosis, transcription inhibition and FACS analysis of mitotic cells

In order to synchronize cells in mitosis, we first arrested cells in G2 by treating cells with the CDK1-inhibitor RO-3306 for 16 hours. Subsequently, RO-3306 was removed and the cells were washed twice with PBS before applying fresh medium supplemented with Taxol, which blocks cells in mitosis. Finally, 45 minutes after Taxol addition, mitotic cells were separated from the interphase cells by shaking of the culture dish (shake-off). This specifically detaches mitotic cells, that were then harvested by collecting the culture medium. To inhibit transcription, we treated cells with 1µg/ml Actinomycin D (Sigma-Aldrich) for the indicated durations. In order to identify mitotic cells, cells were fixed in 80% ethanol (-20°C). Thereafter, cells were stained using an antibody targeting the mitosis-specific marker phosphorylated histone 3 (4N pH3)-ser10 (Upstate, 06-570) and propidium iodide to label DNA content. The mitotic fraction was determined as the fraction of 4N pH3-ser10 positive cells.

### 4.5.4 siRNA transfections

Cells were grown in 96-well microscopy plates (Matriplate, Brooks) or 96-well culturing plate (Greiner Bio-one) and siRNAs were transfected at a final concentration of 10 nM using RNAiMAX (Invitrogen) according to the manufacturer's protocol. Two days post-transfection, cells were either fixed for smFISH or the RNA was harvested for RT-qPCR analysis. For knockdown of CNOT1 and DCP2 we used ON-TARGET plus siRNAs from Dharmacon. As a control, we used a custom siRNA targeting luciferase (5'-CGUACGCGGAAUACUUCGAUU-3') from Dharmacon.

### 4.5.5 CRISPRi

For CRISPR interference (CRISPRi), RPE-FUCCI CRISPRi cells were infected with lentivirus particles expressing a non-targeting single-guide RNA (sgRNA), or a sgRNA targeting CNOT1 [59] and a puromycin resistance cassette followed by BFP. Two days post infection, infected cells were selected with puromycin for 3 days to eliminate uninfected cells.

### 4.5.6 Quantitative reverse transcription PCR (RT-qPCR)

For RT-qPCR analysis, cells were lysed in TriSure (Bioline) and RNA was extracted according to the manufacturers' protocol. First strand synthesis was performed using Bioscript (Bioline). mRNA expression levels were quantified using SYBR-Green Supermix (Bio-Rad) on a Bio-Rad Real-time PCR machine (CFX Connect Real-Time PCR Detection System). Relative mRNA expression levels were calculated using the  $\Delta\Delta C_t$ -method. GAPDH and RPN1 were selected as reference genes for normalization, based on their reported high mRNA stability [37]. RT-qPCR primers were designed using Primer3.

### 4.5.7 Branched DNA single molecule Fluorescent In Situ Hybridization

Single molecule fluorescent in situ hybridization (smFISH) was performed using viewRNA probes targeting TOP2A (probe# VA1-14609) and CDK1 (probe# VA6-17545) (ThermoFisher). Staining was done according to the manufacturer's protocol. In brief, cells were grown in 96-well microscopy plates (Matriplate, Brooks) and fixed for 30 minutes using 4% formaldehyde. Then, cells were permeabilized with detergent solution for 5 minutes at room temperature (RT), and subsequently treated with protease solution for 10 minutes at RT. To label the RNAs, cells were incubated with probes targeting TOP2A and CDK1 for 3 hours at 40°C. Subsequent probes (preAmplifier, Amplifier and Label Probe) were incubated for 1 hour at 40°C. Between probe incubations, cells were washed with wash buffer for 3x 1 minute. After the final incubation (with Label Probe), cells were washed and incubated with DAPI (ThermoFischer, D1306) and wheat germ agglutinin, conjugated to Alexa Fluor 488 (ThermoFisher, W11261) to label DNA and membranes, respectively.

### 4.5.8 Microscopy

For live-cell microscopy, RPE-FUCCI cells were grown on microscopy plates and imaged using a Nikon Ti-E with PFS, equipped with an Andor Zyla 4.2Mpx sCMOS camera, CFI S Plan Fluor ELWD 20x air objective (0.45 NA) and a Lumencor SpectraX light source. Temperature and CO<sub>2</sub> control were provided by an OKO-lab Boldline microscope cage and CO<sub>2</sub> controller. Image analysis was performed using ImageJ software.

For imaging of bDNA-FISH stained samples, we used a Nikon TI2 inverted microscope with a perfect focus system, equipped with a Yokagawa CSU-X1 spinning disc, a 100x oil objective (1.49 NA), and a Prime 95B sCMOS camera (Photometrics).

### 4.5.9 SORT-Seq

SORT-Seq was performed as described previously [60]. Briefly, we sorted in total 104 G2/M phase



cells (FUCCI-G1 negative and FUCCI-G2 positive cells, Fig. 1c) and 893 G1 phase cells (FUCCI-G1 positive and FUCCI-G2 negative cells, Fig. 1c) in three 384-wells plates. Each 384-wells plate contained G1 phase cells from 0-4 hours after the start of G1 phase, but only one plate contained G1 phase cells from 4-9 hours after the start of G1 phase. Therefore, we only used cells from 0-4 hours after the start of G1 phase to identify differentially expressed genes. In subsequent analyses (i.e. the spline analysis and the modelling) we did use all G1 phase cells. After sequencing, we continued with cells (841 in total) that passed quality tests (we removed cells with less than 5900 UMIs or more than 111.000 UMIs to lose low quality cells and doublets, respectively). Finally, we normalized for differences in mRNA recovery per cell using Monocle2 (R package).

#### 4.5.10 Cell cycle timing using the FUCCI system

To obtain a temporal transcriptome profile as cells progress from mitosis into G1 phase, we wanted to compute a cell cycle time for each sorted G1 phase cell (that is, how much time a cell has spent in G1 phase at the moment of sorting). Since FUCCI-G1 levels positively correlate with the amount of time a cell has spent in G1 phase, we reasoned that we could use the measured FUCCI-G1 levels to infer a cell cycle time for a G1 phase cell. To characterize precisely how FUCCI-G1 levels increase during G1 phase, we imaged RPE-FUCCI cells under the microscope with a time-interval of 5 minutes and selected cells that progressed through mitosis into G1 phase. Next, we measured the mean fluorescence intensities of both FUCCI sensors in a region of interest (ROI) in the nucleus using ImageJ and subtracted background signal measured in an extracellular ROI. In each experiment we quantified the fluorescence intensities of the FUCCI sensors in 30 cells.

To compute the average FUCCI-G1 time-trace during G1 phase, we aligned the time-traces of individual cells at the metaphase-to-anaphase transition, which is defined by a sudden decrease in FUCCI-G2 fluorescence [33]. Next, since the total amount of time a cell spends in G1 phase differs for each cell, we clipped individual time-traces at the end of G1 phase, which ends shortly after the FUCCI-G2 levels start to increase [61]. To determine the moment the FUCCI-G2 levels start to increase, we first corrected the FUCCI-G2 traces for fluorescence crosstalk from the FUCCI-G1 marker, which is also excited by the 488 nm laser used for imaging of the FUCCI-G2 marker. To correct the FUCCI-G2 time-traces for crosstalk from the FUCCI-G1 marker, we subtracted at each time point 31% of the FUCCI-G1 fluorescence intensity from the FUCCI-G2 fluorescence intensity. Next, we determined the time point when the mean FUCCI-G2 fluorescence intensity reached a threshold value, which was set by visual inspection, and clipped all FUCCI-G1 time traces at the time point of FUCCI-G2 increase. Finally, we computed for each experiment the average FUCCI-G1 levels from the moment of metaphase and fit the average of three experiments to a third-order polynomial (Fig. 1b).

To directly compare the FUCCI-G1 levels that we measured on the microscope to the FUCCI-G1 levels that we measured on the FACS, we normalized both microscopy- and FACS-measured FUCCI-G1 levels to the average FUCCI-G1 level of early S phase cells. To quantify the mean FUCCI-G1 fluorescence intensity in early S phase cells on the microscope, we analyzed the mean nuclear intensities of at least 700 cells per experiment (Supplemental Fig. 1b). As above, we compensated for fluorescence crosstalk from the FUCCI-G1 marker into the FUCCI-G2 channel by subtracting 31% of the FUCCI-G1 fluorescence intensity from the FUCCI-G2 fluorescence intensity. Next, we determined the range of fluorescence intensities for both the FUCCI-G1 and FUCCI-G2 markers (by subtracting the lowest fluorescence intensity from the highest fluorescence intensity), and defined the early S phase population as those cells with FUCCI-G2 intensities between 2,5% and 10% of the range of FUCCI-G2 intensities and FUCCI-G1 intensities higher than 2,5% of the range of FUCCI-G1 intensities (Supplemental Fig. 1b, yellow dots). We computed the average FUCCI-G1 level of the early S phase cells, and normalized the third-order polynomial fit against the average FUCCI-G1 level of the early S phase cells.

To quantify the mean FUCCI-G1 fluorescence intensity in early S phase cells on FACS, we analyzed the FUCCI sensors on FACS (Fig. 1c). We define the early S phase population as those cells that have high FUCCI-G1 levels and have started to increase the expression of the FUCCI-G2 marker (Fig. 1c), and computed the average FUCCI-G1 level in early S phase cells. To obtain a cell cycle time for each G1 phase cell that was sequenced, we determined the FUCCI-G1 fluorescence intensity level that was obtained by FACS and normalized the FUCCI-G1 level to the average early S phase FUCCI-G1 value. Finally, we used the third-order polynomial fit to infer the cell cycle time of each G1 phase cell from its normalized FUCCI-G1 level.

#### 4.5.11 Cell cycle timing using Monocle2

To rank cells using Monocle2 (R package), we used all G2/M phase cells and G1 phase cells that were from the first 4 hours of G1 phase (based on FUCCI cell cycle timing; see section ‘SORT-seq’). Next, we performed an initial differential transcriptome analysis comparing G2/M phase (FUCCI-G1 marker negative and FUCCI-G2 marker positive) and G1 phase cells (FUCCI-G1 marker positive and FUCCI-G2 marker negative) to select differentially expressed genes that Monocle2 can use in subsequent steps to reconstruct the single-cell trajectory. Monocle2 selected a total of 430 genes that were used to reconstruct the single-cell trajectory, and both the cell rank and the Monocle2 assigned pseudo-times were compared to FUCCI-based ranking and cell cycle timing.

#### 4.5.12 Differential transcriptome analysis

Differential transcriptome analysis was performed with Monocle2 (R package), either using FUCCI-based or Monocle2-based cell cycle time. For the differential transcriptome analysis we used all G2/M phase cells and only G1 phase cells from the first 4 hours of G1 phase (see section ‘SORT-seq’). To increase the confidence of our differential transcriptome analysis, we only selected genes for analysis that were clearly detected in all three 384-wells plates. To select detected genes, we computed for each gene in each single 384-wells plate its average expression in G2/M phase cells (as we didn’t want to bias against genes that were downregulated in G1 phase), and only selected genes which had an average expression of at least 2 reads in each single 384-wells plate. This resulted in a dataset of 3985 genes. Finally, after differential transcriptome analysis, genes that showed at least a twofold increase or decrease in expression and had at least a p-value of  $1E^{-5}$  (based on a Bonferroni correction from a p-value of 0.05) were selected as upregulated or downregulated genes, respectively.

#### 4.5.13 Spline analysis

For the spline analysis (performed in Matlab R2018b), we used the full set of 841 cells (see section ‘SORT-seq’). We selected the 220 genes that were identified in the differential transcriptome analysis as downregulated in G1 phase, and fit each gene profile with a smoothing spline. Next, we computed the derivative of the splines at each time point and determined the time when the derivative was minimal for each gene (i.e. the moment mRNA levels decreased most). To compare different genes to each other, we normalized the derivative of each gene to its minimum value (i.e. setting the minimum value to -1). Finally, we determined for each gene the first time point during which the normalized derivative was at least -0.95 (where -1.0 is the minimum slope after normalization), and divided genes in two groups; one group in which the minimum slope was reached at the first time point (i.e. during mitosis) and one group in which the minimum slope was reached during G1.

#### 4.5.14 Modelling mRNA decrease

mRNA levels ( $m$ ) depend on the synthesis ( $\mu$ ) and degradation ( $\gamma$ ) rate, and the change in mRNA levels over time can be described as follows.

$$dm/dt = \mu - \gamma \cdot m \quad (\text{Equation 4.1})$$

To describe the mRNA levels as cells progress from mitosis into G1 phase, we assumed a simple model in which the observed decrease of mRNA levels is explained by a decrease in the synthesis rate and/or an increase in the degradation rate at a specific time point during M or early G1 phase (referred to as the onset time or  $t_{\text{onset}}$ ). When mRNA levels start at a given value ( $m_0$ ), the solution of equation 4.1 results in the following expression for the mRNA levels over time.

$$m(t) = \mu/\gamma + (m_0 - \mu/\gamma) \cdot e^{-\gamma t} \quad (\text{Equation 4.2})$$

Furthermore, we assumed that mRNA levels remain constant before the onset time, resulting in the following pair of equations to describe the mRNA levels as cells progress from mitosis into G1 phase.

$$\begin{aligned} m(t) &= m_0 & t < t_{\text{onset}} \\ m(t) &= \mu/\gamma + (m_0 - \mu/\gamma) \cdot e^{-\gamma \cdot (t - t_{\text{onset}})} & t \geq t_{\text{onset}} \end{aligned} \quad (\text{Equation 4.3})$$

For each gene, we optimized  $t_{\text{onset}}$  (performed in Matlab R2018B) using an iterative search (between 0 and 370 minutes after metaphase in steps of 10 minutes), in which we optimized  $m_0$ ,  $\mu$ , and  $\gamma$  using least square fitting for each  $t_{\text{onset}}$ . Finally, we computed a sum of squared errors (SSE) between the data (using the full dataset of 841 cells) and model for each  $t_{\text{onset}}$  and selected the  $t_{\text{onset}}$  with the minimal SSE.

#### 4.5.15 Calculating half-lives

We computed mRNA half-lives from the degradation rates ( $\gamma$ ) (that we obtained from the modelling) using equation 4.4.

$$\text{Half-life} = \ln(2)/\gamma \quad (\text{Equation 4.4})$$

#### 4.5.16 Statistics

Statistical comparisons were made using a unpaired one-tailed Student's t-test (Fig. 2e-f, 3d-e, 4a-c, Supplemental Fig. 3k and Supplemental Fig. 4d), a paired one-tailed Student's t-test (Fig. 3c) or a one-tailed Welch's t-test (Fig. 4d)

## 4.6 Acknowledgements

We thank members of the Tanenbaum group for helpful discussions, and Xiaowei Yan for help during the initial stages of the project. We thank the Hubrecht Institute flow cytometry facility and single cell sequencing facility (now Single Cell Discoveries) for their technical support. This work was financially supported by the European Research Council (ERC) through an ERC starting grant (ERCSTG 677936-RNAREG) to M.E.T.; M.E.T. is also supported by the Onco Institute that is partially funded by the Dutch Cancer Society (KWF).

## 4.7 Author contributions

LK and MT conceived the project. LK and SS performed the experiments and analysis, and SS performed the computational modeling. LK and SS prepared the figures, and LK, SS, and MT wrote the manuscript.

## 4.8 References

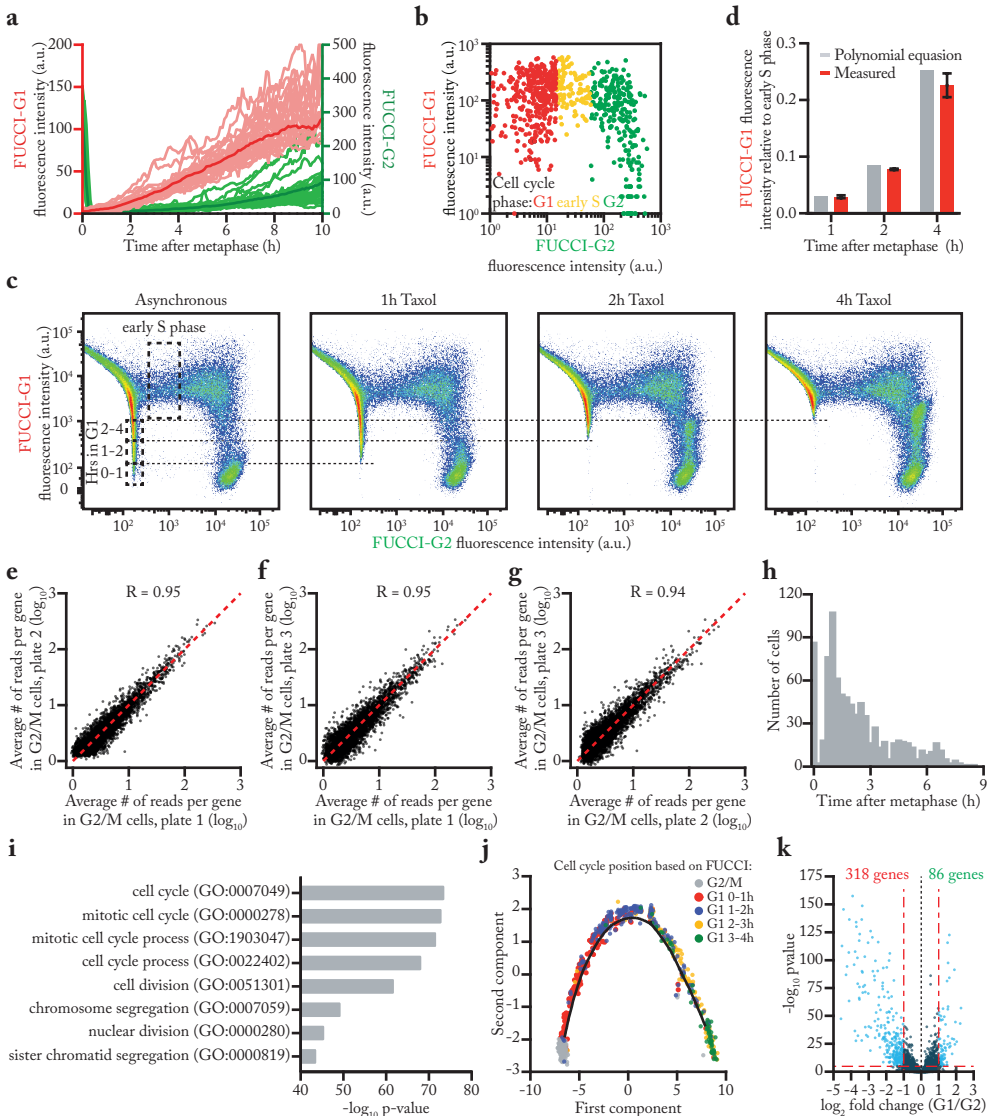
- 1 Hanahan, D. and Weinberg, R.A. (2011) Hallmarks of cancer: the next generation. *Cell* 144, 646–674
- 2 Malumbres, M. and Barbacid, M. (2001) To cycle or not to cycle: a critical decision in cancer. *Nat. Rev. Cancer* 1, 222–231
- 3 Bar-Joseph, Z. et al. (2008) Genome-wide transcriptional analysis of the human cell cycle identifies genes differentially regulated in normal and cancer cells. *Proc. Natl. Acad. Sci. U. S. A.* 105, 955–960
- 4 Chaudhry, M.A. et al. (2002) Gene expression profiling of HeLa cells in G1 or G2 phases. *Oncogene* 21, 1934–1942
- 5 Cho, R.J. et al. (1998) A genome-wide transcriptional analysis of the mitotic cell cycle including control of mRNA transcription. *Mol. Cell* 2, 65–73
- 6 Grant, G.D. et al. (2013) Identification of cell cycle-regulated genes periodically expressed in U2OS cells and their regulation by FOXM1 and E2F transcription factors. *Mol. Biol. Cell* 24, 3634–3650
- 7 Whitfield, M.L. et al. (2002) Identification of genes periodically expressed in the human cell cycle and their expression in tumors. *Mol. Biol. Cell* 13, 1977–2000
- 8 Laoukili, J. et al. (2005) FoxM1 is required for execution of the mitotic programme and chromosome stability. *Nat. Cell Biol.* 7, 126–136
- 9 García-Higuera, I. et al. (2008) Genomic stability and tumour suppression by the APC/C cofactor Cdh1. *Nat. Cell Biol.* 10, 802–811
- 10 Park, H.J. et al. (2008) Anaphase-promoting complex/cyclosome-Cdh1-mediated proteolysis of the forkhead box M1 transcription factor is critical for regulated entry into S phase. *Mol. Cell. Biol.* 28, 5162–5171
- 11 Sigl, R. et al. (2009) Loss of the mammalian APC/C activator FZR1 shortens G1 and lengthens S phase but has little effect on exit from mitosis. *J. Cell Sci.* 122, 4208–4217
- 12 Bortner, D.M. and Rosenberg, M.P. (1997) Induction of mammary gland hyperplasia and carcinomas in transgenic mice expressing human cyclin E. *Mol. Cell. Biol.* 17, 453–459
- 13 Coelho, P.A. et al. (2015) Overexpression of Plk4 induces centrosome amplification, loss of primary cilia and associated tissue hyperplasia in the mouse. *Open Biol.* 5, 150209
- 14 Kalin, T. V. et al. (2006) Increased levels of the FoxM1 transcription factor accelerate development and progression of prostate carcinomas in both TRAMP and LADY transgenic mice. *Cancer Res.* 66, 1712–1720
- 15 Kim, I.M. et al. (2006) The Forkhead Box m1 transcription factor stimulates the proliferation of tumor cells during development of lung cancer. *Cancer Res.* 66, 2153–2161
- 16 Vaidyanathan, S. et al. (2016) In vivo overexpression of Emi1 promotes chromosome instability and tumorigenesis. *Oncogene* 35, 5446–5455
- 17 Nakayama, K.I. and Nakayama, K. (2006) Ubiquitin ligases: cell-cycle control and cancer. *Nat. Rev. Cancer* 6, 369–381
- 18 Vodermaier, H.C. (2004) APC/C and SCF: controlling each other and the cell cycle. *Curr. Biol.* 14, 787–796
- 19 Bertoli, C. et al. (2013) Control of cell cycle transcription during G1 and S phases. *Nat. Rev. Mol. Cell Biol.* 14, 518–528
- 20 Sadasivam, S. and DeCaprio, J.A. (2013) The DREAM complex: master coordinator of cell cycle-dependent gene expression. *Nat. Rev. Cancer* 13, 585–595
- 21 Kronja, I. and Orr-Weaver, T.L. (2011) Translational regulation of the cell cycle: when, where, how and why? *Philos. Trans. R. Soc. B Biol. Sci.* 366, 3638–3652
- 22 Stumpf, C.R. et al. (2013) The translational landscape of the mammalian cell cycle. *Mol. Cell* 52, 574–582

- 23 Tanenbaum, M.E. et al. (2015) Regulation of mRNA translation during mitosis. *Elife* 4, e07957
- 24 Trcek, T. et al. (2011) Single-molecule mRNA decay measurements reveal promoter-regulated mRNA stability in yeast. *Cell* 147, 1484–1497
- 25 Eser, P. et al. (2014) Periodic mRNA synthesis and degradation co-operate during cell cycle gene expression. *Mol. Syst. Biol.* 10, 717
- 26 Rambout, X. et al. (2016) The transcription factor ERG recruits CCR4-NOT to control mRNA decay and mitotic progression. *Nat. Struct. Mol. Biol.* 23, 663–672
- 27 Battich, N. et al. (2020) Sequencing metabolically labeled transcripts in single cells reveals mRNA turnover strategies. *Science* 367, 1151–1156
- 28 Castro, A. et al. (2005) The anaphase-promoting complex: a key factor in the regulation of cell cycle. *Oncogene* 24, 314–325
- 29 Wade Harper, J. et al. (2002) The anaphase-promoting complex: it's not just for mitosis any more. *Genes Dev.* 16, 2179–2206
- 30 Peters, J.M. (2002) The anaphase-promoting complex: proteolysis in mitosis and beyond. *Mol. Cell* 9, 931–943
- 31 Garneau, N.L. et al. (2007) The highways and byways of mRNA decay. *Nat. Rev. Mol. Cell Biol.* 8, 113–126
- 32 Yamashita, A. et al. (2005) Concerted action of poly(A) nucleases and decapping enzyme in mammalian mRNA turnover. *Nat. Struct. Mol. Biol.* 12, 1054–1063
- 33 Sakaue-Sawano, A. et al. (2008) Visualizing spatiotemporal dynamics of multicellular cell-cycle progression. *Cell* 132, 487–498
- 34 Qiu, X. et al. (2017) Reversed graph embedding resolves complex single-cell trajectories. *Nat. Methods* 14, 979–982
- 35 Qiu, X. et al. (2017) Single-cell mRNA quantification and differential analysis with Censur. *Nat. Methods* 14, 309–315
- 36 Trapnell, C. et al. (2014) The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat. Biotechnol.* 32, 381–386
- 37 Schwanhäusser, B. et al. (2011) Global quantification of mammalian gene expression control. *Nature* 473, 337–342
- 38 Tani, H. et al. (2012) Genome-wide determination of RNA stability reveals hundreds of short-lived noncoding transcripts in mammals. *Genome Res.* 22, 947–956
- 39 Herzog, V.A. et al. (2017) Thiol-linked alkylation of RNA to assess expression dynamics. *Nat. Methods* 14, 1198–1204
- 40 Eisen, T.J. et al. (2020) The dynamics of cytoplasmic mRNA metabolism. *Mol. Cell* 77, 786–799
- 41 Blomen, V.A. et al. (2015) Gene essentiality and synthetic lethality in haploid human cells. *Science* 350, 1092–1096
- 42 Hart, T. et al. (2015) High-resolution CRISPR screens reveal fitness genes and genotype-specific cancer liabilities. *Cell* 163, 1515–1526
- 43 Wang, T. et al. (2015) Identification and characterization of essential genes in the human genome. *Science* 350, 1096–1101
- 44 Park, J.E. et al. (2016) Regulation of poly(A) tail and translation during the somatic cell cycle. *Mol. Cell* 62, 462–471
- 45 Gilbert, L.A. et al. (2013) CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. *Cell* 154, 442–451
- 46 Gilbert, L.A. et al. (2014) Genome-scale CRISPR-mediated control of gene repression and activation. *Cell* 159, 647–661
- 47 Saelens, W. et al. (2019) A comparison of single-cell trajectory inference methods. *Nat. Biotechnol.* 37, 547–554
- 48 Haghverdi, L. et al. (2016) Diffusion pseudotime robustly reconstructs lineage branching. *Nat. Methods* 13, 845–848
- 49 Tian, L. et al. (2019) Benchmarking single cell RNA-sequencing analysis pipelines

- using mixture control experiments. *Nat. Methods* 16, 479–487
- 50 Hsiao, C.J. et al. (2020) Characterizing and inferring quantitative cell cycle phase in single-cell RNA-seq data analysis. *Genome Res.* 30, 611–621
- 51 Mahdessian, D. et al. (2021) Spatiotemporal dissection of the cell cycle with single-cell proteogenomics. *Nature* 590, 649–654
- 52 Cho, R.J. et al. (2001) Transcriptional regulation and function during the human cell cycle. *Nat. Genet.* 27, 48–54
- 53 Alfieri, C. et al. (2017) Visualizing the complex functions and mechanisms of the anaphase promoting complex/cyclosome (APC/C). *Open Biol.* 7, 170204
- 54 Sivakumar, S. and Gorbsky, G.J. (2015) Spatiotemporal regulation of the anaphase-promoting complex in mitosis. *Nat. Rev. Mol. Cell Biol.* 16, 82–94
- 55 Banaszynski, L.A. et al. (2006) A rapid, reversible, and tunable method to regulate protein function in living cells using synthetic small molecules. *Cell* 126, 995–1004
- 56 Nishimura, K. et al. (2009) An auxin-based degron system for the rapid depletion of proteins in nonplant cells. *Nat. Methods* 6, 917–922
- 57 Yesbolatova, A. et al. (2020) The auxin-inducible degron 2 technology provides sharp degradation control in yeast, mammalian cells, and mice. *Nat. Commun.* 11, 5701
- 58 Jost, M. et al. (2017) Combined CRISPRi/a-based chemical genetic screens reveal that rigosertib is a microtubule-destabilizing agent. *Mol. Cell* 68, 210–223
- 59 Horlbeck, M.A. et al. (2016) Compact and highly active next-generation libraries for CRISPR-mediated gene repression and activation. *Elife* 5, e19760
- 60 Muraro, M.J. et al. (2016) A single-cell transcriptome atlas of the human pancreas. *Cell Syst.* 3, 385–394
- 61 Grant, G.D. et al. (2018) Accurate delineation of cell cycle phase transitions in living cells with PIP-FUCCI. *Cell Cycle* 17, 2496–2516



## Supplemental figures

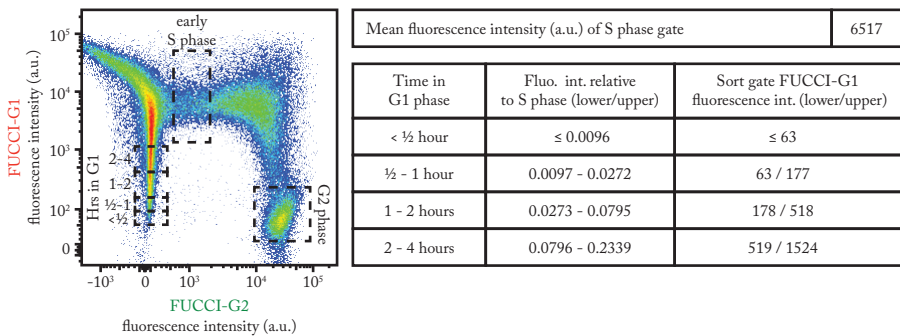


**Supplemental Fig. 1 | FUCCI-based and Monocle2-based ordering of cells during the M-G1 phase transition.**

**a**, Fluorescence microscopy time traces of RPE-1 cells expressing FUCCI-G1 and FUCCI-G2 markers. Asynchronously growing RPE-FUCCI cells were imaged every 5 minutes and the average nuclear intensities for both FUCCI markers was measured using Image J (red and green lines, respectively). Dark red and green lines represent the average of 30 cells from one experiment, light red and green lines represent individual cells. **b**, Microscopy-based analysis of fluorescence intensities of FUCCI-G1 marker in RPE-FUCCI cells. These data were used to identify the average FUCCI-G1 fluorescence

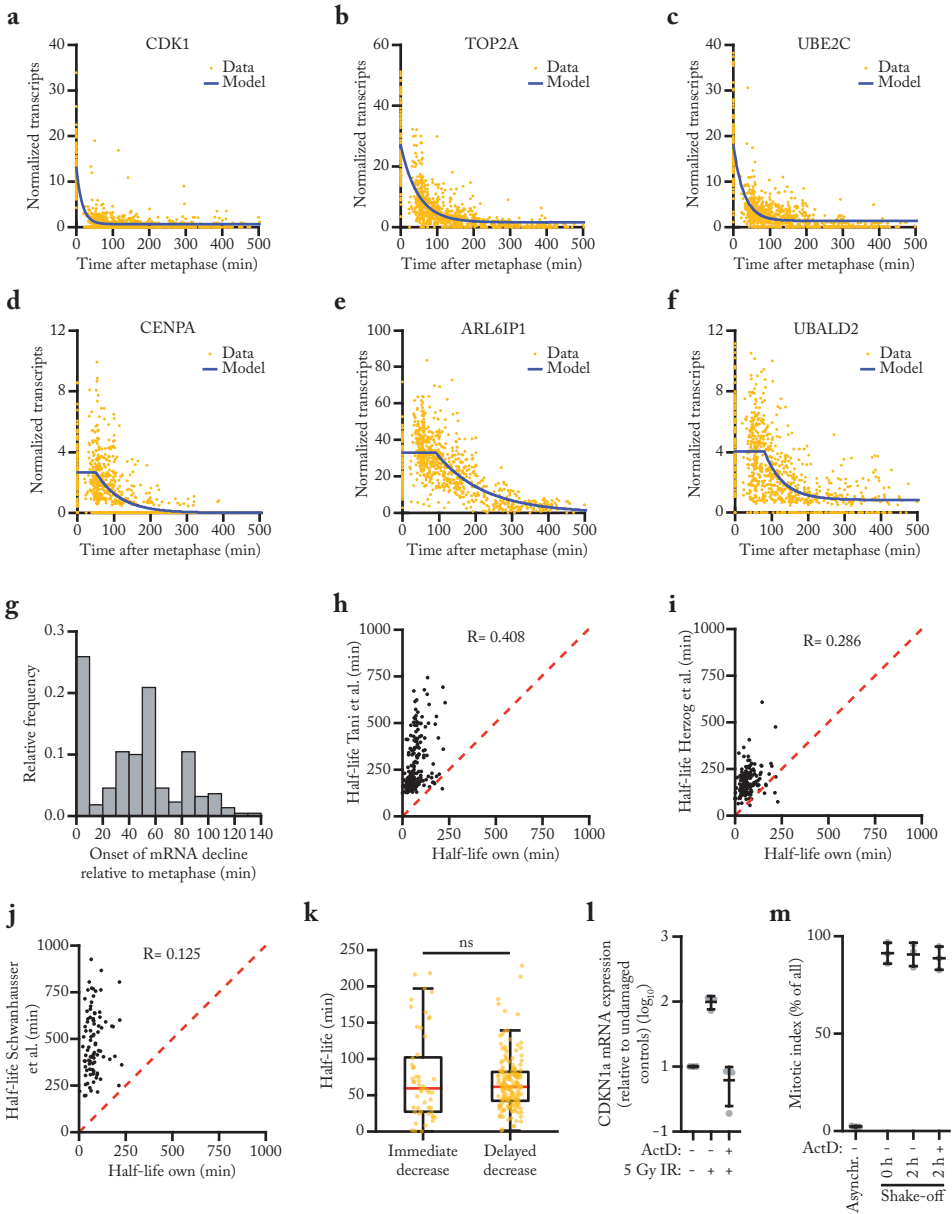


level in early S phase cells (yellow dots, see Methods for details). **c**, FACS plots of FUCCI fluorescence after different durations of Taxol treatment. To quantify the increase in FUCCI-G1 fluorescence over time, asynchronously growing RPE-FUCCI cells were treated with Taxol for the indicated durations to arrest cells in mitosis, thus preventing new cells from entering G1 phase. Over time, FUCCI-G1 fluorescence increases, which results in the gradual loss of cells with low FUCCI-G1 fluorescence. The lowest FUCCI-G1 fluorescence intensity after a 1-, 2- or 4-hour incubation with Taxol were identified (dotted lines), and used to calculate the FUCCI-G1 fluorescence intensity relative to early S phase (see Methods). **d**, Comparison of the relative FUCCI-G1 fluorescence intensities as determined by the polynomial equation and by Taxol treatment of cells followed by FACS (c) at 1, 2 and 4 hours after mitosis. Error bars indicate SEM of 3 experiments **e-g**, Comparison of normalized read counts from three sequencing experiments. Each dot represents the average G2 level of a gene. Dotted red line indicates similar average read counts in both plates. **h**, Histogram showing the position in the cell cycle of all cells subjected to scRNA-Seq. **i**, Gene Ontology of genes downregulated in G1 phase compared to G2/M phase, identified when positioning cells along the cell cycle based on FUCCI marker fluorescence. **j**, Single-cell trajectory of the M-G1 phase transition constructed by Monocle2. Colors indicate the cell cycle position based on FUCCI marker fluorescence. **k**, Differential transcriptome analysis of G1 vs G2/M phase RPE-FUCCI cells, aligned based on Monocle2 trajectory inference.



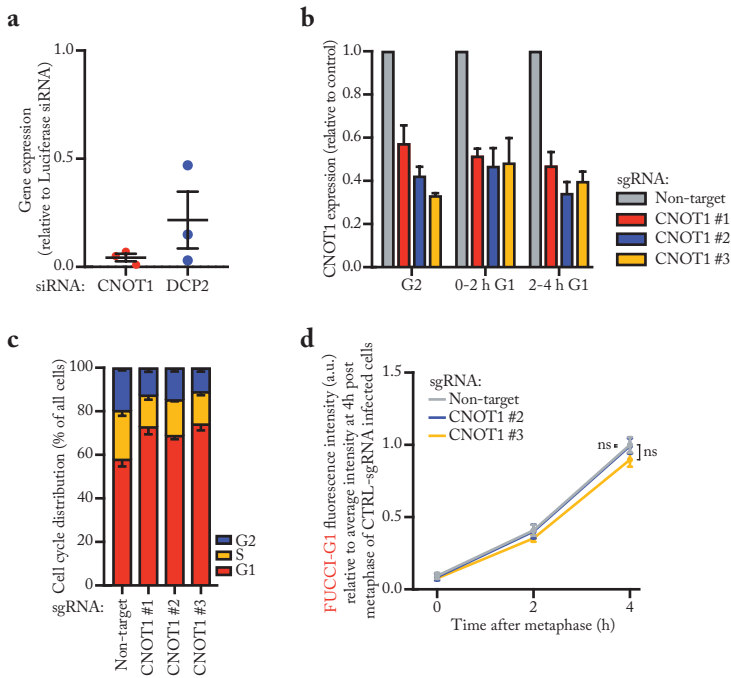
**Supplemental Fig. 2 | Gating strategy for G1 phase cells based on FUCCI-G1 fluorescence intensity.**

FACS analysis of asynchronously growing RPE-FUCCI cells, including gating strategy for the identification of various populations of G1, early S and G2 phase cells (left). The table displays the FUCCI-G1 fluorescence intensity relative to early S phase (middle column) that is associated with the time a cell has spent in G1 phase (left column), based on the polynomial equation (Fig. 1b). The relative FUCCI-G1 fluorescence intensity values (middle column) are converted to absolute FUCCI-G1 fluorescence intensity values (right column) using the mean FUCCI-G1 fluorescence intensity during early S phase (value listed above the table). The absolute fluorescence intensities of FUCCI-G1 (right column) were then used to isolate cells of a specific age with the corresponding gating strategy shown in the FACS-plot on the left.



**Supplemental Fig. 3 (opposite page) | Modelling mRNA decrease during the M-G1 phase transition.**

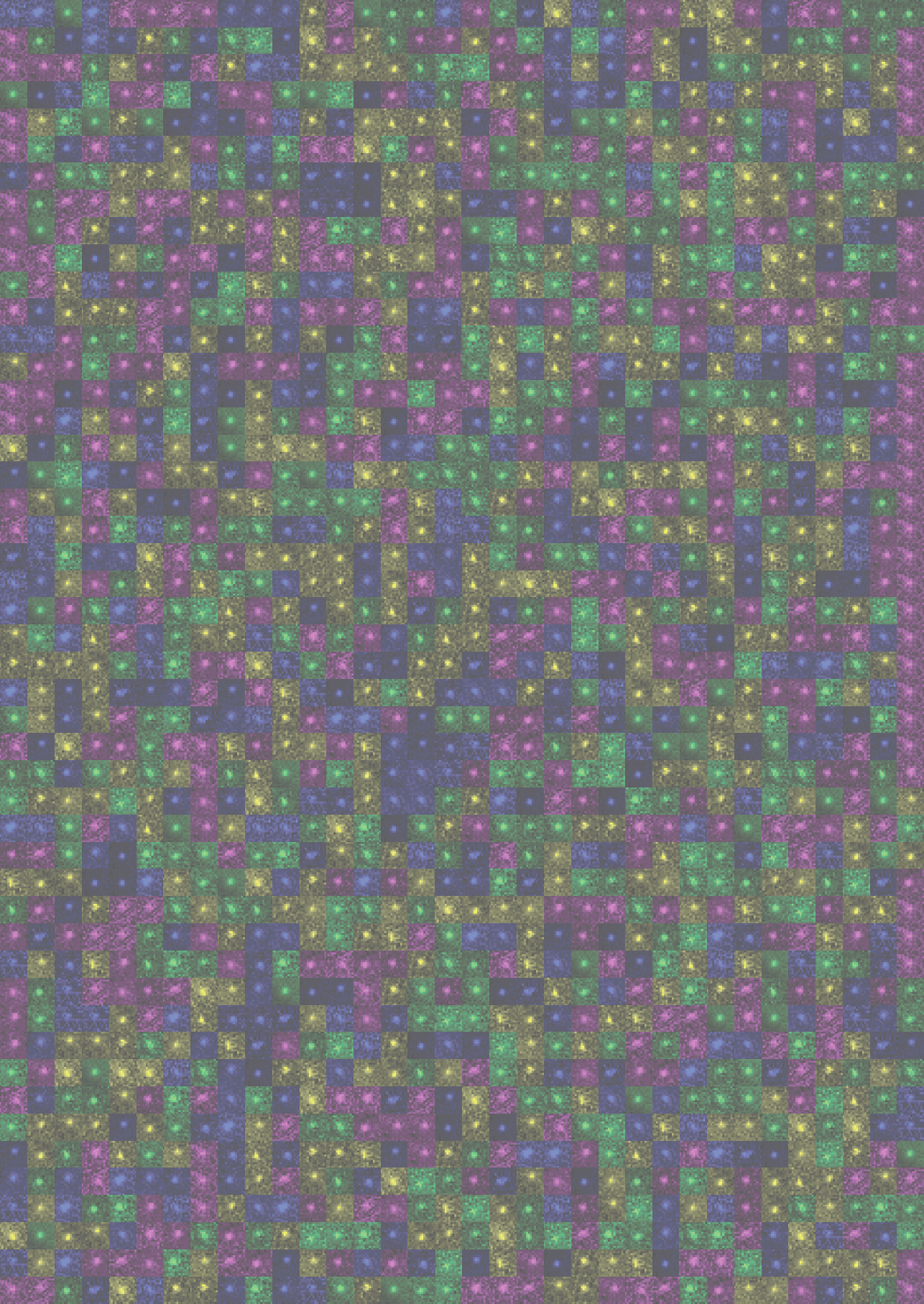
**a-f**, mRNA abundance over time of genes subject to post-mitotic mRNA decay. Blue lines indicate best fit obtained using the mathematical model described in Fig. 3a. Example genes are shown that are representative for the immediate decrease group (CDK1, TOP2A and UBE2C) and for the delayed decrease group (CENPA, ALR6IP1 and UBALD2). **g**, mRNA levels over time were fit as in (a-f) and the onset time of the decline in mRNA levels was determined for each of the 220 downregulated genes. **h-j**, Comparison of mRNA half-lives during the M-G1 phase transition as calculated in Fig. 3b with mRNA half-lives in asynchronous cells determined in previous studies [37–39]. Dashed lines indicate identical half-lives. Note that the half-lives of most genes are shorter during the M-G1 phase transition than in asynchronous growing cells. **k**, Boxplot of mRNA half-lives of immediate and delayed decrease genes. For each gene, the half-life was determined from the moment mRNA levels start to decrease. P-value is based on a one-tailed Student's t-test. **l**, Analysis of transcription inhibition by Actinomycin D. Expression levels of the DNA damage-induced gene CDKN1a were measured by RT-qPCR in cells that were DNA damaged (exposed to 5 Gy ionizing radiation), in the presence or absence of Actinomycin D, relative to non-irradiated cells. Each dot represents a single experiment and lines with error bars indicate mean  $\pm$  SEM. **m**, Mitotic index of RPE-1 cells treated with the transcription inhibitor Actinomycin D. RPE-1 cells were arrested in G2 using a CDK1 inhibitor (RO 3306). After 16h, the CDK1 inhibitor was removed and replaced by Taxol, thereby blocking cells in mitosis. 45 minutes later, mitotic cells were collected through mitotic shake-off, after which Actinomycin D was added for up to 2 hours. Cells were fixed and the fraction of mitotic cells was determined by FACS (by staining cells for DNA content and the mitosis-specific marker phosphorylated histone 3 at ser 10). Each dot represents a single experiment and lines with error bars indicate mean  $\pm$  SEM. P-values are indicated as \* ( $p < 0.05$ ), \*\* ( $p < 0.01$ ), \*\*\* ( $p < 0.001$ ), ns = not significant.



### Supplemental Fig. 4 | CNOT1 depletion levels and effects on the cell cycle.

**a**, Validation of siRNA-mediated knockdown. RPE-1 cells were transfected with indicated siRNAs, or a control siRNA targeting luciferase. mRNA levels relative to control were measured by RT-qPCR at 48 hours post siRNA transfection. Lines with error bars indicate the average  $\pm$  SEM of three independent experiments. **b**, CNOT1-depletion levels of cells depicted in Fig. 4d. RPE-FUCCI CRISPRi cells were infected with the indicated sgRNAs. 5 days post infection, cells were sorted into the indicated cell cycle fractions, and CNOT1 expression was measured by RT-qPCR. Error bars indicate the SEM of three individual experiments. **c**, Cell cycle distribution of control and CNOT1-depleted RPE-FUCCI cells. Cells were treated as in Fig. 4b, except their cell cycle distribution was determined by FACS, based on FUCCI fluorescence. Error bars represent SEM of three independent experiments. **d**, Accumulation of the FUCCI-G1 marker in control or CNOT1-depleted RPE-FUCCI cells. RPE-FUCCI CRISPRi cells were infected with the indicated sgRNAs. 5 days post infection, the cells were imaged using time-lapse microscopy, and FUCCI-G1 fluorescence was determined for cells at metaphase and 2 or 4 hours thereafter. Fluorescence intensities were normalized against the average fluorescence intensity of control cells at 2 hours post-metaphase. Lines and error bars indicate the average  $\pm$  SEM of three independent experiments. At least 8 cells per condition per experiment were quantified. P-values are based on a one-tailed Student's t-test. P-values are indicated as \* ( $p < 0.05$ ), \*\* ( $p < 0.01$ ), \*\*\* ( $p < 0.001$ ), ns = not significant.





# 5

## Live-cell imaging of promoter-enhancer communication

Stijn Sonneveld<sup>1,2</sup>, Sjoerd J. D. Tjalsma<sup>1,2</sup>, Erica S. M. Vos<sup>1,2</sup>, Niels J. Rinzema<sup>1</sup>, Peter H. L. Krijger<sup>1</sup>, Christian Valdes-Quezada<sup>1</sup>, Wouter de Laat<sup>1,\*</sup>, Marvin E. Tanenbaum<sup>1,\*</sup>

<sup>1</sup> Oncode Institute, Hubrecht Institute–KNAW and University Medical Center Utrecht, Utrecht, The Netherlands

<sup>2</sup> These authors contributed equally to this work.

\* These authors are corresponding author and lead contact. Email address for correspondence: w.delaat@hubrecht.eu and m.tanenbaum@hubrecht.eu.

## 5.1 Abstract

Enhancers are genetic elements that regulate the expression of distal target genes by activating the gene promoter. The mechanisms that underlie promoter-enhancer communication and result in activation of the target promoter remain elusive. To investigate this, we have developed a method that simultaneously visualizes promoter-enhancer contacts and transcriptional activity of the target promoter in human erythroleukemia K562 cells. Our initial results suggest that transcriptional activity of the target promoter is not triggered by spatial proximity of the promoter and enhancer. Finally, we establish a proof-of-concept method that aims to measure the co-diffusion between the promoter and enhancer. We hypothesize that co-diffusion may act as a more sensitive measurement of chromosome topology and as such may uncover a functional link between chromosome topology and gene activity.

## 5.2 Introduction

Enhancers are genetic regulatory elements, typically around 100-1000 base pairs (bp) in length, that confer strong transcriptional activation of linked genes [1]. Enhancers are cell-type specific, acting as drivers of transcription programs, based on the availability of lineage specific transcription factors. They are orientation-independent, and can be found at or near gene promoters, within introns, or far away from the genes they control [2-4]. By now the human genome is estimated to contain hundreds of thousands putative enhancers [5,6], which regulate gene expression programs during development and when misregulated can drive disease [7-10]. Interestingly, enhancers are not always in close proximity to their target gene, and can even be found up to several mBs upstream or downstream. In addition, a single gene is often regulated by multiple enhancers, resulting in complex spatial and temporal control. Therefore, understanding the mechanisms behind enhancer communication (i.e. when and how does an enhancer activate its target gene) is essential for a better understanding of gene expression regulation.

The mechanisms behind enhancer communication are still an area of intense debate. Many of the questions revolve around the nature of the contacts made between the enhancer and its target gene. In the canonical model (referred to as the contact model) enhancers are thought to activate the target gene by forming a direct contact with the target gene promoter [1]. Evidence for this model comes from studies using chromosome conformation capture-based methods, which have shown that distal enhancers come into close proximity of their target gene promoter [11,12], and that spatial proximity is correlated with gene expression [13,14]. Several studies have also shown that inducing direct interactions between an enhancer and promoter is



sufficient to drive transcriptional activation, further supporting the contact model [15–18]. However, recent findings have started to question the contact model. Firstly, a study that used 3D-FISH combined with super-resolution microscopy reported increased separation between the Sonic hedgehog (*Sbb*) gene promoter and its enhancers [19]. Secondly, it was found that a single enhancer was able to activate two gene promoters simultaneously [20,21], suggesting that direct interactions are not required for transcriptional activation. Thirdly, a number of recent studies have found that many transcription factors and RNA polymerase II form so-called liquid droplets [22–27]. Together, this data supports a second model for promoter-enhancer contacts (referred to as the condensation model) in which some enhancers (sometimes referred to as ‘super enhancers’) are thought to nucleate the formation of phase-separated condensates that contain high concentrations of transcription factors [28]. These condensates can reach sizes of 200–500 nm, removing the requirement for direct promoter-enhancer contacts for transcriptional activation.

While the distance over which promoter-enhancer contacts are formed is heavily studied, much less is known about the dynamics of promoter-enhancer contacts. In the classic model the promoter and enhancer are thought to form a stable contact, resulting in transcriptional activation (referred to as the stable interaction model). Alternatively, it could be that brief promoter-enhancer contacts drive transcriptional activation (referred to as the hit-and-run model). To assess the real-time dynamics and functional consequences of promoter-enhancer contacts, methods are required that provide high spatial and temporal resolution. One method that can provide a precise and dynamic measurement of promoter-enhancer contact is live-cell imaging. However, live-cell imaging of individual loci has been challenging and only recently have technological developments enabled dynamic tracking of multiple genetic loci at the single-cell level [29,30]. By merging these disciplines with existing methods to follow nascent gene expression [30], recent research has begun to reveal the biophysical and molecular mechanisms underlying four-dimensional gene control [31–35].

To distinguish between the stable interaction and hit-and-run model, two recent studies visualized promoter-enhancer contacts and transcription in developing *Drosophila* embryos and mouse embryonic stem cells, respectively [33,34]. Both studies found promoter-enhancer distances of ~200–400 nm in transcriptionally active cells, in support of the condensate model in which functional promoter-enhancer contacts are made over larger distances. Surprisingly, while in *Drosophila* embryos transcriptional activation was preceded by the formation of a promoter-enhancer contact, this was not observed in the mouse embryonic stem cells in which promoter-enhancer distance was similar in transcriptionally active and inactive cells.

These conflicting findings suggest that promoter-enhancer communication is more complex and could be locus-specific. The results in the mouse embryonic cells also raise the question how transcriptional activation is regulated in these cells. It could be that small conformational changes in the topology of the promoter and enhancer underlie transcriptional activation and that the distance measurements that were performed are not sensitive enough to detect such changes. Alternatively, it could be that other factors such as transcription factor binding triggers transcriptional activation. Taken together, to obtain a better understanding of the dynamics of promoter-enhancer contacts, it is necessary to investigate promoter-enhancer interactions in other model systems and to develop new and sensitive analysis methods that can interrogate chromosome topology.

Here, we lay the foundations for using K562 cells, a human erythroleukemia cell line, to visualize promoter-enhancer contacts as well as transcription dynamics in real-time in live cells, and we develop a new (and potentially more sensitive) method to measure promoter-enhancer contacts. We labeled the promoter in our system using the ANCHOR system [32], and used PP7 RNA hairpin repeats for live-cell imaging of transcription. Next, we developed a 2-kb binding platform to which a fluorescently labeled nuclease-dead Cas9 is recruited to follow enhancer position. Our initial results suggest that functional promoter-enhancer contacts are formed at ~200 nm and that promoter-enhancer distance does not correlate with nascent transcription. Finally, we develop a new method that aims to measure the co-diffusion of the promoter and enhancer. We hypothesize that this method may provide a more sensitive measurement of the conformational status of the promoter and enhancer and as such may reveal conformational changes that drive transcriptional activation.

## 5.3 Results

### 5.3.1 An enhancer reporter system in K562 cells

To investigate promoter-enhancer communication in live cells, we made use of an enhancer reporter system in K562 cells that was previously established (Rinzema et al., in preparation). Briefly, this system is composed of an eGFP reporter gene, which is driven by the  $\gamma$ -globin gene promoter and a downstream (relative to the genome) enhancer based on the endogenous  $\beta$ -globin super-enhancer. The enhancer (referred to as the  $\mu$ LCR enhancer) is a 6.5-kb form of the endogenous  $\beta$ -globin LCR (Locus Control Region, ~20 kb), which acts as a super-enhancer in erythroid cells to confer strong gene transcription via its four juxtaposed hypersensitive sites [36]. The reporter gene and  $\gamma$ -globin promoter were integrated in K562 cells, a human erythroleukemia cell line, at chromosome 18 (hg coordinates: chr18:22,029,048). This locus was selected as the surrounding region is largely devoid of other genes and regulatory

sequences, thereby minimizing the interference of other genetic regulatory elements. Next, the  $\mu$ LCR enhancer was integrated at approximately 100 kB downstream of the  $\gamma$ -globin promoter and 3 CTCF binding sites were positioned directly downstream of the enhancer, allowing DNA looping between the  $\mu$ LCR enhancer and CTCF binding sites upstream of the  $\gamma$ -globin gene promoter. The  $\mu$ LCR enhancer and the CTCF binding sites were flanked by FRT and loxP sites, respectively, allowing for efficient deletion of one or both elements.

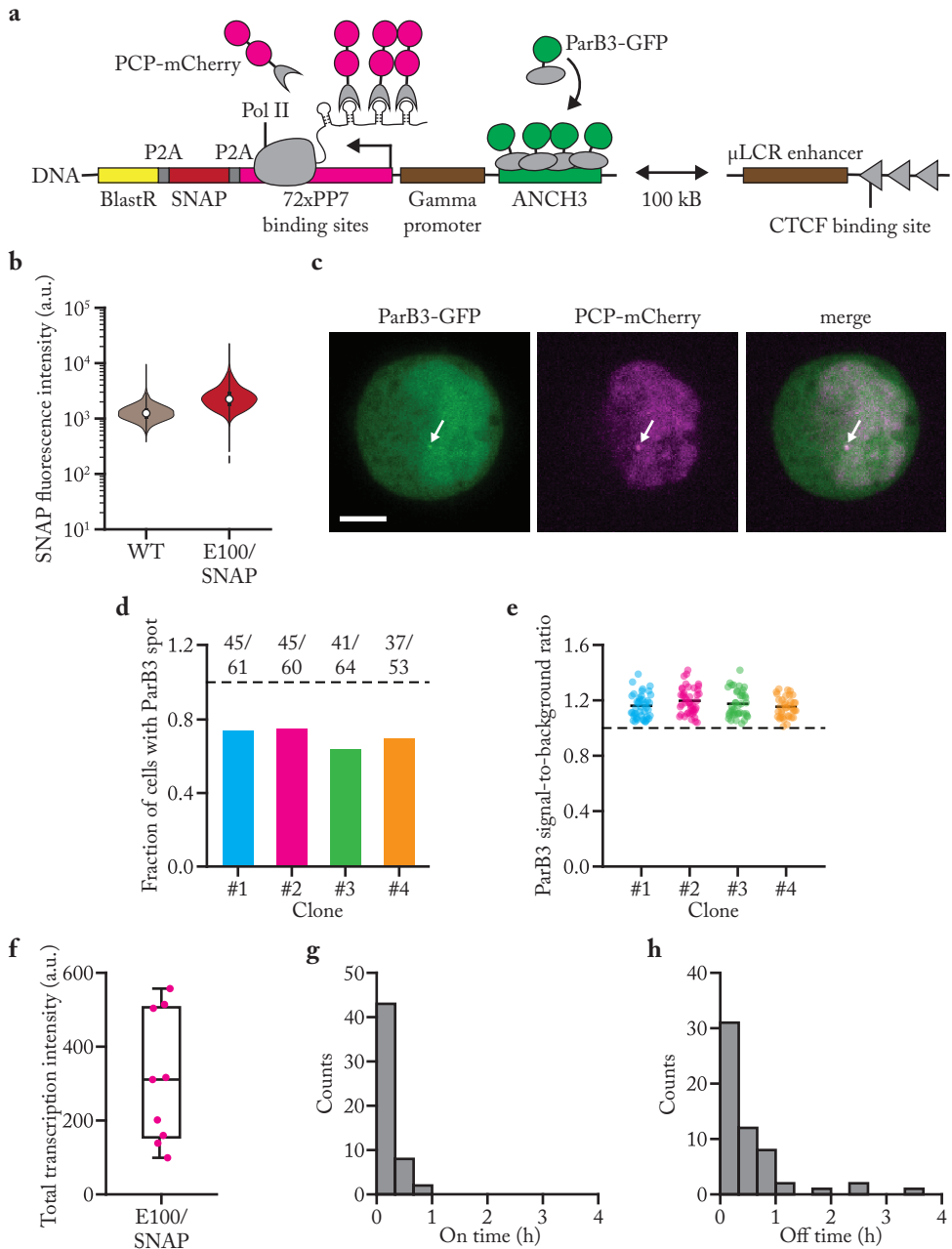
FACS analysis of the resulting cell line (referred to as the E100/eGFP cell line) showed high expression of eGFP compared to wild-type K562 cells (Supplemental Fig. 1a). Importantly, eGFP expression was enhancer dependent as deletion of the enhancer reverted eGFP fluorescence levels back to wild-type (data not shown; Rinzema et al., in preparation). To further confirm that eGFP expression was caused by promoter-enhancer interactions, chromosome conformation capture (4C) experiments were performed that showed specific contacts between the  $\gamma$ -globin promoter and the  $\mu$ LCR enhancer (data not shown; Rinzema et al., in preparation). Taken together, this enhancer reporter system provides a good model system to study promoter-enhancer interactions using a live-cell imaging approach.

### 5.3.2 Modifying the enhancer reporter system to visualize promoter location and activity

To visualize promoter-enhancer interactions and follow the transcriptional activity of the promoter, we adapted the enhancer reporter system in two steps. In the first step we introduced tools to visualize the promoter location and activity (Fig. 1a). To visualize the promoter location we adopted the ANCHOR DNA labeling system (specifically ANCHOR3). The ANCHOR3 system is based on the insertion of a short (~ 1kB) non-repetitive DNA sequence (ANCH3) that is recognized by OR proteins (ParB3) [32]. Despite the fact that the ANCH3 platform contains few ParB3 protein binding sites, fluorescent foci are readily visible by fluorescent microscopy thanks to the ability of the ParB3 protein to oligomerize around the binding site. To visualize promoter activity we introduced PP7 hairpins at the start of our reporter gene. Upon transcription of the PP7 hairpins, fluorescently labeled PCP molecules directly bind to these hairpins, resulting in a bright fluorescent spot that is visible by fluorescent microscopy. Taken together, by employing the ANCHOR3 and PP7 labeling systems, we aim to simultaneously visualize promoter location and activity.

To introduce the visualization tools in E100/eGFP cells, we designed a targeting strategy (see Methods) in which the  $\gamma$ -globin promoter and eGFP reporter gene is replaced by a novel cassette containing the ANCH3 platform immediately adjacent

to the  $\gamma$ -globin promoter, which drives expression of a reporter gene containing 72xPP7 repeats, a Blastidicin resistance gene (for antibiotic selection), and a SNAP tag (for protein quantification). Once we established a clonal cell line (referred to as the E100/SNAP cell line) and genetically confirmed the integration, we performed



FACS analysis to measure SNAP expression levels. After labeling E100/SNAP cells with SNAP-JF-646, FACS analysis showed significant expression of the SNAP tag compared to wild-type K562 cells (Fig. 1b), albeit that expression levels were much lower compared to E100/eGFP cells (compare Fig. 1b with Supplemental Fig. 1a). The low SNAP levels were not associated with the clonal cell line that we selected as multiple clonal cell lines, which were all genetically confirmed to contain our cassette, displayed similar or lower SNAP levels (data not shown).

To assess if the differences in protein expression level between the E100/eGFP and E100/SNAP cells is also present at the mRNA level, we performed single-molecule FISH (smFISH). Qualitatively in agreement with the FACS data, E100/SNAP cells contain less mRNA molecules compared to E100/eGFP cells (Supplemental Fig. 1b-d). These results suggest that either the transcription rate or mRNA stability of the reporter gene is lower in E100/SNAP cells compared to E100/eGFP cells. Quantitatively, the smFISH data cannot fully explain the differences observed with the FACS: while the number of mRNA molecules only differs ~5 fold between E100/eGFP and E100/SNAP cells, E100/eGFP cells have fluorescence levels that are ~50 fold above wild-type levels and E100/SNAP cells have fluorescence levels that are less than 2 fold above wild-type cells. One explanation for the differences on the protein level could be that a higher fraction of mRNA molecules resides in the cytoplasm in E100/eGFP cells compared to the E100/SNAP cells (i.e. it could be that the reporter gene in the E100/SNAP cells is retained more in the nucleus and therefore less translated.) (Supplemental Fig. 1e).

The FACS data also revealed a large proportion (more than 50%) of E100/SNAP cells that display SNAP fluorescence levels similar to wild-type K562 cells. We wondered if the cells with very low SNAP levels resemble cells in a temporary state or if a proportion of cells has definitively turned off the  $\gamma$ -globin promoter. To test

**Fig. 1 (opposite page) | An enhancer reporter system in K562 cells.**

**a**, Schematic of the enhancer reporter system and the tools used to visualize the location of the promoter and its transcriptional activity. **b**, Wild-type K562 and E100/SNAP cells were labeled with SNAP-JF-646 and the fluorescence levels were measured by FACS. Dot and thick error bar indicate median and interquartile range. **c**, Representative image of an E100/SNAP cell. Arrow indicates co-localizing ParB3-GFP and PCP-mCherry signal. Scale bar, 5  $\mu$ m. **d-e**, E100/SNAP cells were transduced with ParB3-GFP and four clones were selected. Fraction of cells with a ParB3 spot (d) and signal-to-background ratio of ParB3 spots (e) for the four different clones. (e) Line indicates the mean and dashed line indicates signal-to-background ratio of 1. **f**, Boxplot of the total intensity of the PCP-mCherry transcription spot for each cell during the four hour movie. **g**, Histogram of on times (i.e. the durations of transcriptionally active periods). **h**, Histogram of off times (i.e. the time between transcriptionally active periods).

this, we FACS sorted E100/SNAP cells in a bottom, middle, and top population (each population was ~33% of all cells) based on their SNAP fluorescence levels. We cultured the different populations for ~2 weeks and measured the SNAP fluorescence level of each population again after 3 and 17 days on FACS (Supplemental Fig. 1f-g). While small differences in the SNAP fluorescence levels between the three populations were present 3 days after sorting, all populations displayed very similar SNAP fluorescence levels after 17 days, suggesting that the cells in the bottom population resemble a temporary state and are not inherently different from the cells in the top population.

### 5.3.3 Visualizing promoter location and activity

Next, we wanted to test if the ANCHOR and PCP labeling systems would allow us to track the promoter location and activity. To this end, we transiently expressed ParB3-GFP and PCP-mCherry in E100/SNAP cells and imaged the cells. While most cells did not display ParB3 or PCP spots, we found some cells that showed a ParB3 and PCP spot in close proximity to each other (Fig. 1c). Importantly, most cells contained only one or two ParB3 spots, in agreement with a single integration event. As we only found few cells with ParB3 and PCP spots after transient expression of ParB3-GFP and PCP-mCherry, we decided to further investigate which expression levels of ParB3-GFP and PCP-mCherry are optimal for imaging. We generated a series of clones with varying ParB3-GFP levels or PCP-mCherry levels. For the ParB3-GFP clones, we selected 4 clones with the best ParB3 spots and quantified in each clone the fraction of cells with a ParB3 spot and the signal-to-background ratio (Fig. 1d-e). Not all the cells displayed a ParB3 spot (~25-35% of the cells did not have a ParB3 spot), possibly this is caused by removal of ParB3 proteins from the DNA, for instance by the replication machinery.

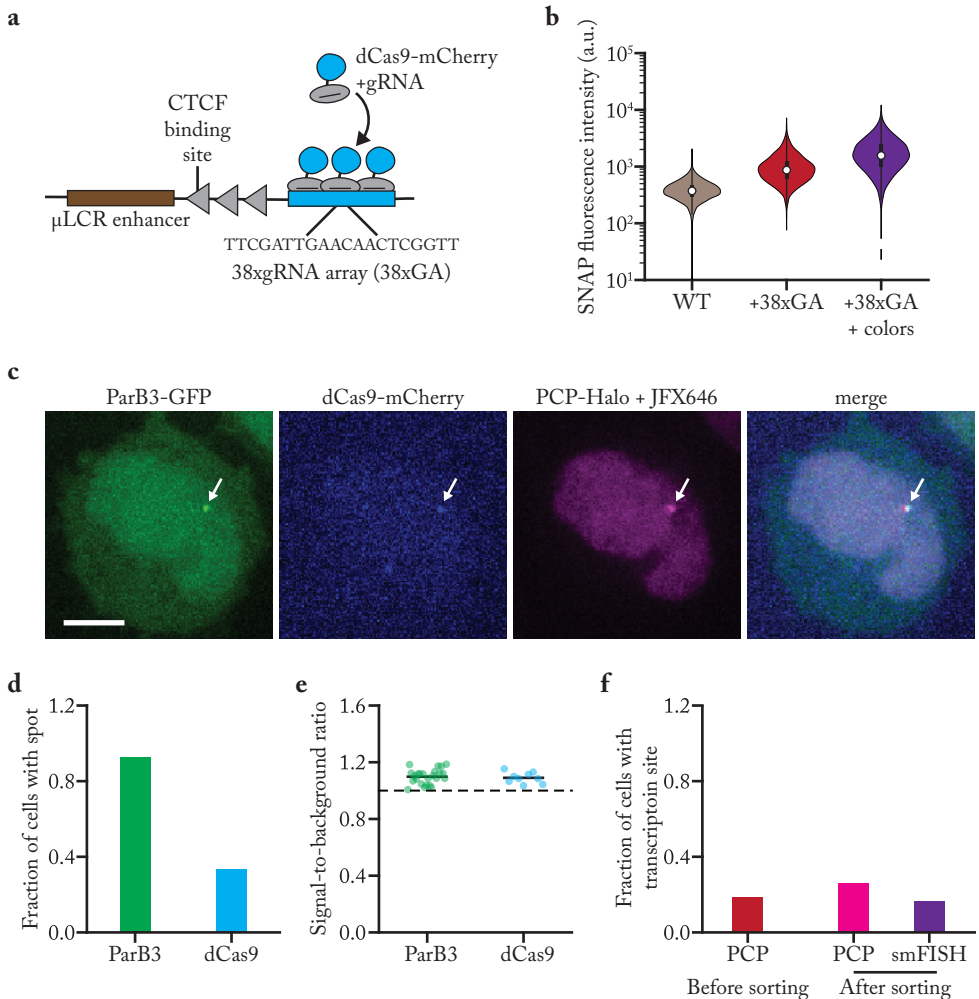
For the PCP-mCherry clones, we selected a single clone that displayed a single bright PCP-mCherry spot that fluctuated in intensity over time, indicative of nascent transcription. To characterize the transcriptional dynamics of our reporter gene, we followed cells for 4 hours and quantified the transcription rate as measured by the presence of a PCP-mCherry spot. Interestingly, we found extensive heterogeneity in transcriptional activity between different cells during the entire 4 hours (Fig. 1f and Supplemental Fig. 1h-i). We quantified the duration of active (on time) and inactive (off time) periods (see Methods) and found a narrow distribution of on times (<20 minutes) (Fig. 1g). For the duration of inactive periods we found a somewhat similar distribution, except for a tail towards longer times (inactive periods could take up to 3 hours) (Fig. 1h). These results are reminiscent of a recent study which reported prolonged inactive periods ranging from minutes to days for the *TFF1* gene [37].

Although more data is required to obtain a detailed view of the transcriptional dynamics of our reporter gene, these initial data suggest that E100/SNAP cells transition between transcriptional active and inactive states over longer periods of time (i.e. potentially on the >hour time scale).

### 5.3.4 Visualizing promoter-enhancer interactions

To visualize promoter-enhancer contacts, we required a visualization tool to follow the location of the enhancer as well. To this end, we decided to use CRISPR-Cas9 imaging, in which a catalytically inactive version of Cas9 (dCas9) is fluorescently labeled and recruited to a specific region in the genome. CRISPR-Cas9 can be readily used to visualize repetitive regions in the DNA, since a single gRNA can recruit multiple dCas9 molecules to the repetitive sequence [38]. On the other hand, imaging non-repetitive regions requires the introduction of multiple gRNAs (at least 12) and often results in a lower signal-to-noise ratio compared to inserted arrays [31,39]. Therefore, we decided to design an array (referred to as the 38xGA platform) that contained 38 repeats of a short sequence that could be targeted by a single gRNA (Fig. 2a).

We designed a targeting strategy to integrate the 38xGA directly downstream of the 3 CTCF binding sites (Fig. 2a, see Methods). We decided to insert the 38xGA platform downstream of the enhancer and CTCF sites to prevent interference of the 38xGA platform with promoter-enhancer communication. We established a clonal cell line (referred to as the E100/SNAP-38xGA cell line), and confirmed correct integration by PCR. We performed FACS analysis to measure the SNAP fluorescence levels and found that expression of the reporter gene was not affected by introduction of the 38xGA next to the  $\mu$ LCR enhancer (compare Fig. 2b with Fig. 1b). To visualize promoter-enhancer contacts and transcriptional activity of the gene promoter simultaneously, we transduced E100/SNAP-38xGA cells with ParB3-GFP, dCas9-mCherry, gRNA-BFP, and PCP-Halo. We selected clonal cell lines that were positive for gRNA-BFP, contained ParB3-GFP and PCP-Halo levels that were comparable to the levels we established in the E100/SNAP cells, and with a wide range of dCas9-mCherry fluorescence levels. After screening, we selected a clone in which we could visualize triple localization of a ParB3, dCas9, and PCP spot (Fig. 2c). We verified that these cells were still expressing the reporter gene (Fig. 2b) and quantified the fraction of cells in which we could detect a ParB3 spot and/or dCas9 spot (Fig. 2d-e). We found that in ~90% of the cells we could detect a ParB3 spot, while only ~30% of cells displayed a dCas9 spot. The lower percentage of cells with a dCas9 spot is likely caused by the low signal intensity of the dCas9 spot, suggesting that less dCas9 molecules compared to ParB3 molecules are recruited to



**Fig. 2 | Visualizing promoter-enhancer communication.**

**a**, Schematic of the tool used to visualize the location of the  $\mu$ LCR enhancer. **b**, Cells were labeled with SNAP-JF-646 and the fluorescence levels were measured by FACS. Dot and thick error bar indicate median and interquartile range. **c**, Representative image of an E100/SNAP-38xGA cell. Arrow indicates co-localization of ParB3-GFP, dCas9-mCherry, and PCP-Halo. Cells were labeled with Halo-JFX-650. Scale bar, 5  $\mu$ m. **d-e**, A single E100/SNAP-38xGA clone was selected that expressed ParB3-GFP, dCas9-mCherry, gRNA-BFP, and PCP-Halo. Fraction of cells with a ParB3 and dCas9 spot (**d**) and signal-to-background ratio of ParB3 and dCas9 spots (**e**). **f**, Transcriptional activity of E100/SNAP-38xGA cells before or after sorting strategy (see Methods). Transcriptional activity was determined by live-cell imaging of PCP-Halo and calculating the fraction of cells with a PCP spot that co-localized with a ParB3 spot. For cells after sorting, transcriptional activity was also determined by smFISH using probes against the SNAP reporter. Fraction of cells was determined for which a co-localizing ParB3 and smFISH spot was observed.



the binding platforms.

While maintaining E100/SNAP-38xGA cells over time, we noticed that SNAP fluorescence levels decreased over time and even reverted back to wild-type K562 levels (Supplemental Fig. 2a). This effect is likely caused by epigenetic silencing of the region in which we integrated our enhancer reporter system. We also observed this phenotype at several other occasions. For instance, after transduction of E100/SNAP cells with ParB3-GFP, some clonal cell lines were found to display SNAP fluorescence levels comparable to wild-type K562 cells (data not shown). Similarly, after integration of the 38xGA platform in E100/SNAP cells, several clones (in which the integration was genetically confirmed) contained SNAP fluorescence levels comparable to wild-type K562 cells (data not shown). Although not extensively tested, we did not notice a reduction of SNAP fluorescence levels in E100/SNAP cells, suggesting that the silencing may be induced by the introduction of repetitive sequences (e.g. the 38xGA platform) or DNA binding proteins (e.g. ParB3-GFP or dCas9-mCherry).

Since epigenetic silencing is known to affect chromosome topology [40], this effect could make it difficult to interpret any results on measuring promoter-enhancer contacts. Especially when measuring promoter-enhancer distances in transcriptionally inactive cells, it could be that these cells are (1) un-looped and therefore inactive or (2) epigenetically silenced and therefore inactive. As a consequence, the measured promoter-enhancer distance could reflect the un-looped state or the epigenetically silenced state. To alleviate the epigenetic silencing, we designed a sorting strategy in which we (FACS) sort E100/SNAP-38xGA cells with SNAP fluorescence levels that exceed wild-type K562 SNAP fluorescence levels (threshold was set at >99th percentile of K562 wild-type cells). We reasoned that these cells, as they display SNAP fluorescence levels, are not epigenetically silenced at the moment of sorting. By performing the experiments within a short time window after the sorting (i.e. within one week), we expect that only a small fraction of the cells will be epigenetically silenced during the analysis.

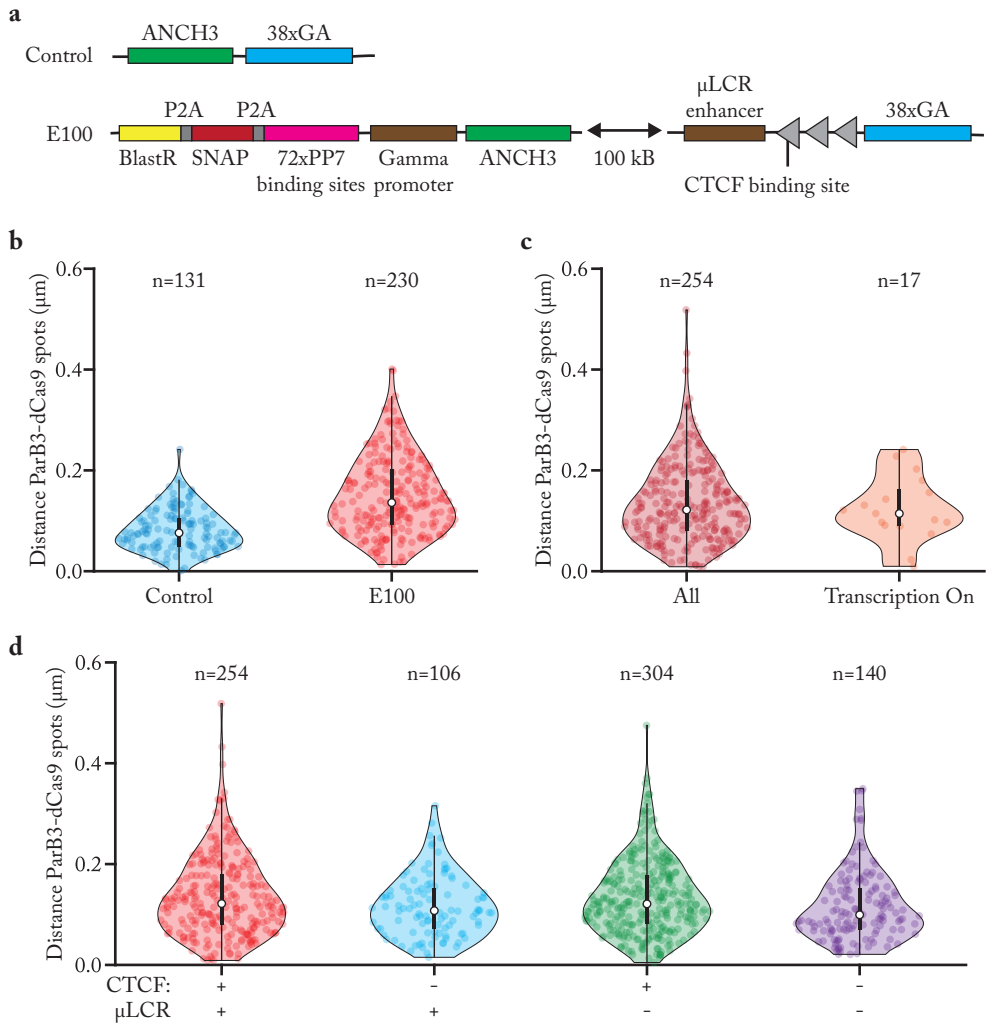
To test how gene expression of the reporter gene is affected by this sorting strategy, we sorted E100/SNAP-38xGA cells on SNAP fluorescence levels and performed FACS analysis, smFISH, and live-cell PCP-Halo imaging four days after sorting. SNAP fluorescence levels were higher compared to unsorted cells (Supplemental Fig. 2b) and the number of mRNA molecules per cell was higher compared to unsorted E100/SNAP cells (Supplemental Fig. 2c). Using the PCP-Halo live-cell imaging, we found that also an increased fraction of the cells showed active transcription after sorting (Fig. 2f). Using the smFISH, we could also quantify the fraction of cells

in which a ParB3 spot co-localized with a smFISH signal (Supplemental Fig. 2d), indicative of nascent transcription. The values found in the live-cell imaging were in a similar range as found in smFISH (Fig. 2f), indicating that live-cell imaging of PCP-Halo is a sensitive method to detect nascent transcription. Together, these results further suggest that cells can switch between transcriptionally active and inactive states over durations that can last several days.

### **5.3.5 Promoter-enhancer distance does not correlate with promoter activity**

Several studies have measured promoter-enhancer distance in relation to the transcriptional activity of the gene promoter [33,34]. Interestingly, while one study found a strong correlation between the promoter-enhancer distance and transcriptional activation, another study did not find this correlation. To investigate if transcriptional activation and promoter-enhancer proximity are correlated, we set out to measure the promoter-enhancer distance in cells that are either transcriptionally active or inactive (please note that all experiments described from here onwards did not employ the sorting strategy described in the previous section). When measuring promoter-enhancer distance, not only the physical distance between the promoter and enhancer contributes to the measured distance, but also other factors such as imaging errors (i.e. due to chromatic aberrations) and localization errors. Moreover, ParB3-GFP and dCas9-mCherry proteins are recruited to two distinct platforms and not to the promoter and enhancer directly. Assuming that the centers of the ParB3 and dCas9 foci are located in the middle of their respective platforms, this will result in an 'extra distance' of ~1.5 kB, which could add significantly to the measured promoter-enhancer distance (assuming a distance of 0.34 nm between base pairs this could add ~500 nm).

To take in account the physical distance between the ANCH3 and 38xGA platforms, we generated a control cell line which contained a reporter in which the ANCH3 platform and the 38xGA platform were inserted directly adjacent to each other (ANCH3-38xGA reporter) (Fig. 3a, top). After quantification of the distance between ParB3 and dCas9 foci in control cells (see Methods), we found a narrow distribution centered around ~76 nm (Fig. 3b) which represents thus the minimal distance between a ParB3 and dCas9 spot. Next, we measured the distance between the ParB3 and dCas9 spots in E100/SNAP-38xGA cells and found a median distance of ~136 nm (Fig. 3b). The median ParB3-dCas9 distance is slightly larger compared to control cells ( $p$ -value < 0.05), suggesting that the promoter and enhancer don't make a direct contact, although the measured ParB3-dCas9 distance could also be affected by the orientation in which the promoter and enhancer interact.



**Fig. 3 | Promoter-enhancer communication does not correlate with promoter activity.**

a, Schematic of the control reporter (top) and the E100 reporter (bottom). b-d, The distance between the ParB3-GFP and dCas9-mCherry spot (see Methods). Dot and thick error bar indicate median and interquartile range.

The distribution of ParB3-dCas9 distances in E100/SNAP-38xGA cells was wider compared to control cells, suggesting that the promoter and enhancer take on a wide variety of configurations.

The relatively wide distribution of promoter-enhancer distances in E100/SNAP-38xGA cells suggests that the promoter and enhancer take on different configurations. We wondered if those configurations would be predictive of the transcriptional status

of the cell. To investigate this, we simultaneously measured the promoter-enhancer distance and the transcriptional status of the promoter. After labeling the cells with Halo-JFX-650, we found that only few cells (~10%) contained PCP-Halo spots (Supplemental Fig. 3a). To increase the number of cells with an active transcription site, we pre-treated E100/SNAP-38xGA cells for two days with hemin. Treatment of K562 cells with hemin results in erythroid differentiation [41–43] and upregulation of many genes including our reporter gene (upregulation is mediated through the  $\mu$ LCR enhancer, data not shown). Interestingly, in a first experiment in which hemin was removed ~2h before live-cell imaging a similar fraction of cells compared to untreated cells displayed a PCP-Halo spot (Supplemental Fig. 3b, cells 1-84). In a second experiment, during which cells were continuously treated with hemin, a larger fraction of cells did show a PCP-Halo spot (Supplemental Fig. 3b, cells 85-111). Although more data would be required, these results suggest that hemin directly affects  $\mu$ LCR activity. Next, we measured the promoter-enhancer distance in presence and absence of hemin, and found that both distributions were very similar (Supplemental Fig. 3c). As still only a small number of cells displayed a PCP-Halo spot (8 cells in the absence of hemin and 9 cells in the presence of hemin), we decided to pool both data-sets together and compare the cells that are transcriptionally active with all other cells. Comparison of transcriptionally active cells with all other cells revealed very similar promoter-enhancer distances (Fig. 3c). Thus, this initial data-set suggests that there is no correlation between promoter-enhancer distance and the transcriptional status of the cell.

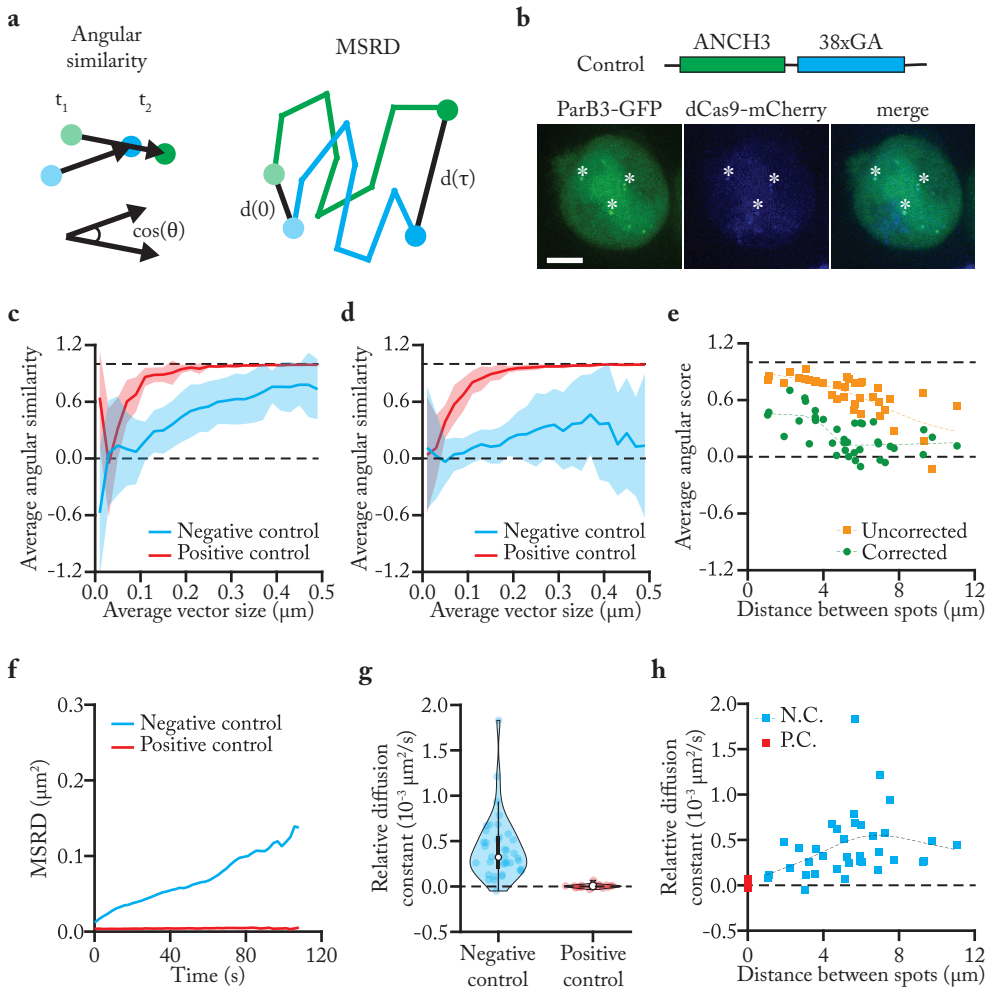
The  $\gamma$ -globin promoter and  $\mu$ LCR enhancer are thought to be brought in close proximity through DNA looping. Both the  $\mu$ LCR enhancer and CTCF binding sites play an important role in DNA looping, and in E100/eGFP cells removal of both results in loss of interactions between the  $\gamma$ -globin promoter and the surrounding region as measured by 4C (data not shown). To investigate the role of the  $\mu$ LCR enhancer and CTCF binding sites in the distance between the ParB3 and dCas9 spots, we removed the  $\mu$ LCR enhancer and/or the CTCF binding sites. We transfected E100/SNAP-38xGA cells with the Cre and FLP recombinases and selected 2-4 clones in which the  $\mu$ LCR enhancer, the CTCF binding sites, or both were removed. We measured the distance between the ParB3 and dCas9 spots in these clones, but surprisingly did not find any significant differences in the measured distances (Fig. 3d and Supplemental Fig. 3d-f). We also measured SNAP fluorescence levels and found that all clones had fluorescence levels comparable to wild-type K562 cells (Supplemental Fig. 3g). While removal of the  $\mu$ LCR enhancer is expected to result in loss of SNAP fluorescence levels, this is not expected for cells in which only the CTCF binding sites were removed. These results suggest that at least in the cells in which the CTCF binding sites were removed, the reporter gene has been

epigenetically silenced, making it difficult to interpret these results. A solution for this problem would be to employ the sorting strategy that was described above (e.g. see Supplemental Fig. 2b), subsequently remove the  $\mu$ LCR enhancer and/or CTCF binding sites and perform the distance measurements directly.

### 5.3.6 Co-diffusion as a more sensitive measurement for promoter-enhancer interactions

Measuring the distance between promoter and enhancer did not show a clear correlation between the distance and the transcriptional activity of the gene promoter. The failure to detect any differences in the distance could suggest that transcriptional activation is not mediated through conformational changes of the promoter and enhancer. Alternatively, it could mean that the measurements of promoter-enhancer distance are not sensitive enough to detect the conformational changes that underlie transcriptional activation. One model for transcriptional activation is that the promoter and enhancer already form a 'pre-looped' state and that transcriptional activation is the result of minor reconfigurations that establish a 'functional' contact between the promoter and enhancer [1]. We reasoned that such small changes may remain undetectable by measuring promoter-enhancer distance because of the large variability observed in distances (e.g. see Fig. 3b). To obtain a more sensitive measurement of a promoter-enhancer interaction, we hypothesized to measure the co-movement or co-diffusion of the promoter and enhancer over time. We reasoned that a promoter and enhancer that interact together would show increased co-movement or co-diffusion, while a promoter and enhancer that do not interact would move independently and thus show a reduced co-movement or co-diffusion.

To measure the level of co-diffusion between the promoter and enhancer, we developed two independent analysis methods that are based on the angular similarity (Fig. 4a, left) and the mean square relative displacement (Fig. 4a, right), respectively (see Methods). Briefly, the angular similarity reports on the direction of two vectors; in case two vectors point in the same direction the angular similarity is 1 and in case two vectors point in opposite direction the angular similarity is -1. By following promoter and enhancer movement over time, we can compute vectors for the movement of the promoter and enhancer over time and determine the angular similarity between the vectors at each time point. In case the promoter and enhancer move independently, the angular similarity is expected to be random and the average angular similarity will be  $\sim 0$ . In contrast, in case the promoter and enhancer move together the angular similarities are expected to be predominantly around 1 and as a result the average angular similarity will be  $\sim 1$  as well. Thus, by measuring the average



angular similarity we can determine the extent of co-diffusion for two spots (i.e. between 0 for no co-diffusion and 1 for perfect co-diffusion).

While the angular similarity provides a sensitive measurement of co-diffusion, one disadvantage is that the angular similarity is affected by cellular drift (i.e. if both spots move in the same direction due to cellular drift, they will display an average angular similarity close to 1 and appear to co-diffuse). As K562 cells are expected to drift while tracking the promoter and enhancer over time, measuring a parameter for co-diffusion that is independent of cellular drift is preferred. One parameter that is not affected by cellular drift is the mean square relative displacement (MSRD) [44]. The MSRD reports on how the relative distance between two spots changes over time. In case two spots move independently, the MSRD is expected to increase linearly with

**Fig. 4 (opposite page) | Co-diffusion as a new measure for promoter-enhancer interactions.**

**a**, Schematic of the angular similarity and the mean square relative displacement (MSRD) method to measure the co-diffusion of two spots. **b**, Schematic of control reporter (top) and a representative image of a control cell (bottom). Image is a maximum intensity projection of 5 slices. Asterisks indicate co-localizing ParB3-GFP and dCas9-mCherry spots. Scale bar, 5  $\mu\text{m}$ . **c-d**, The average angular similarity before (c) and after (d) drift correction for the negative control (spots at different loci in the same cell) and positive control (spots at the same locus in the same cell) (see Methods). Solid lines and corresponding shaded regions represent mean  $\pm$  SEM. For the negative control the co-diffusion was determined for 38 spot pairs and for the positive control for 18 spot pairs. Dashed lines indicate score associated with no co-diffusion (at 0) or perfect co-diffusion (at 1). **e**, The average angular score (see Methods) for the negative control for different distances between the two loci. Each point is the average angular score of two spots and yellow and green dashed lines are a smoothing spline fitted to all points. Black dashed lines indicate score associated with no co-diffusion (at 0) or perfect co-diffusion (at 1). **f**, The average MSRD for the negative control (n=38) and positive control (n=18). **g**, The relative diffusion constant as calculated from the MSRD for each pair of spots. Dashed line indicates relative diffusion constant of 0. Dot and thick error bar indicate median and interquartile range. **h**, The relative diffusion constant for each pair of spots for different distances between the loci. Each point is the relative diffusion constant of two spots and thin dashed line is a smoothing spline fitted to negative control points. Thick dashed line indicates relative diffusion constant of 0.

a rate proportional to the average diffusion constant of the two spots. In contrast, in case the two spots move together the MSRD is expected to remain at 0. Thus, the MSRD reports on the co-diffusion of two spots independent of cellular drift.

Next, we set out to validate both analysis methods. For validation, we made use of the ANCH3-38xGA control cells that contain the reporter in which the ANCH3 platform and the 38xGA platform are inserted directly adjacent to each other (Fig. 4b, top). We found that control cells contained multiple integrations of the ANCH3-38xGA reporter as we detected many co-localizing ParB3 and dCas9 spots in a single cell (Fig. 4b, bottom). This feature allowed us to examine the co-diffusion between ParB3 and dCas9 spots at the same locus, which act as a positive control for co-diffusion, and between ParB3 and dCas9 spots at different loci within the same cell, which act as a negative control for co-diffusion. To measure the co-diffusion of positive control spot pairs and negative control spot pairs, we performed dual-camera time-lapse imaging, following ANCH3-38xGA cells for 50 time points with an interval of  $\sim 2.25$  seconds. We obtained images in 5 z-planes and made maximum intensity projections.

We computed the average angular similarity for negative and positive control spots and plotted the average angular similarity at different vector sizes (i.e. for this the average movement of the ParB3 and dCas9 spots was computed) (Fig. 4c). At low vector sizes ( $< 50$  nm), we observed that both for negative and positive control spots,

the average angular similarity is  $\sim 0$ , indicative of no co-diffusion. At these small movements, vectors are likely to be dominated by localization errors and are thus expected to show no co-diffusion. At larger vector sizes, we measured higher average angular similarities for positive control spots compared to negative control spots, suggesting that positive control spots co-diffuse stronger than negative control spots. However, negative control spots also displayed average angular similarities larger than 0, which is likely caused by cellular drift as negative control spots are not expected to display any co-diffusion.

To compensate for the cellular drift, we decided to employ an *in silico* drift correction algorithm. For drift correction we used the ParB3 spots as fiducial markers to compute the translational and rotational drift between subsequent time points (Supplemental Fig. 4a-c, see Methods). After drift correction, we again computed the average angular similarity for negative and positive control spots (Fig. 4d), and found a reduction of the average angular similarity for negative control spots. Importantly, drift correction did not reduce the average angular similarity for positive control spots, indicating that positive control spots display real co-diffusion. Next, we wanted to compute the amount of co-diffusion for individual spot pairs. As such, we computed for each spot pair an average angular similarity score (see Methods) and found that drift correction improved the separation between negative and positive control spot pairs (Supplemental Fig. 4d-e). Interestingly, for negative control spot pairs we observed a strong correlation between the distance between the two spots and the average angular score (Fig. 4e). This correlation was observed both before and after drift correction, suggesting that local movement of the nucleus or chromatin contributes to the average angular similarity score.

Next, we computed the time-averaged MSRD for negative and positive control spots (Fig. 4f and Supplemental Fig. 4f-g). As expected negative control spot pairs displayed a linear increase of MSRD over time, while positive control spot pairs displayed no increase in MSRD. To compare negative and positive control spot pairs, we computed for each MSRD a relative diffusion constant based on the first 5 time points of the MSRD (see Methods) and found a clear difference in the relative diffusion constant of negative and positive control spot pairs (Fig. 4g). Based on the time-averaged MSD of individual traces of the ParB3 and dCas9 spots, we also computed before and after drift correction the diffusion constants of individual ParB3 and dCas9 spots (Supplemental Fig. 4h-i). As expected, we found that the relative diffusion constant was very similar to the individual diffusion constants of ParB3 and dCas9 spots after drift correction (compare Fig. 4g and Supplemental Fig. 4i). Finally, for the MSRD we also observed a correlation between the distance between a negative control spot pair and the relative diffusion constant, suggesting



that spots close to each other display higher co-diffusion than spots that are far apart from each other (Fig. 4h).

Taken together, these validation experiments have shown that both the angular similarity and the MSRD analysis methods can distinguish between negative and positive control spot pairs. The next step will be to see if these methods are able to reveal differences in promoter-enhancer co-diffusion in transcriptionally active and inactive cells.

## 5.4 Discussion

The mechanisms that underlie promoter-enhancer communication are still elusive. One important question that remains unanswered is whether stable or dynamic promoter-enhancer contacts drive transcriptional activation. To distinguish between both models approaches are necessary that provide high spatial and temporal resolution. Here, we have developed a live-cell imaging approach in human K562 cells that allows us to visualize promoter-enhancer contacts and transcriptional activity of the gene promoter simultaneously.

Initial analysis of promoter-enhancer distance in transcriptionally active and inactive cells revealed that the promoter-enhancer distance is very similar in active and inactive cells, in agreement with a previous study performed in mouse embryonic stem cells [34]. The failure to detect a significant difference in promoter-enhancer distance in transcriptionally active and inactive cells raises the question what event triggers transcriptional activation. We consider in total 4 explanations that could be responsible for these results. Firstly, it could be that promoter-enhancer interactions are not formed in E100/SNAP cells as we observed that reporter gene expression is relatively low (i.e. maybe the low SNAP levels the E100/SNAP cells display, are enhancer-independent). We also observed that cells are susceptible to (epigenetic) silencing of the reporter gene over time, which may alter chromosome topology. Together, this may compromise our results. Secondly, it could be that promoter-enhancer contacts are short-lived and result in delayed transcriptional activation (the hit-and-run model). In such a scenario promoter-enhancer distance and transcriptional activity may be poorly correlated. Thirdly, promoter and enhancer topology may only undergo minor conformational changes that result in transcriptional activation. Such minor changes may be below the sensitivity threshold when measuring promoter-enhancer distance. Finally, other molecular events such as transcription factor recruitment could form the trigger for transcriptional activation. Below, we will discuss these four possibilities in more detail.

### 5.4.1 Reporter gene expression is low and susceptible to silencing

FACS analysis of E100/SNAP cells revealed lower reporter gene expression levels compared to E100/eGFP cells. The low expression levels were confirmed using smFISH, suggesting that the transcription rate or the stability of the reporter gene is lower in E100/SNAP cells compared to E100/eGFP cells. In case the low expression levels are caused by a reduced transcription rate, the low activity of the  $\gamma$ -globin promoter raises the question if promoter activity is still enhancer mediated in E100/SNAP cells. If promoter activity is no longer enhancer mediated, this could explain the absence of a correlation between promoter-enhancer distance and transcriptional activity. To test this hypothesis, a chromatin conformation capture (4C) experiment could be employed to investigate if promoter-enhancer contacts are still made in E100/SNAP cells. We believe, however, that most likely the 4C experiment in E100/SNAP cells will not differ significantly from E100/eGFP cells, because we did observe that in a large fraction of the cells (>50%) the promoter and enhancer come close together (within 150 nm). A previous study showed that in case of Hi-C, the strongest interactions seen correspond to co-localizations within 150 nm in ~20% of the cells [45], which would comply with our distance measurements.

We also observed that some clonal cell lines lost their SNAP expression over time, an effect which is likely due to epigenetic silencing of the reporter gene. As the integration site of the enhancer reporter system is in a region of the chromosome that is devoid of other (active) genes, this area may be especially susceptible to epigenetic gene silencing. Reporter gene silencing is a problem in analyzing the results from the  $\mu$ LCR enhancer and CTCF binding site knock-out experiments. As deletion of the  $\mu$ LCR enhancer also results in loss of SNAP expression, it is no longer possible to assess if epigenetic silencing of the region has occurred. Since epigenetic silencing could potentially have major effects on chromosome topology, this makes it difficult to interpret the results from these experiments. A potential solution for this problem could be to employ the sorting strategy. In this strategy, E100/SNAP cells are first sorted on SNAP expression to exclude potentially silenced cells, subsequently the  $\mu$ LCR enhancer is removed and cells are directly imaged. By minimizing the time between  $\mu$ LCR enhancer removal and imaging, the effects of gene silencing may be minimized.

### 5.4.2 A more sensitive method to measure promoter-enhancer contacts

One model for promoter-enhancer communication is that brief contacts between the promoter and enhancer result in transcriptional activation (i.e. the hit-and-run model). This may also result in a time delay between formation of the promoter-enhancer contact and transcriptional activation. In addition, a recent study proposed

that promoter-enhancer contacts are being memorized into longer-lived promoter states [46], although the exact mechanism of memory is unclear. Together, short interactions and long-lived effects from the interaction may explain why we don't observe a correlation between promoter-enhancer distance and transcriptional activity. To investigate this model further, promoter-enhancer distance and transcriptional activity need to be followed simultaneously for longer time periods. Interestingly, a recent study followed the distance between two linked genes that are regulated by the same enhancer [35]. They found several instances in which the two transcription spots were separated for the entire duration of a transcriptionally active period, suggesting that transcription is not triggered by transient promoter-enhancer interactions. However, in their assay they could not exclude the formation of a promoter-enhancer contact before transcriptional activation, so it will be interesting to follow promoter-enhancer distance before and during transcription in our enhancer reporter system.

Transient promoter-enhancer contacts could explain why we don't observe a correlation between promoter-enhancer distance and transcriptional activity. Another reason why our measurements are not sensitive enough is that the distance measurements are not accurate enough to detect differences between the un-looped and the looped state. It has been proposed that a promoter and enhancer engage in a pre-looped state in which they are brought into close proximity before transcriptional activation [1]. Subsequently, small conformational changes in the promoter and enhancer topology may result in the formation of a functional promoter-enhancer contact. In this model it could very well be that in the pre-looped state and in the looped state, promoter-enhancer distances are in a very similar range. Taking in account the imaging and localization errors in measuring the distance, it will be difficult to detect changes that occur in the <25 nm range. To reduce imaging errors and localization errors, it is possible to correct for chromatic shifts and to average multiple measurements of promoter-enhancer distance, respectively [47,48].

Another approach to increase the sensitivity of detecting different promoter-enhancer conformations is to develop new analysis methods that provide a more sensitive read-out of promoter-enhancer interactions. Here, we hypothesized that measuring the co-movement or co-diffusion between the promoter and enhancer may provide such a sensitive method. We have validated our analysis method and showed that we can distinguish between two spots that move independently and two spots that are at the same locus. One worry is that we observed a strong correlation between the distance between the spots and the amount of co-diffusion (see Fig. 4e,h). These results suggest that the measured co-diffusion originates to some extent from local nuclear and chromatin movement and raises the question if we have enough sensitivity to distinguish different promoter-enhancer topologies. For this,

it will be best to compare E100/SNAP-38xGA cells with and without the  $\mu$ LCR enhancer, because based on 4C data this is expected to have the largest impact on chromosome topology.

### 5.4.3 What is the trigger for transcriptional activation?

The absence of correlation between promoter-enhancer distance and the transcriptional status of a cell may be caused by the cells currently used or by a lack of sensitivity. However, another explanation could be that it is not the formation of a promoter-enhancer contact that is the trigger for transcriptional activation but it is another molecular event such as the recruitment of an essential transcription factor. We observed extensive heterogeneity in transcriptional activity between cells that we followed for a 4 hour period (see Fig. 1f) and found that cells sorted on high SNAP fluorescence levels still showed increased transcriptional activity 4 days after sorting (see Fig. 2f). Together, these results suggest that cells transition between periods of active and inactive transcription on the time scale of minutes to hours. On the other hand, DNA loop formation, which is thought to be a main driver of promoter-enhancer contacts, is a dynamic process during which loops are continuously formed and broken [45,49–51], and CTCF and cohesin, two proteins that play a key role in DNA looping, have a residence time on DNA of  $\sim$ 1-2 minutes and  $\sim$ 22 minutes, respectively [52]. While the data on the transcription dynamics of our reporter gene needs to be further substantiated, the initial data suggests that the formation of promoter-enhancer contacts may be an unlikely trigger for transcriptional activation.

If promoter-enhancer contacts are made continuously, what determines then the ‘success rate’ of a contact that is made? Interestingly, a recent study that followed nascent transcription of the *TFF1* gene under endogenous regulation also observed extensive heterogeneity in the duration of inactive transcription periods, ranging from minutes to days [37]. Using a computational model, they found that transcriptional dynamics could be best captured using a 3 gene state model: (1) a repressive state, (2) a poised un-looped state, and (3) an active looped state. In their model the rate-limiting step was the transition from the repressive state to the poised state, which had an average transition rate of  $\sim$ 43 days. The molecular state of the repressive state remained elusive, though, and to test this hypothesis it will be interesting to see if we can visualize the chromatin state around our promoter and enhancer.

Taken together, while the mechanisms underlying transcriptional activation and the role of promoter-enhancer contacts remain elusive, we think we have established a live-cell imaging approach with which we can hopefully answer some of the questions regarding promoter-enhancer communication.

## 5.5 Methods

### 5.5.1 Cell culture and cell lines

K562 cells were cultured in RPMI 1640 medium (Gibco) supplemented with 10% FBS (Sigma-Aldrich) and 1% penicillin/streptomycin (Gibco). Cells were incubated at 37°C and 5% CO<sub>2</sub>. In case of hemin treatment, cells were incubated for two days with hemin (Sigma-Aldrich) at a final concentration of 30 μM. HEK293T (ATCC) were grown in DMEM (4.5 g/L glucose, Gibco) containing 5% fetal bovine serum (Sigma-Aldrich) and 1% penicillin/streptomycin (Gibco). Cells were incubated at 37°C and 5% CO<sub>2</sub>.

To generate ANCH3-38xGA control cells, wild-type K562 cells were transfected with two plasmids: one plasmid containing the ANCH3 and 38xGA platforms (and a puromycin resistance gene) flanked by inverted terminal repeats, and a second plasmid containing the piggyBac transposase. To select for stable integration, cells were selected for 2 weeks on puromycin at a final concentration of 1 μg/ml. Cells were subsequently transduced with 2 lentiviruses expressing ParB3-GFP, and gRNA-GA-BFP, respectively. GFP and BFP positive cells were isolated by FACS and used for further downstream analysis. One day before each experiment, cells were transfected with a plasmid containing dCas9-mCherry.

E100/SNAP and E100/SNAP-38xGA cells were generated using a CRISPR knock-in strategy (see section 5.5.2). Subsequently, the visualization tools (i.e. ParB3-GFP, dCas9-mCherry, gRNA-GA-BFP, and PCP-mCherry/PCP-Halo) were introduced either by transient transfection or transduction.

### 5.5.2 Targeting strategy

For imaging promoter-enhancer communication, we made use of the previously established E100/eGFP cell line (Rinzema et al., in preparation). In this cell line, an eGFP reporter gene, together with the  $\gamma$ -globin gene promoter, is integrated at chromosome 18 (hg coordinates: chr18:22,029,048) in human K562 cells. The  $\mu$ LCR enhancer, a 6.5-kb form of the endogenous  $\beta$ -globin LCR (Locus Control Region, ~20 kb), is integrated at approximately 100 kb downstream of the  $\gamma$ -globin promoter and 3 CTCF binding sites are positioned directly next of the enhancer. To visualize promoter-enhancer contacts and transcription, we needed to integrate visualization tools for the promoter and enhancer location, and for transcriptional activity. To this end, we designed a CRISPR knock-in strategy that was composed of two steps; in the first step we knocked-in the ANCH3 targeting construct, enabling visualization of the promoter location and of the transcriptional activity. In the second step we knocked-in the 38xGA targeting construct, enabling visualization of the enhancer location. The design of the targeting constructs and knock-in strategies is discussed in detail below.

#### *ANCH3 targeting strategy*

For the first CRISPR knock-in step we designed a targeting construct (referred to as the ANCH3 targeting construct) that allowed visualization of the promoter location and of transcriptional activity. The ANCH3 targeting construct contained the following elements, in sequential order: the ANCH3 sequence (a kind gift from K. Bystricky), the  $\gamma$ -globin promoter, 72xPP7 hairpin repeats, a blasticidin resistance gene, and a SNAP-tag sequence. The 72xPP7 hairpin repeats, blasticidin resistance gene, and SNAP-tag sequence were separated by 2 P2A sequences. To integrate the ANCH3 targeting construct in E100/eGFP cells, we selected two sgRNAs that deleted the  $\gamma$ -globin promoter and eGFP reporter gene. We introduced these sgRNAs with wild-type Cas9 into E100/eGFP cells by transfection. We then sorted a bulk population of cells that was negative for eGFP (referred to as  $\Delta$ eGFP cells). Next, we designed two new sgRNAs and two homology arms flanking the ANCH3 targeting construct that would allow for integration of the ANCH3-PP7 targeting construct to the introduced cut sites by

homology directed repair. We delivered the 4 plasmids (containing the two sgRNAs, wild type Cas9, and the ANCH3 targeting construct) into  $\Delta$ eGFP cells by transfection and sorted single cells by FACS that were positive for SNAP (using SNAP-JF-646), and screened for positive clones by PCR.

### *38xGA targeting strategy*

For the second step we designed a targeting construct (referred to as the 38xGA targeting construct) that allowed us to visualize the enhancer location. To this end, we synthesized the 38xGA array (Genewiz) containing a repetitive sequence to which either a fluorescently labeled TALE or dCas9 protein can bind. The repeats are based on the GATTGAACAACCTCGGT sequence targetable by an engineered TALE protein [53], modified to include 5 NGG PAM sequences. The final repeat sequence was CCACCTAAGTTCGATTGAACAACCTCGGTTAGGTGG. We designed an array containing 38 copies of the repeat sequence and each copy was separated by a random 15 nt linker. To integrate the 38xGA targeting construct in E100/SNAP cells, we selected a sgRNA that targeted a sequence directly upstream of the 3 CTCF sites (which directly flank the  $\mu$ LCR enhancer). We designed two homology arms flanking the 38xGA targeting construct allowing for integration at the cut site by homology directed repair. We delivered the 3 plasmids (containing the sgRNA, wild-type Cas9 and the 38xGA targeting construct) into E100/SNAP cells by transfection. We sorted single cells by FACS that were positive for SNAP (using SNAP-JF-646), and screened for positive clones by performing a PCR. To confirm integration of the 38xGA targeting construct on the correct allele (i.e. the allele that contained the  $\mu$ LCR enhancer and the ANCH3 targeting construct), we used in the PCR one primer complementary to the 38xGA targeting construct and one primer complementary to the 3xCTCF sites. Finally, we expressed ParB3-GFP, PCP-Halo, gRNA-GA-BFP, and dCas9-mCherry in the positive clones and selected a single clone (referred to as E100/SNAP-38xGA cells) that displayed clear ParB3, PCP, and dCas9 spots.

### **5.5.3 K562 transfections**

Plasmids were transfected by nucleofection into K562 cells using AMAXA Nucleofector II, program T-003, using sterile nucleofection buffer composed of 90 mM sodium phosphate buffer pH 7.2, 5 mM KCl, 10 mM MgCl<sub>2</sub>, 20 mM HEPES pH 7.2; final pH 7.2. Per nucleofection, we used 1 million cells and 5  $\mu$ g DNA.

### **5.5.4 Lentivirus production and infection**

For lentivirus production, HEK293T cells were plated in a 6-well plate (Greiner Bio-one) at 30% confluency, and transfected 24 hr after plating with a mixture of 50  $\mu$ l OptiMEM (Sigma-Aldrich), 10  $\mu$ l Polyethylenimine (PEI) (Polysciences Inc) (1mg/ml), 0.4  $\mu$ g pMD2.g, 0.6  $\mu$ g psPAX2, and 1  $\mu$ g of lentiviral vector. The medium was replaced with 2 ml fresh culture medium 24 hr after transfection, and 72 hr after transfection, viral supernatant was collected. In case of dCas9-mCherry, the viral supernatant was further concentrated using an Amicon Ultra-4 100 kDa ultrafiltration device (Sigma-Aldrich).

For lentiviral infections, cells were seeded in a 6-well plate (Greiner Bio-one) at 70% confluency. Viral supernatant was added to the cells along with Polybrene (10  $\mu$ g/ml) (Santa Cruz Biotechnology Inc) and the cells were spun at 2000 rpm for 90 min at 37°C (spin infection). After the spin infection, the culture medium was replaced with fresh medium, and cells were incubated for at least 48 hr before further analysis.

### **5.5.5 Isolating genomic DNA from clones in 96 wells plates**

DirectPCR Lysis Reagent (Viagen Biotech, Catalog No: 301-C) was used to isolate genomic DNA from clones in 96 wells plates. Briefly, 50-100  $\mu$ l Viagen/Prot K mix (ProtK stock 10 mg/mL, at 50X dilution) was added to each well and cells were resuspended. Plates were sealed with 3M aluminum foil

tape (3M HD 431), placed at 55°C for 1 hr, then at 85°C for 30 min. 2  $\mu$ L lysate was used as template for each PCR.

### 5.5.6 FACS

In case of analyzing Halo or SNAP levels using FACS, cells were labeled either with Halo-JFX-650 or SNAP-JF-646, respectively, at a final concentration of 50 nM for at least 30 minutes and up to 2 hours. After labeling, cells were re-suspended in fresh medium or in PBS prior to analysis on the FACS.

### 5.5.7 smFISH

Single-molecule Fluorescence In Situ Hybridization (smFISH) was performed as described previously [54,55]. We designed 48 oligonucleotide probes (20 nt in length) against the eGFP reporter gene and 48 probes against the SNAP reporter gene, using the Stellaris Probe Designer. Since the eGFP reporter gene was too short to contain 48 probes, we designed some probes to be partially overlapping. Oligonucleotide probes were ordered from IDT and labeled using terminal deoxynucleotidyl transferase (ThermoFisher) and an Amino-11-ddUTP (Lumiprobe) coupled to an ATTO 633 NHS ester (Sigma-Aldrich), as described before [56]. Probes were dissolved in 50  $\mu$ L TAE and used at a final dilution of 1:2000.

For hybridization, all steps were performed in suspension and after each step cells were centrifuged at 200 g for 4 minutes. Cells were fixed in PBS with 4% paraformaldehyde (Electron Microscopy Science) for 15 minutes at Room Temperature (RT), washed twice with PBS and incubated for 30 min in 100% ethanol at 4°C. After fixation, cells were washed twice in wash buffer with 10% formamide (ThermoFisher) at RT, followed by overnight incubation with the probes in hybridization buffer at 37°C. Following overnight incubation, samples were washed 2x for 1 hour in wash buffer at 37°C. DAPI (Sigma-Aldrich) was added to the final wash step in order to stain the nuclei. Finally, cells were placed in anti-bleach buffer [54,55] to reduce fluorescence bleaching, and plated on a poly-L-lysine (Sigma-Aldrich) coated 96-well microscopy plate (Matriplate, Brooks).

### 5.5.8 Microscopy

All imaging experiments were performed using a Nikon TI inverted microscope with perfect focus system equipped with a Yokagawa CSU-X1 spinning disc, a 100x 1.49 NA objective and two iXon Ultra 897 EM-CCD cameras (Andor) using NIS elements software. All live-cell experiments were performed at 37°C, while smFISH experiments were imaged at RT.

For live-cell imaging experiments, cell culture medium was replaced with pre-warmed CO<sub>2</sub>-independent Leibovitz's-15 medium (Gibco) containing 5% fetal bovine serum (Sigma-Aldrich) and 1% penicillin and streptomycin (Gibco). Camera exposure times of 100 ms were used for BFP (405 laser), GFP (488 laser), mCherry (561 laser), and Cy5 (644 laser), and images were acquired in multiple Z-planes with an inter-slice distance of 0.8  $\mu$ m each. For transcription imaging (Fig. 1f-h and Supplemental Fig. 1h-i), images were acquired every 5 minutes for 4 hours. For co-diffusion experiments (Fig. 4 and Supplemental Fig. 4), images were obtained with 2 cameras, enabling simultaneous acquisition of GFP and mCherry images. A total of 50 images in 5 Z-planes with an inter-slice distance of 0.8  $\mu$ m each were acquired with an interval time of ~2.25 s. After acquisition, maximum intensity projections were generated in NIS elements and used for downstream analyses.

For the smFISH experiments, images for all 3 colors (DAPI, ParB3-GFP, and ATTO 633) were acquired with a camera exposure time of 100 ms. Z-stacks were acquired for all 3 colors with an inter-slice distance of 0.8  $\mu$ m each. After acquisition, maximum intensity projections were generated in NIS elements and used for further analysis.

### 5.5.9 Quantification of smFISH experiments

We used TransTrack [57] for quantification of the smFISH experiments. First, we selected cells and identified the nucleus based on DAPI. Finally, we quantified the number of mRNA molecules based on the smFISH in the cytoplasm and nucleus in each cell.

### 5.5.10 Transcription dynamics

We first generated maximum intensity projections. For each cell that we could track for the entire 4 hour duration of the movie, we manually determined at each time point if a transcription spot was present. Next, we measured the PCP-mCherry intensity in an ROI with a width and height of 3 pixels centered around the PCP-mCherry spot. For frames in which no PCP-mCherry spot was detected, we extrapolated the location based on the previous and next spot location. For each spot, local background fluorescence intensity was measured using a second ROI at multiple locations directly next to the spot of interest, and the mean background intensity was subtracted from the mean spot intensity. Finally, we determined the duration of transcriptionally active and inactive periods (on and off times, respectively) by computing the length of periods that were manually annotated as transcriptionally active (on times) and the time between periods that were manually annotated as transcriptionally active (off times). We also included the start and the end of the movie (e.g. if the movie started with an inactive period we computed the time from the start of the movie until the start of the active period as the off time).

### 5.5.11 Measuring spot intensity

To determine the signal to background ratio (Fig. 1e and 2e), intensities were measured in FIJI in an ROI with a width and height of 4 pixels. The measured intensities were divided by the background intensity, which was measured using a second ROI (with the same dimensions) at a location directly next to the spot of interest. To determine the peak to background ratio, an ROI with a width and height of 20 pixels was drawn around the spot of interest, and the maximum intensity in the ROI was divided by the mean intensity of the four ROI corner pixels.

### 5.5.12 Subpixel localization and distance measurements

To localize ParB3 and dCas9 spots, we made use of an algorithm that employs radial symmetry to provide an accurate estimation of the subpixel localization of a spot [58]. We selected an ROI with a width and height of 20-30 pixels, centered around the location of the ParB3 and dCas9 spots, and used the radial symmetry algorithm to determine the subpixel location of both spots. Finally, we computed the distance between the ParB3 and dCas9 spots, using Equation 5.1, and converted the distance from pixels to  $\mu\text{m}$ .

$$\text{distance} = \sqrt{(x_{\text{ParB3}} - x_{\text{dCas9}})^2 + (y_{\text{ParB3}} - y_{\text{dCas9}})^2} \quad (\text{Equation 5.1})$$

### 5.5.13 Drift correction

For the co-diffusion analysis in ANCH3-38xGA control cells we performed a drift correction to compensate for cellular drift. For drift correction we made use of the fact that in a single ANCH3-38xGA control cell multiple ParB3 spots are visible. As such, we reasoned that we could use the ParB3 spots as fiducial markers to correct for cellular drift. We corrected for drift by comparing the locations of the ParB3 spots between the first time point and any other time point. To this end, we computed a cost function (Equation 5.2) for each time point.

$$\text{cost}(n) = \sum_{i=1}^m \sqrt{(x_i^{t_n} - x_i^{t_1})^2 + (y_i^{t_n} - y_i^{t_1})^2} \quad (\text{Equation 5.2})$$

Here, the distance between the location of a ParB3 spot at time point  $t_n$  and  $t_j$  is computed for all ParB3 spots present in the cell (where  $m$  is the number of spots in the cell), and all the distances are



summed together to compute the cost score at time point  $n$ . Next, we minimized the cost function for each time point by applying an image transformation (with translation parameters  $t_x$  and  $t_y$ , and rotation parameter  $\theta$ ) and finding the transformation parameters at which the cost function was minimized. The minimization search was performed in two steps. In the first step a fast and bounded search was employed around the transformation parameters of the previous time point (to drift correct the second time point, transformation parameters of 0 were used as a starting value). In the second step, a Matlab minimization function was employed for a refined search in which the parameters found in step 1 were used as initial values for the minimization function. In this way, we obtained for each time point the transformation parameters that minimized the cost function for that particular time point. These transformation parameters were then used to drift correct the images and the subpixel localizations of ParB3 and dCas9 spots.

#### 5.5.14 MSD and diffusion constant

We determined the time-averaged mean squared displacement (MSD) for uncorrected and drift corrected tracks. We computed the MSD by using equation 5.3.

$$\text{MSD}(\tau) = \langle [r(t+\tau) - r(t)]^2 \rangle \quad (\text{Equation 5.3})$$

Here,  $r$  is the position of a particle and  $\tau$  is the lag time between the two time points. For a single track we calculated for each lag time the average squared displacement. From the MSD we also computed the diffusion constants of particles. As the slope of the MSD directly relates to the diffusion constant (Equation 5.4, two-dimensional diffusion), we used the slope between time points 1 and 5 to compute the diffusion constant (since the start of the track is more reliable, we didn't use later time points).

$$\text{MSD}(t) = 4 \cdot D \cdot t \quad (\text{Equation 5.4})$$

#### 5.5.15 Angular similarity

To measure the co-movement or co-diffusion between two spots, we designed a method that measured the angular similarity between two traces. The angular or cosine similarity is defined as follows (Equation 5.5):

$$\text{Angular similarity} = \cos\theta = (A \cdot B) / (\|A\| \cdot \|B\|) \quad (\text{Equation 5.5})$$

Here,  $\theta$  is the angle between two vectors  $A$  and  $B$ . If two vectors point in the same direction,  $\theta$  is  $0^\circ$  and the angular similarity is  $\sim 1$  (i.e. cosine of  $0^\circ$  is 1), while if two vectors point in opposite direction,  $\theta$  is  $180^\circ$  and the angular similarity is  $\sim -1$  (i.e. cosine of  $180^\circ$  is  $-1$ ). We reasoned that if two spots move independently, the direction of the two vectors is expected to be uncorrelated. Hence, the angular similarity is randomly distributed between  $-1$  and  $1$ , and the average angular similarity should yield a result of  $\sim 0$ . On the other hand, if two spots move together, the direction of the two vectors should be predominantly in the same direction. Computing the average angular similarity is then expected to yield a result close to  $\sim 1$ . Thus, by computing the average angular similarity it is possible to distinguish if two spots move together or not.

We used the angular similarity method to measure the co-diffusion of spots in ANCH3-38xGA control cells. As a negative control we measured the angular similarity between ParB3 and dCas9 spots that were in the same cell but not at the same locus and as a positive control we measured the angular similarity between ParB3 and dCas9 spots that were in the same cell and at the same locus. We computed the angular similarity for uncorrected and drift corrected tracks. For each trajectory we computed the angular similarity between each time point, and we calculated the average length of the two vectors. Finally, for each trajectory we determined the average angular score by computing the average angular

similarity of all data points which had an average vector size of at least 150 nm.

### 5.5.16 MSRD

To measure the co-diffusion between two spots, we also designed a second method that measures the time-averaged mean squared relative distance (MSRD) [44]. The MSRD is a measure much similar to the MSD, except that it uses the relative distance between two particles. As a result, the MSRD is insensitive to cellular drift as this does not affect the distance between two particles (as long as only translational and rotational drift is taking place). For two spots that move independently, the distance between the two spots is expected to increase on average at a rate that is proportional to the average diffusion constant of the two spots. On the other hand, if two spots move together, the two spots are expected to remain at the same distance from each other. Therefore, the MSRD is expected to remain zero. The time-averaged MSRD can be computed by using equation 5.6:

$$\text{MSRD}(\tau) = \langle [d_{\text{ParB3} - \text{dCas9}}(t+\tau) - d_{\text{ParB3} - \text{dCas9}}(t)]^2 \rangle \quad (\text{Equation 5.6})$$

Here,  $d_{\text{ParB3} - \text{dCas9}}$  is the distance between the ParB3 and dCas9 spots and  $\tau$  is the lag time between the two time points. For every lag time, we average the relative distance (i.e. the change in distance between the ParB3 and dCas9 spots during this lag time) for all possible starting points along the trajectory.

We used the MSRD method to measure the co-diffusion of spots in ANCH3-38xGA control cells. As a negative control we computed the MSRD between ParB3 and dCas9 spots that were in the same cell but not at the same locus and as a positive control we computed the MSRD between ParB3 and dCas9 spots that were in the same cell and at the same locus. For all negative control spot pairs, we also measured the distance between the ParB3 and dCas9 spots. For all spot pairs we also computed the average diffusion constant based on the slope of the MSRD curve between time points 1 and 5.

## 5.6 Acknowledgements

We would like to thank Kerstin Bystricky for kindly sharing the ANCHOR system plasmids and advice for working with them. We also thank Ive Logister and the Hubrecht Institute FACS facility for sorting. We would like to thank Alexander van Oudenaarden for the recommendation on using the mean square relative distance as a measure for co-diffusion. We would like to thank members of the Tanenbaum and de Laat labs for helpful discussions. All authors are supported by the Oncode Institute that is partly funded by the Dutch Cancer Society (KWF).

## 5.7 Author contributions

S.S., S.T., E.V., W.L., and M.T. conceived the experiments. S.T. and E.V. designed genome editing, screening, and isolation of K562 clones. S.T. and E.V. performed FACS experiments. S.S. and S.T. performed microscopy and computational analyses. N.R., P.K., and C.V.Q. generated the original E100/eGFP cell line. S.S., S.T., E.V., W.L., and M.T. wrote the manuscript.

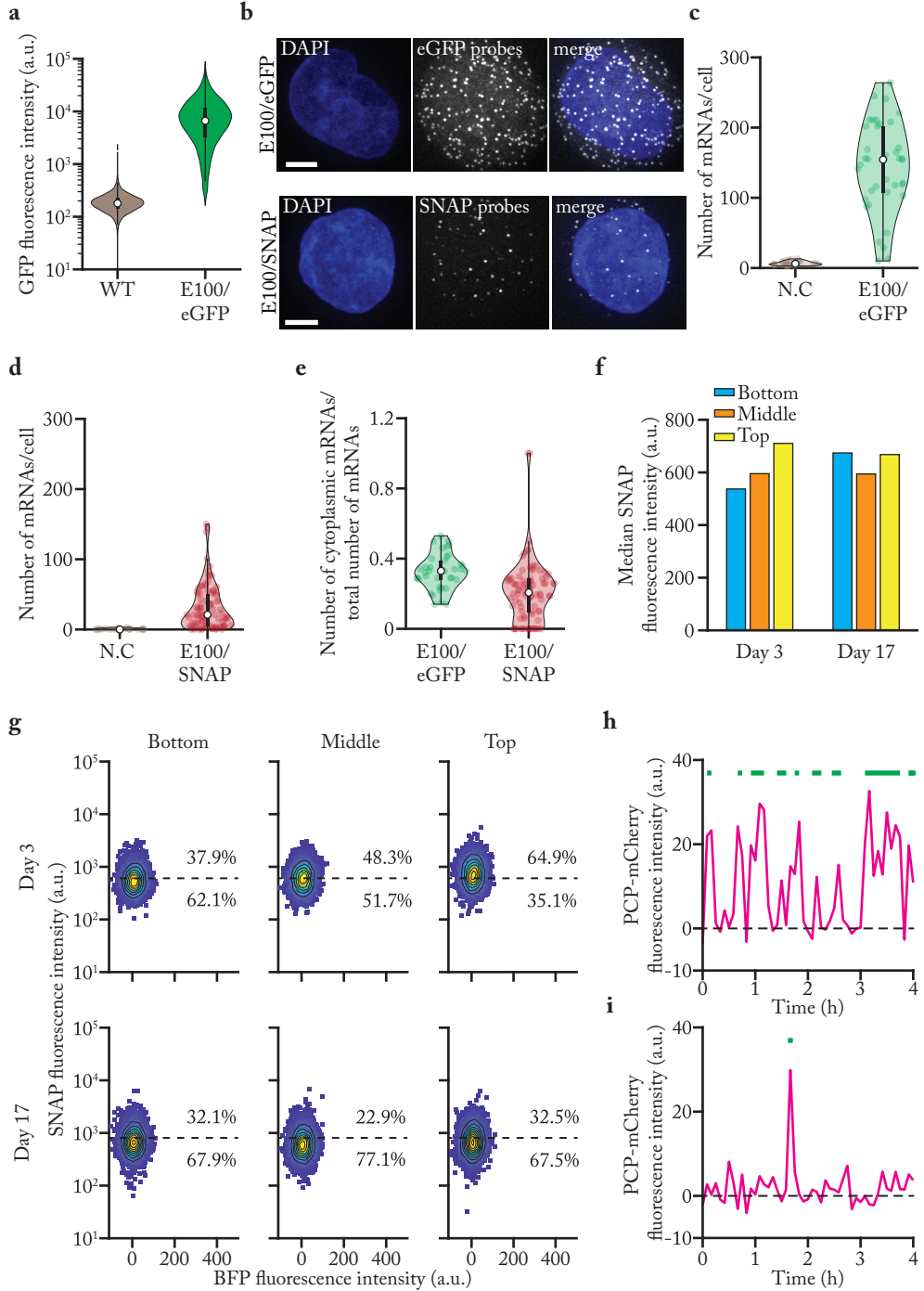
## 5.8 References

- 1 Furlong, E.E.M. and Levine, M. (2018) Developmental enhancers and chromosome topology. *Science* 361, 1341–1345
- 2 Gillies, S.D. et al. (1983) A tissue-specific transcription enhancer element is located in the major intron of a rearranged immunoglobulin heavy chain gene. *Cell* 33, 717–728
- 3 Kong, S. et al. (1997) Transcription of the HS2 enhancer toward a cis-linked gene is independent of the orientation, position, and distance of the enhancer relative to the gene. *Mol. Cell. Biol.* 17, 3955–3965
- 4 Serfling, E. et al. (1985) Enhancers and eukaryotic gene transcription. *Trends Genet.* 1, 224–230
- 5 The ENCODE Project Consortium (2012) An integrated encyclopedia of DNA elements in the human genome. *Nature* 489, 57–74
- 6 Shen, Y. et al. (2012) A map of the cis-regulatory sequences in the mouse genome. *Nature* 488, 116–120
- 7 Levine, M. (2010) Transcriptional enhancers in animal development and evolution. *Curr. Biol.* 20, R754–R763
- 8 Long, H.K. et al. (2016) Ever-changing landscapes: transcriptional enhancers in development and evolution. *Cell* 167, 1170–1187
- 9 Spielmann, M. et al. (2018) Structural variation in the 3D genome. *Nat. Rev. Genet.* 19, 453–467
- 10 Smith, E. and Shilatifard, A. (2014) Enhancer biology and enhanceropathies. *Nat. Struct. Mol. Biol.* 21, 210–219
- 11 Tolhuis, B. et al. (2002) Looping and interaction between hypersensitive sites in the active  $\beta$ -globin locus. *Mol. Cell* 10, 1453–1465
- 12 Carter, D. et al. (2002) Long-range chromatin regulatory interactions in vivo. *Nat. Genet.* 32, 623–626
- 13 Sanyal, A. et al. (2012) The long-range interaction landscape of gene promoters. *Nature* 489, 109–113
- 14 Mifsud, B. et al. (2015) Mapping long-range promoter contacts in human cells with high-resolution capture Hi-C. *Nat. Genet.* 47, 598–606
- 15 Bartman, C.R. et al. (2016) Enhancer regulation of transcriptional bursting parameters revealed by forced chromatin looping. *Mol. Cell* 62, 237–247
- 16 Deng, W. et al. (2012) Controlling long-range genomic interactions at a native locus by targeted tethering of a looping factor. *Cell* 149, 1233–1244
- 17 Deng, W. et al. (2014) Reactivation of developmentally silenced globin genes by forced chromatin looping. *Cell* 158, 849–860
- 18 Morgan, S.L. et al. (2017) Manipulation of nuclear architecture through CRISPR-mediated chromosomal looping. *Nat. Commun.* 8, 1–9
- 19 Benabdallah, N.S. et al. (2019) Decreased enhancer-promoter proximity accompanying enhancer activation. *Mol. Cell* 76, 473–484
- 20 Fukaya, T. et al. (2016) Enhancer control of transcriptional bursting. *Cell* 166, 358–368
- 21 Lim, B. et al. (2018) Visualization of transvection in living *Drosophila* embryos. *Mol. Cell* 70, 287–296
- 22 Boija, A. et al. (2018) Transcription factors activate genes through the phase-separation capacity of their activation domains. *Cell* 175, 1842–1855
- 23 Sabari, B.R. et al. (2018) Coactivator condensation at super-enhancers links phase separation and gene control. *Science* 361, eaar3958
- 24 Cho, W.K. et al. (2018) Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. *Science* 361, 412–415
- 25 Chong, S. et al. (2018) Imaging dynamic and selective low-complexity domain interactions that control gene transcription. *Science* 361, eaar2555

- 26 Han, X. et al. (2020) Roles of the BRD4 short isoform in phase separation and active gene transcription. *Nat. Struct. Mol. Biol.* 27, 333–341
- 27 Cisse, I.I. et al. (2013) Real-time dynamics of RNA Polymerase II clustering in live human cells. *Science* 341, 664–667
- 28 Hnisz, D. et al. (2017) A phase separation model for transcriptional control. *Cell* 169, 13–23
- 29 Lakadamyali, M. and Cosma, M.P. (2020) Visualizing the genome in high resolution challenges our textbook understanding. *Nat. Methods* 17,
- 30 Sato, H. et al. (2020) Imaging of DNA and RNA in Living Eukaryotic Cells to Reveal Spatiotemporal Dynamics of Gene Expression. *Annu. Rev. Biochem.* 89, 159–187
- 31 Gu, B. et al. (2018) Transcription-coupled changes in nuclear mobility of mammalian cis-regulatory elements. *Science* 359, 1050–1055
- 32 Germier, T. et al. (2017) Real-Time Imaging of a Single Gene Reveals Transcription-Initiated Local Confinement. *Biophys. J.* 113, 1383–1394
- 33 Chen, H. et al. (2018) Dynamic interplay between enhancer–promoter topology and gene activity. *Nat. Genet.* 50, 1296–1303
- 34 Alexander, J.M. et al. (2019) Live-cell imaging reveals enhancer-dependent Sox2 transcription in the absence of enhancer proximity. *Elife* 8, e41769
- 35 Heist, T. et al. (2019) Large distances separate coregulated genes in living *Drosophila* embryos. *Proc. Natl. Acad. Sci. U. S. A.* 116, 15062–15067
- 36 Tuan, D. et al. (1985) The “ $\beta$ -like-globin” gene domain in human erythroid cells. *Proc. Natl. Acad. Sci. U. S. A.* 82, 6384–6388
- 37 Rodriguez, J. et al. (2019) Intrinsic dynamics of an endogenous human gene reveal the basis of expression heterogeneity. *Cell* 176, 213–226.e18
- 38 Chen, B. et al. (2013) Dynamic imaging of genomic loci in living human cells by an optimized CRISPR/Cas system. *Cell* 155, 1479–1491
- 39 Brandão, H.B. et al. (2021) Tracking and interpreting long-range chromatin interactions with super-resolution live-cell imaging. *Curr. Opin. Cell Biol.* 70, 18–26
- 40 Mizuguchi, T. et al. (2014) Cohesin-dependent globules and heterochromatin shape 3D genome architecture in *S. pombe*. *Nature* 516, 432–435
- 41 Nakajima, O. et al. (1997) Hemin-induced erythroid differentiation of human myeloleukemia K562 cell line and its modification by bioresponse modifiers. *Cell. Mol. Biol. (Noisy-le-grand)*. 43, 115–134
- 42 Rutherford, T.R. et al. (1979) K562 human leukaemic cells synthesise embryonic haemoglobin in response to haemin. *Nature* 280, 164–165
- 43 Dean, A. et al. (1981) Induction of hemoglobin accumulation in human K562 cells by hemin is reversible. *Science* 212, 459–461
- 44 Pönisch, W. and Ziburdaev, V. (2018) Relative distance between tracers as a measure of diffusivity within moving aggregates. *Eur. Phys. J. B* 91,
- 45 Finn, E.H. et al. (2019) Extensive heterogeneity and intrinsic variation in spatial genome organization. *Cell* 176, 1502–1515
- 46 Zuin, J. et al. (2021) Nonlinear control of transcription through enhancer-promoter interactions. *bioRxiv*
- 47 Niekamp, S. et al. (2019) Nanometer-accuracy distance measurements between fluorophores at the single-molecule level. *Proc. Natl. Acad. Sci. U. S. A.* 116, 4275–4284
- 48 Matsuda, A. et al. (2018) Accurate and fiducial-marker-free correction for three-dimensional chromatic shift in biological fluorescence microscopy. *Sci. Rep.* 8, 1–3
- 49 Cattoni, D.I. et al. (2017) Single-cell absolute contact probability detection reveals chromosomes are organized by multiple low-frequency yet specific interactions. *Nat. Commun.* 8, 1753

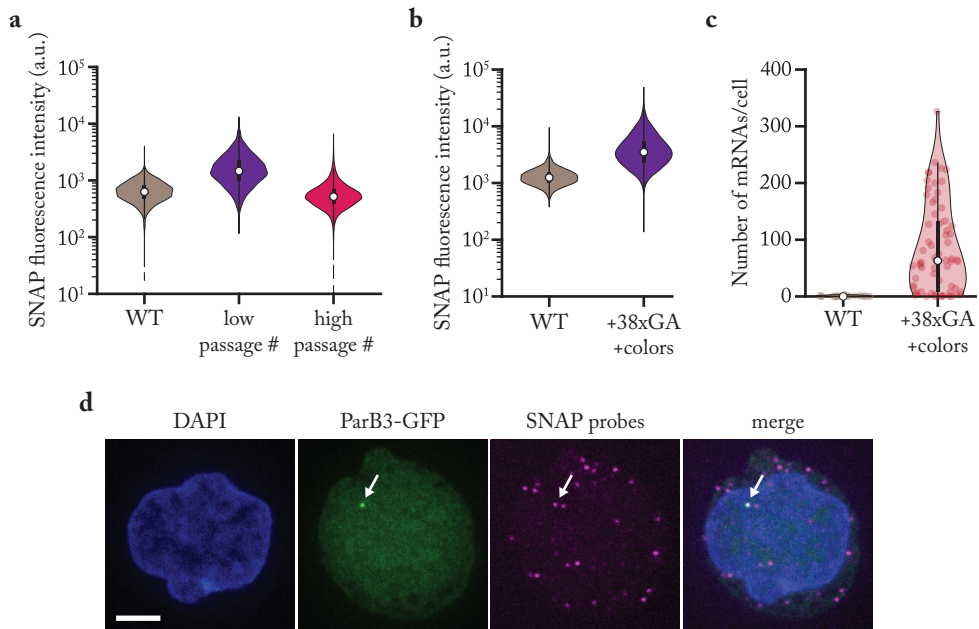
- 50 Nagano, T. et al. (2013) Single-cell Hi-C reveals cell-to-cell variability in chromosome structure. *Nature* 502, 59–64
- 51 Bintu, B. et al. (2018) Super-resolution chromatin tracing reveals domains and cooperative interactions in single cells. *Science* 362, eaau1783
- 52 Hansen, A.S. et al. (2017) CTCF and cohesin regulate chromatin loop stability with distinct dynamics. *Elife* 6, e25776
- 53 Aryan, A. et al. (2013) TALEN-based gene disruption in the dengue vector *Aedes aegypti*. *PLoS One* 8, e60082
- 54 Raj, A. et al. (2008) Imaging individual mRNA molecules using multiple singly labeled probes. *Nat. Methods* 5, 877–879
- 55 Lyubimova, A. et al. (2013) Single-molecule mRNA detection and counting in mammalian tissue. *Nat. Protoc.* 8, 1743–1758
- 56 Gaspar, I. et al. (2017) Enzymatic production of single-molecule FISH and RNA capture probes. *RNA* 23, 1582–1591
- 57 Boersma, S. et al. (2019) Multi-color single-molecule imaging uncovers extensive heterogeneity in mRNA decoding. *Cell* 178, 458–472
- 58 Parthasarathy, R. (2012) Rapid, accurate particle tracking by calculation of radial symmetry centers. *Nat. Methods* 9, 724–726

Supplemental figures



**Supplemental Fig. 1 (opposite page) | Gene expression analysis of E100/SNAP cells.**

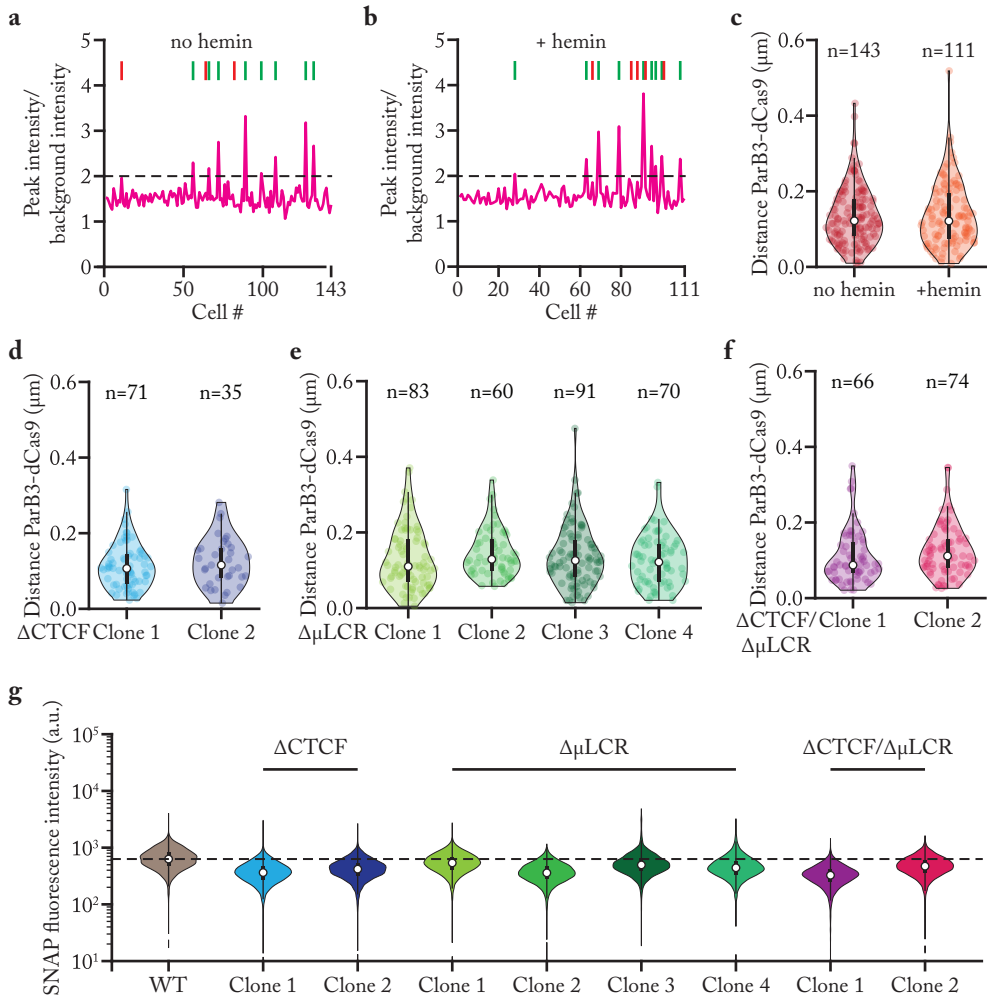
**a**, GFP fluorescence levels as measured by FACS for K562 wild-type and E100/eGFP cells. Dot and thick error bar indicates median and interquartile range. **b**, Representative images of cells incubated with smFISH probes targeting the eGFP reporter (top panel) or targeting the SNAP reporter (bottom panel). Scale bar, 5  $\mu\text{m}$ . **c-d**, Number of mRNA molecules in E100/eGFP (c) or E100/SNAP (d) cells ( $n=38$  and  $n=91$ , respectively) based on smFISH using probes against the eGFP or SNAP reporter, respectively. Negative control was determined by incubating E100/SNAP cells ( $n=30$ ) with probes against eGFP reporter (c) or E100/eGFP cells ( $n=23$ ) with probes against SNAP reporter (d). Dot and thick error bar indicate median and interquartile range. **e**, Fraction of mRNA molecules in cytoplasm based on smFISH using probes against eGFP (E100/eGFP) or SNAP (E100/SNAP) reporter. Dot and thick error bar indicate median and interquartile range. **f**, Median SNAP fluorescence levels as measured by FACS after labeling cells with SNAP-JF-646. E100/SNAP cells were first FACS sorted in a bottom, middle, and top fraction based on their SNAP fluorescence levels (each fraction was ~33% of the entire population). For each fraction, SNAP fluorescence levels were subsequently measured at day 3 and day 17. **g**, SNAP fluorescence levels as measured by FACS for bottom, middle and top fraction 3 days and 17 days after sorting. Dashed line indicates the maximum SNAP fluorescence level of K562 wild-type cells (<1% of wild-type cells has fluorescence levels higher than dashed line) and percentages indicate fraction of cells above or below dashed line. **h-i**, Representative PCP-mCherry intensity traces of a transcriptionally active (h) and transcriptionally inactive (i) cell. Green lines indicate time points during which a PCP-mCherry spot was detected. At time points when no PCP-mCherry spot was detected, PCP-mCherry intensity was measured at the location between the last observed PCP-mCherry spot and the next observed PCP-mCherry spot.



### Supplemental Fig. 2 | Gene expression analysis of E100/SNAP-38xGA cells.

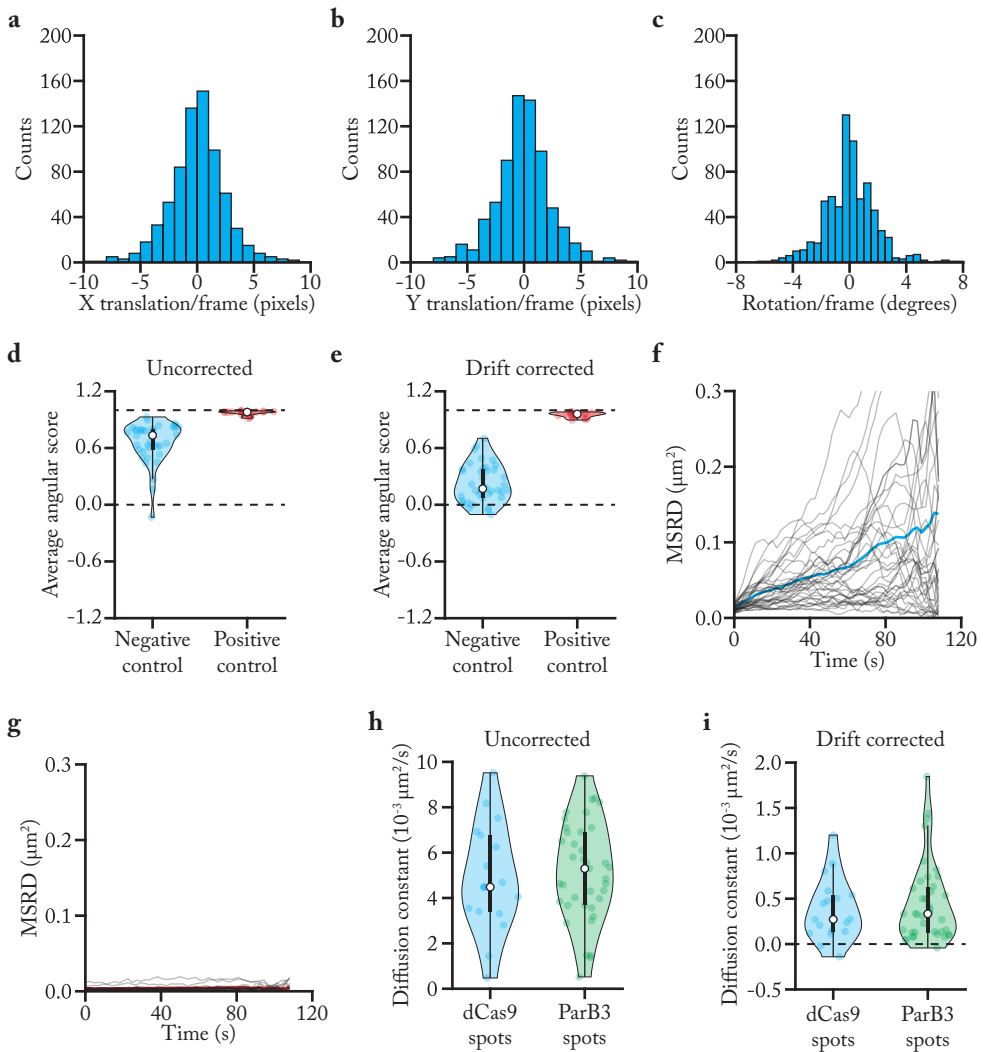
**a**, Cells were labeled with SNAP-JF-646 and the fluorescence levels were measured by FACS. SNAP fluorescence levels of E100/SNAP-38xGA cells were either measured directly after thawing a vial (low passage #) or after keeping the cells in culture for ~2 months (high passage #). Dot and thick error bar indicate median and interquartile range. **b-d**, E100/SNAP-38xGA cells were FACS sorted on having SNAP levels higher than wild-type cells (data not shown). Four days later SNAP fluorescence levels were measured on FACS (**b**) and mRNA levels were measured with FISH (**c-d**). **b**, Cells were labeled with SNAP-JF-646 and the fluorescence levels were measured by FACS. Dot and thick error bar indicate median and interquartile range. **c-d**, E100/SNAP-38xGA cells expressing ParB3-GFP were fixed and incubated with smFISH probes targeting the SNAP reporter to visualize reporter mRNAs. **c**, Number of mRNAs in E100/SNAP-38xGA cells ( $n=66$ ) based on smFISH probes targeting the SNAP reporter. Negative control was determined by incubating E100/eGFP cells ( $n=23$ ) with probes targeting the SNAP reporter (data is replotted from Supplementary figure 1d). Dot and thick error bar indicate median and interquartile range. **d**, Representative image of cells incubated with smFISH probes targeting the SNAP reporter. Arrow indicates co-localization of a ParB3-GFP spot and a smFISH spot, suggesting that this cell is transcriptionally active. Scale bar, 5  $\mu\text{m}$ .





### Supplemental Fig. 3 | Promoter-enhancer distance analysis.

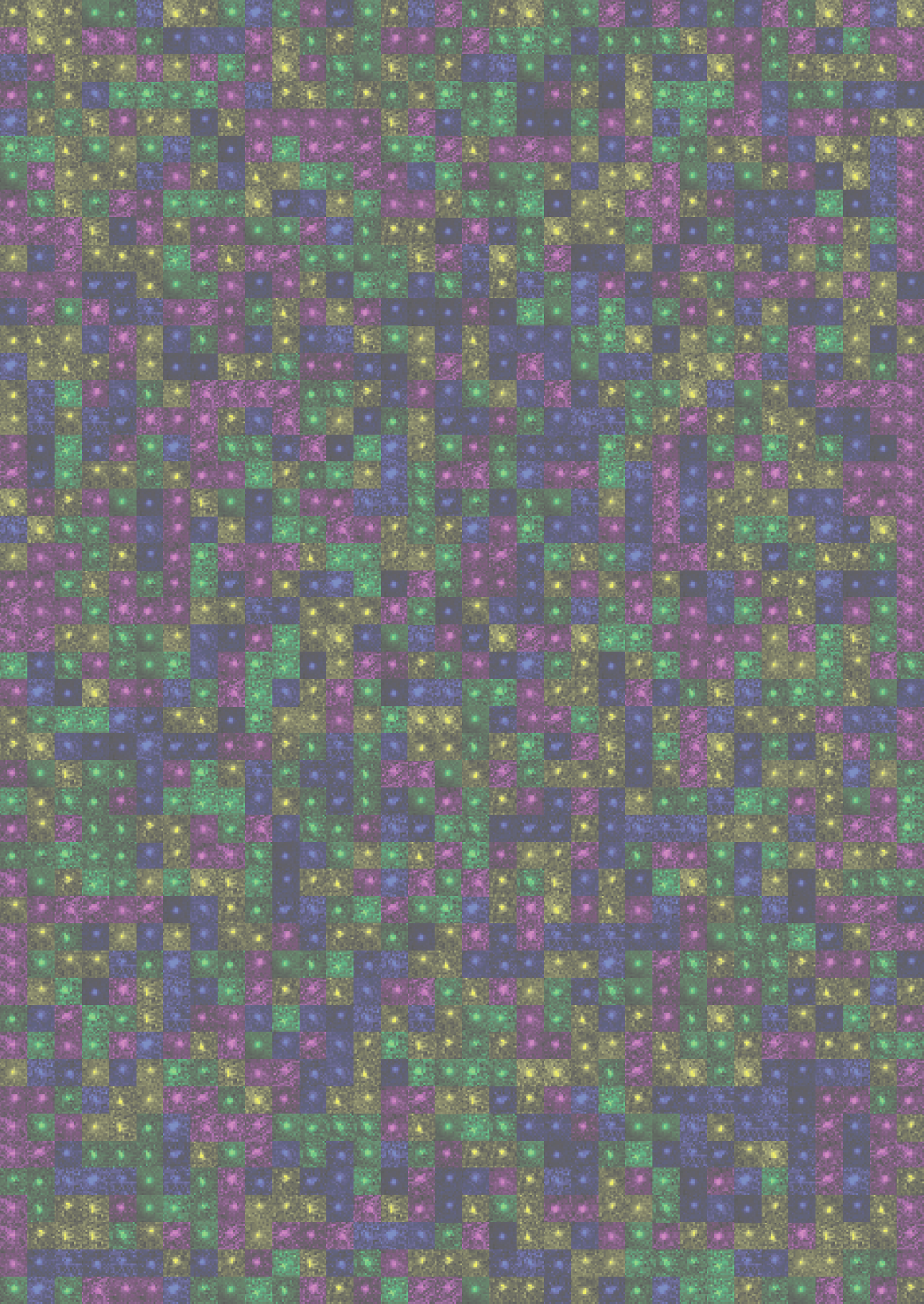
**a-b**, E100/SNAP-38xGA cells were labeled with Halo-JFX-650 and PCP-Halo intensity was measured in a 20x20 pixel ROI centered around the ParB3-GFP spot. Peak intensity over background intensity of PCP-Halo was computed in the ROI. Dashed line at 2 indicates threshold used to determine if a cell has a PCP-Halo spot or not (i.e. if it was transcriptionally active or not). Lines on top of the graph show those cells that were manually annotated to have a PCP-Halo spot. Green lines indicate that the cell had a peak over background ratio higher than the threshold, red lines indicate that the cell had a peak-over-background ratio lower than the threshold. **c-f**, The distance between the ParB3-GFP and dCas9-mCherry spot (see Methods). Dot and thick error bar indicate median and interquartile range. **g**, Cells were labeled with SNAP-JF-646 and the fluorescence levels were measured by FACS. Dot and thick error bar indicate median and interquartile range. Dashed line indicates the median SNAP levels of wild-type K562 cells. The wild-type data is replotted from Supplementary Fig. 2a.



### Supplemental Fig. 4 | Drift correction and co-diffusion analysis.

**a-c**, Cells were drift corrected by using the ParB3 spots in the cell as fiducial markers (see Methods). The drift correction algorithm computed the optimal translation (a-b) and rotation (c) values between each frame of each cell. **d-e**, The average angular score (see Methods) before (d) and after (e) drift correction. The dashed lines indicate score associated with no co-diffusion (at 0) and perfect co-diffusion (at 1). Dot and thick error bar indicate median and interquartile range. **f-g**, The time-averaged mean square relative distance (MSRD) for negative control spots (spots at different loci but in the same cell) (f) and positive control spots (spots in the same locus and in the same cell) (g). Gray thin lines are the MSRD of single spot pairs. Thick blue and red line are the average MSRD for the negative control spots and positive control spots, respectively, and are replotted from Fig. 4f. **h-i**, Diffusion constant as determined from time-averaged MSD measurements before (h) and after (i) drift correction. Dot and thick error bar indicate median and interquartile range. Dashed line in (i) indicates diffusion constant of 0.





# 6

## General discussion

Stijn Sonneveld<sup>1</sup>

<sup>1</sup> Oncode Institute, Hubrecht Institute–KNAW and University Medical Center Utrecht, Utrecht, The Netherlands

## 6.1 A short summary of this thesis

In this work, we have used single cell and single-molecule methods to study various processes that play a role in gene expression regulation. In **chapter 3**, we investigated how Argonaute2 silences its mRNA targets inside a living cell. Next, in **chapter 4** we found a role for mRNA degradation during the cell cycle. Finally, in **chapter 5** we have developed new techniques to study promoter-enhancer interactions in live cells.

In **chapter 1**, we posed several questions; here, we will reflect on those questions to find out how far we got in answering them and what new questions have been invoked.

### 6.1.1 In vivo Argonaute-target interactions

In **chapter 3**, we have measured the cleavage kinetics of single mRNA molecules, revealing that some mRNAs are cleaved within a minute while other mRNAs are still not cleaved after 20 minutes. Interestingly, we found that mRNAs, in which the AGO2 target site was located in the coding sequence, were generally cleaved much faster than mRNAs, in which the AGO2 target site was present in the 3' UTR. We showed that AGO2 target sites are generally masked by RNA structure, but that in the coding sequence translating ribosomes unwind RNA structure, thereby increasing target site accessibility and cleavage efficiency. Finally, we modelled mRNA structural dynamics and found that slow transitions (>10 min) between conformations in which the AGO2 target site becomes accessible, best matched our experimental results.

While we have revealed one important factor (i.e. mRNA structure) that affects AGO2 functioning in live cells, there could still be other factors (such as mRNA modifications) that play an important role. Furthermore, mRNA structure itself is also regulated by different factors; for example, ribosomes can unfold mRNA structure and as a result affect AGO2 functioning. It will be interesting to investigate if there are factors that, similar to ribosomes, can unfold mRNA structure and thereby increase AGO2 target site accessibility. For instance, it could be that other molecules such as RNA helicases assist AGO2 target binding [1]. This could be especially important for AGO2 target sites that reside in the 3' UTR, as ribosomes are not present here. Such a mechanism might not only be important for AGO2 binding, but may also be important for other RNA binding proteins (RBPs) as many RBPs target the 3' UTR of mRNAs [2]. Our results also raise new questions. It still remains elusive how mRNA structure obstructs AGO2 binding exactly. We find it most likely that this is not caused by a single, strong interaction between the AGO2 target site and another sequence in the mRNA. Instead our data suggests that many weak interactions together result in target site masking. Further investigation is

necessary to investigate if it is direct base-pairing of the AGO2 target site or the 3D position of the AGO2 target site within the mRNA molecule (i.e. is the AGO2 target site directly exposed to the outside or is the target site deeply buried inside the RNA structure) that is driving target site masking.

### 6.1.2 mRNA degradation during the cell cycle

In **chapter 4**, we have found 220 genes that are downregulated during the M-G1 phase transition. The downregulation of these genes seemed to occur predominantly at two moments during the M-G1 phase transition, namely during mitosis and after ~80 minutes into G1 phase. Downregulation of mRNA levels can be mediated through a decrease of the synthesis rate or an increase in the degradation rate. We showed that for many mRNAs the half-life during the M-G1 phase transition is significantly shorter than the half-life in unsynchronized cells, and found that the downregulation of mRNA levels depends on CNOT1, a central factor in deadenylation-dependent decay [3]. Together, these results suggest that mRNA degradation plays an important role in the downregulation of mRNA levels during the M-G1 phase transition.

While we have detected genes that are subject to mRNA degradation during the M-G1 phase transition there are still many questions that remain unanswered. Firstly, the exact mechanism underlying the mRNA degradation during the M-G1 phase transition remains elusive. We identified CNOT1 as an important factor suggesting that mRNA decay is mediated through the deadenylation-dependent decay pathway. However, as CNOT1 and deadenylation-dependent decay governs mRNA degradation of many genes, it remains unclear how specific genes are targeted for degradation during the M-G1 phase transition. The two waves of mRNA decay suggest that (at least) two different mechanisms exist to target these two groups of mRNAs. It could be that specific RBPs target the mRNAs and recruit the deadenylation machinery (i.e. CNOT1). As RBPs are able to distinguish between RNA molecules containing very similar sequences [4], such a mechanism could provide the specificity to target a subset of genes for mRNA degradation. To test this hypothesis, a CRISPR screening approach [5,6] could be employed in order to find either the RBP or the sequence motif within the mRNA that is responsible for the degradation during the M-G1 phase transition.

Since many genes are subject to mRNA degradation during the M-G1 phase transition, the question arises what function mRNA degradation plays in the cell cycle. There are already many processes in place that regulate the expression of cell cycle genes, among which transcriptional activation and scheduled protein degradation are most extensively studied [7–10]. If transcriptional activation and protein

degradation are able to control gene expression, what additional role could mRNA degradation play? One possibility could be that increased mRNA degradation results in a rapid clearance of transcripts, while transcriptional shutdown alone may result in prolonged gene expression [11–13]. To test the function of mRNA degradation during the cell cycle, it would be interesting to inhibit mRNA decay of the genes that are downregulated during the M-G1 phase transition, and to examine the effects on cell cycle progression. However, to inhibit mRNA degradation during the M-G1 phase transition, we will first need to find the mechanisms that are governing this mRNA degradation.

### 6.1.3 Towards visualizing promoter-enhancer interactions

In **chapter 5**, we have developed a live-cell imaging method to visualize promoter and enhancer location, and nascent transcription simultaneously. Initial results show that promoter and enhancer are often in close proximity (i.e. within 150 nm), suggesting that promoter-enhancer proximity is required for transcriptional activation. However, we also often observed promoter-enhancer proximity without transcriptional activity, raising the question why promoter-enhancer proximity not always results in transcriptional activation. One possibility could be that promoter-enhancer topology only undergoes minor conformational changes to activate transcription, and that these conformational changes are accompanied with small changes in promoter-enhancer distance. Small changes in promoter-enhancer distance may remain undetected, especially since measurements of promoter-enhancer distance are affected by imaging and localization errors [14].

In addition, it is unclear if a promoter and enhancer form a stable interaction or only bind transiently. In case interactions only occur transiently, this may further complicate the detection of a correlation between promoter-enhancer proximity and transcriptional activation. Recently, it was also proposed that promoter-enhancer contacts are remembered for a long period of time [15,16], which may result in a delayed correlation between promoter-enhancer proximity and transcriptional activation. Together, these reasons may explain why we and others [17,18] have not observed a correlation between promoter-enhancer proximity and transcriptional activation. To overcome these challenges, new and more sensitive measurements of promoter-enhancer topology are required. In **chapter 5**, we have developed a new assay that uses co-diffusion as a measure for promoter-enhancer interaction. Further experiments are required to test if this method can be a better parameter to measure promoter-enhancer proximity. Then, by carefully quantifying transcriptional dynamics (i.e. what are the durations of transcriptionally active and inactive periods) and promoter-enhancer topology, we aim to be able to distinguish between the many



models of promoter-enhancer communication [19].

## 6.2 Heterogeneity in gene expression

Using single cell and single-molecule techniques, we have studied the mechanisms behind 3 different processes that play an important role in gene expression. In all three studies we have found extensive heterogeneity between single cells or single molecules. Does this observed heterogeneity fulfill a function? In some cases it would seem that heterogeneity is indeed functional and the origin of heterogeneity is controlled by the cell. For instance, in **chapter 4** we found that many mRNAs are subject to mRNA decay during a specific moment in the cell cycle (i.e. during the M-G1 phase transition). As a result, in an asynchronous cell population (i.e. a cell population that consists of cells that are in different phases of the cell cycle) it is expected that for these genes there will be a large cell-to-cell heterogeneity of the mRNA degradation rate. This cell-to-cell heterogeneity reflects the different cell cycle stages these cells are in, and it would seem that the cell thus controls the heterogeneity in this case.

On the other hand, in **chapter 3** we found that for one reporter gene, some mRNA molecules are cleaved rapidly (within minutes) by AGO2 while other mRNA molecules remain intact for long times (>20 minutes). We showed that the heterogeneity in cleavage times originates predominantly from differences in target site accessibility. If the AGO2 target site is not masked by mRNA structure, AGO2 can quickly bind and cleave the mRNA molecule. In contrast, in case mRNA structure masks the AGO2 target site, AGO2 binding and mRNA cleavage is prevented. Thus, the structural conformation of an mRNA molecule determines if an mRNA molecule is cleaved rapidly or not. While a functional role of this heterogeneity cannot be excluded (i.e. maybe only a fraction of mRNA molecules needs to be cleaved), it seems more likely that mRNA structure is not functional and only impairs AGO2 functioning. In this case, the origin of the observed heterogeneity is a consequence of the nature of RNA molecules to form intra and intermolecular base-pair interactions and not because of regulatory mechanisms that are controlled by the cell.

To better understand heterogeneity in gene expression at the single cell or single-molecule level, it is essential to investigate the origins of heterogeneity. In general, there are two sources that contribute to the heterogeneity or variability in gene expression, commonly referred to as intrinsic and extrinsic variability [20,21]. Intrinsic variability originates from the inherent stochastic nature of the biochemical processes that underlie gene expression such as transcription or translation, while extrinsic variability originates from actual differences between cells or single

molecules. Interestingly, intrinsic and extrinsic variability is also often referred to as ‘noise’ as if it only results in unwanted fluctuations of gene expression. However, both intrinsic and extrinsic sources of heterogeneity may fulfill a function in the cell. For instance, in **chapter 2** we discussed that intrinsic variability during translation initiation start site selection may have an important function in the cell as it can result in functional proteome diversification [22]. On the other hand, heterogeneity in start site selection may also be ‘unwanted’ as it can result in the synthesis of aberrant or toxic proteins. By carefully studying the origins and consequences of heterogeneity, we can obtain a better understanding of the processes underlying gene expression.

### 6.3 Towards a better understanding of gene expression heterogeneity

To study the origins and consequences of gene expression heterogeneity, methods are required that allow us to assess gene expression at the single cell and single-molecule level. While in recent years technological advancements in single cell sequencing and single-molecule microscopy techniques have rapidly increased our ability to investigate chromatin architecture, transcription, and translation [23–29], there are still limitations to these methods. For instance, it remains challenging to visualize the interaction of a single RBP with a single mRNA molecule of interest. While studies have performed single-particle tracking experiments of several RBPs [30], this is only possible under conditions in which a subset of the molecules is labeled (which is obtained by partial labeling or photo-bleaching). Such an approach, however, complicates detection of a binding event to a specific mRNA molecule as only a small fraction of the molecules is labeled. In **chapter 3**, this limitation precluded visualization of the binding of a single AGO2 molecule to its target molecule. Direct visualization of AGO2 target binding would be very valuable as it could present direct evidence for our model that AGO2 binding is limiting and not the AGO2 slicing or release step. Thus, development of new methods that operate at the single cell and single-molecule level are key to investigate gene expression heterogeneity.

Not only developing new methods is necessary to obtain a better understanding of gene expression heterogeneity, but also combining existing methods may prove to be very valuable. Since many factors collectively control the processes underlying gene expression, multiple parameters together may explain heterogeneity in gene expression. As such, if all parameters are measured simultaneously, a more complete understanding of the origins and consequences of gene expression heterogeneity can be obtained. For instance, in **chapter 2** we looked at translation heterogeneity and how several factors may explain intragenic heterogeneity (i.e. different mRNAs derived from one gene in a single cell show translational heterogeneity). Two mRNA molecules may contain different mRNA modifications, adapt different

RNA structures, and bind different RBPs. Together, these differences may explain translational heterogeneity between two mRNA molecules. Thus, to understand the origins of translational heterogeneity, it is important that these parameters are measured simultaneously on the same mRNA molecule.

Measuring multiple parameters simultaneously is also important to better understand promoter-enhancer communication. In **chapter 5** we have developed a method that allows us to simultaneously assess promoter and enhancer location as well as transcriptional activity. By integrating all the information together we hope to obtain a better understanding of promoter-enhancer communication. Promoter activity, however, is not only affected by the conformation of the promoter and enhancer. Instead, it is likely that binding of different transcription factors and DNA and histone modifications in the region surrounding the promoter and/or enhancer also play a key role. Thus, for a better understanding of promoter-enhancer communication, it could be informative to measure (some of) these factors as well. In single-molecule imaging methods, measurement of multiple parameters is not trivial though, as the number of imaging channels (which sets the number of parameters that can be measured simultaneously) is limited. There are different solutions to circumvent this problem. For instance, in case of fixed cells, different parameters can be measured in the same cell or on the same molecule by employing multiple rounds of imaging during which the parameters are visualized one-by-one [31]. Such an approach is not possible for live cells and for such cases other solutions exist. For example, it is possible to increase the number of available imaging channels by using spectral decomposition methods, which allow the usage of different colors that partially overlap in their emission spectrum [32].

Looking back, it has only been little over 20 years ago when the first measurements of single RNA molecules in a single cell were obtained [33]. Since then, new methods have been developed at a staggering pace resulting in an exponential increase of our toolbox to study gene expression at the single cell and single-molecule level. As more and more methods are being developed and existing methods are being combined, we can better interrogate the causes and consequences of gene expression heterogeneity and as such we are expected to obtain a better understanding of the fundamental processes underlying gene expression.

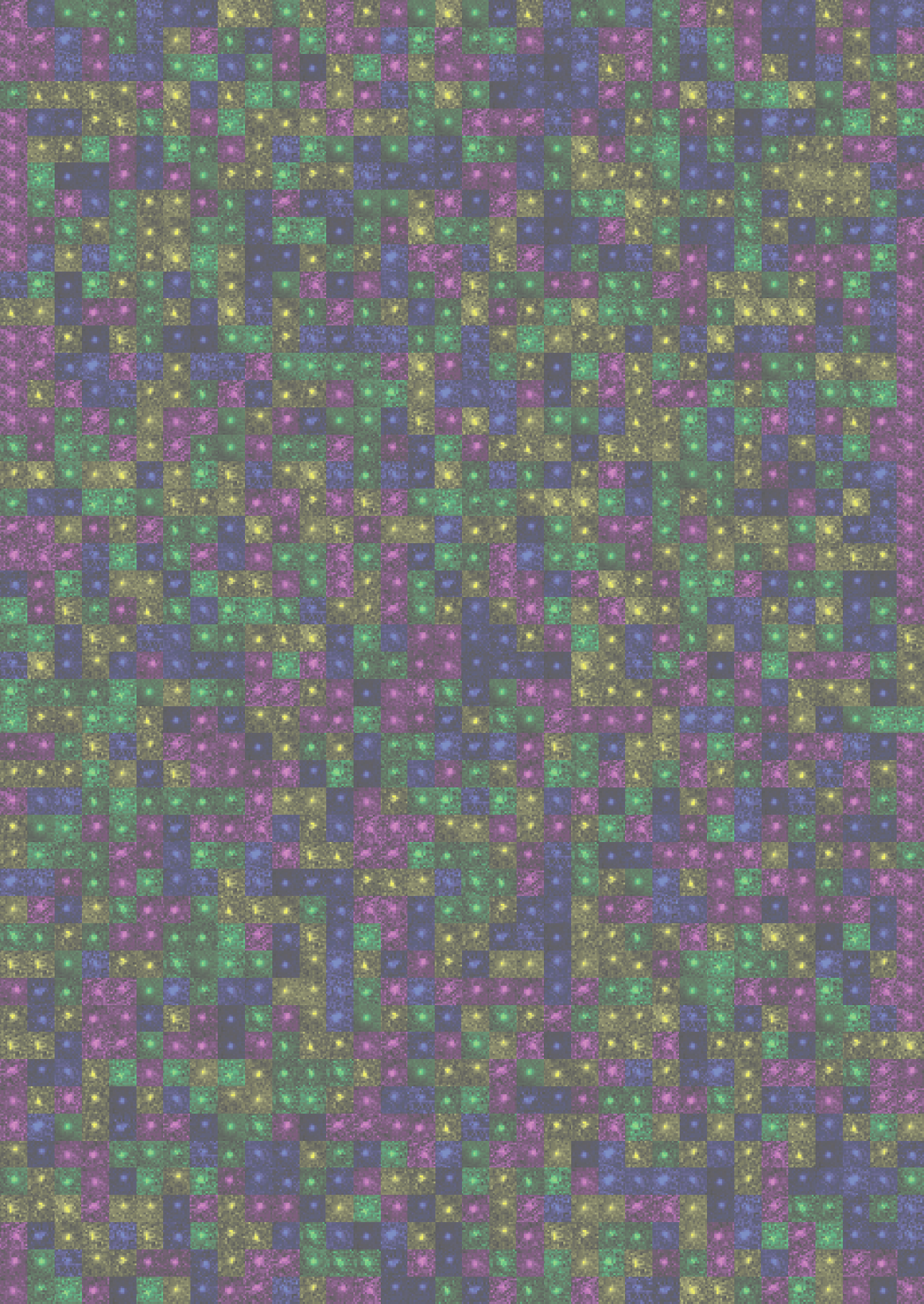
## 6.4 References

- 1 Tauber, D. et al. (2020) Modulation of RNA condensation by the DEAD-box protein eIF4A. *Cell* 180, 411–426
- 2 Hentze, M.W. et al. (2018) A brave new world of RNA-binding proteins. *Nat. Rev. Mol. Cell Biol.* 19, 327–341

- 3 Tucker, M. et al. (2001) The transcription factor associated Ccr4 and Caf1 proteins are components of the major cytoplasmic mRNA deadenylase in *Saccharomyces cerevisiae*. *Cell* 104, 377–386
- 4 Webster, M.W. et al. (2019) RNA-binding proteins distinguish between similar sequence motifs to promote targeted deadenylation by Ccr4-Not. *Elife* 8, e40670
- 5 Kampmann, M. (2018) CRISPRi and CRISPRa screens in mammalian cells for precision biology and medicine. *ACS Chem. Biol.* 13, 406–416
- 6 Adli, M. (2018) The CRISPR tool kit for genome editing and beyond. *Nat. Commun.* 9, 1911
- 7 Nakayama, K.I. and Nakayama, K. (2006) Ubiquitin ligases: cell-cycle control and cancer. *Nat. Rev. Cancer* 6, 369–381
- 8 Vodermaier, H.C. (2004) APC/C and SCF: controlling each other and the cell cycle. *Curr. Biol.* 14, 787–796
- 9 Bertoli, C. et al. (2013) Control of cell cycle transcription during G1 and S phases. *Nat. Rev. Mol. Cell Biol.* 14, 518–528
- 10 Sadasivam, S. and DeCaprio, J.A. (2013) The DREAM complex: master coordinator of cell cycle-dependent gene expression. *Nat. Rev. Cancer* 13, 585–595
- 11 Rabani, M. et al. (2014) High-resolution sequencing and modeling identifies distinct dynamic RNA regulatory strategies. *Cell* 159, 1698–1710
- 12 Elkon, R. et al. (2010) Major role for mRNA stability in shaping the kinetics of gene induction. *BMC Genomics* 11, 259
- 13 Shalem, O. et al. (2008) Transient transcriptional responses to stress are generated by opposing effects of mRNA production and degradation. *Mol. Syst. Biol.* 4, 223
- 14 Niekamp, S. et al. (2019) Nanometer-accuracy distance measurements between fluorophores at the single-molecule level. *Proc. Natl. Acad. Sci. U. S. A.* 116, 4275–4284
- 15 Xiao, J. et al. (2020) How subtle changes in 3D structure can create large changes in transcription. *bioRxiv*
- 16 Zuin, J. et al. (2021) Nonlinear control of transcription through enhancer-promoter interactions. *bioRxiv*
- 17 Alexander, J.M. et al. (2019) Live-cell imaging reveals enhancer-dependent Sox2 transcription in the absence of enhancer proximity. *Elife* 8, e41769
- 18 Mateo, L.J. et al. (2019) Visualizing DNA folding and RNA in embryos at single-cell resolution. *Nature* 568, 49–54
- 19 Furlong, E.E.M. and Levine, M. (2018) Developmental enhancers and chromosome topology. *Science* 361, 1341–1345
- 20 Swain, P.S. et al. (2002) Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proc. Natl. Acad. Sci. U. S. A.* 99, 12795–12800
- 21 Singh, A. and Soltani, M. (2013) Quantifying intrinsic and extrinsic variability in stochastic gene expression models. *PLoS One* 8, e84301
- 22 Chen, J. et al. (2020) Pervasive functional translation of noncanonical human open reading frames. *Science* 367, 1140–1146
- 23 Shapiro, E. et al. (2013) Single-cell sequencing-based technologies will revolutionize whole-organism science. *Nat. Rev. Genet.* 14, 618–630
- 24 Brandão, H.B. et al. (2021) Tracking and interpreting long-range chromatin interactions with super-resolution live-cell imaging. *Curr. Opin. Cell Biol.* 70, 18–26
- 25 Darzacq, X. et al. (2009) Imaging transcription in living cells. *Annu. Rev. Biophys.* 38, 173–196
- 26 Vera, M. et al. (2016) Single-cell and single-molecule analysis of gene expression regulation. *Annu. Rev. Genet.* 50, 267–291
- 27 Biswas, J. et al. (2019) Fluorescence imaging methods to investigate translation in single cells. *Cold Spring Harb. Perspect. Biol.* 11,

a032722

- 28 Cialek, C.A. et al. (2020) Lighting up single-mRNA translation dynamics in living cells. *Curr. Opin. Genet. Dev.* 61, 75–82
- 29 Pichon, X. et al. (2018) A growing toolbox to image gene expression in single cells: sensitive approaches for demanding challenges. *Mol. Cell* 71, 468–480
- 30 Katz, Z.B. et al. (2016) Mapping translation “hot-spots” in live cells by tracking single molecules of mRNA and ribosomes. *Elife* 5, e10415
- 31 Moffitt, J.R. et al. (2016) High-throughput single-cell gene-expression profiling with multiplexed error-robust fluorescence in situ hybridization. *Proc. Natl. Acad. Sci. U. S. A.* 113, 11046–11051
- 32 Valm, A.M. et al. (2016) Multiplexed spectral imaging of 120 different fluorescent labels. *PLoS One* 11, e0158495
- 33 Femino, A.M. et al. (1998) Visualization of single RNA transcripts in situ. *Science* 280, 585–590



# Appendix

## Tracking translation with TransTrack

Stijn Sonneveld<sup>1</sup> and Marvin E. Tanenbaum<sup>1</sup>

<sup>1</sup> Oncode Institute, Hubrecht Institute–KNAW and University Medical Center  
Utrecht, Utrecht, The Netherlands

TransTrack has been used in the following publications (in order of appearance):

Boersma, S., Khuperkar, D., Verhagen, B.M.P. et al. Multi-color single-molecule imaging uncovers extensive heterogeneity in mRNA decoding. *Cell* 178, 458–472 (2019).  
<https://doi.org/10.1016/j.cell.2019.05.001>

Hoek, T.A., Khuperkar, D., Lindeboom, R.G.H. et al. Single-molecule imaging uncovers rules governing nonsense-mediated mRNA decay. *Mol Cell* 75, 324–339 (2019).  
<https://doi.org/10.1016/j.molcel.2019.05.008>

Ruijtenberg, S., Sonneveld, S., Cui, T.J. et al. mRNA structural dynamics shape Argonaute-target interactions. *Nat Struct Mol Biol* 27, 790–801 (2020).  
<https://doi.org/10.1038/s41594-020-0461-1>

## A.1 Abstract

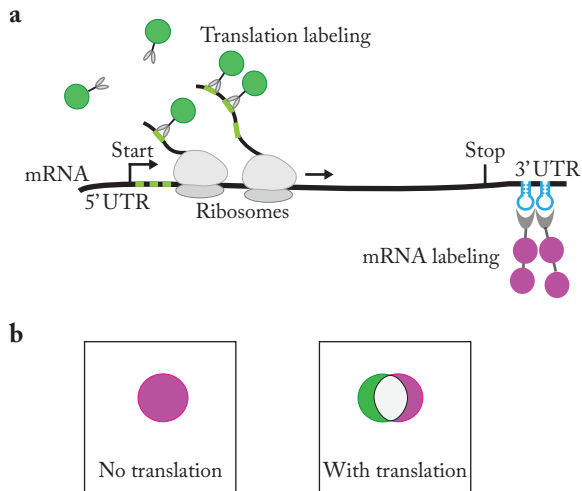
Recently, a new method was developed that allows live-cell imaging of translation of individual mRNA molecules. One strength of this method is that it enables quantification of translational activity on single mRNA molecules. However, manual tracking and quantification of translational activity on single mRNA molecules is a labor intensive process and can also introduce potential selection bias. Here, we describe a new software package, named ‘TransTrack’, that facilitates quantification of translational activity on single mRNA molecules over time. TransTrack combines automated single-particle tracking with a suite of manual actions. These actions allow the user to manually select mRNA molecules for quantification and to correct the mRNA tracks that contain errors. Together, this semi-automated method ensures the generation of high-quality tracks while minimizing manual labor.

## A.2 Introduction

Over the last two decades development of new imaging techniques and genetically encoded tags have strongly increased the use of microscopy-based methods in gene expression analysis [1]. Recently, technical advancements have resulted in a new microscopy-based live-cell imaging method that visualizes the process of translation of individual mRNA molecules (Fig. 1a; referred to as single-molecule translation

**Fig. 1 | A method for single-molecule translation imaging.**

**a**, Schematic of the single-molecule translation imaging method. Translation is detected through the introduction of multiple copies of a short peptide epitope (e.g. SunTag or HA frankenbody) at the start of the coding sequence. Upon translation, these peptide epitopes are bound by fluorescently-labeled antibodies resulting in a bright fluorescent spot. The mRNA molecule is labeled independently by the introduction of a repeated stem-loop structure (e.g. PP7 or MS2 hairpin structures) in the 3' UTR. The stem-loop structure is directly bound by a fluorescently-labeled protein. **b**, Schematic representation of the signal as observed with a (confocal) microscope. In the absence of translation (left) only an mRNA signal will be present. In the presence of translation (right), there will be co-localization of the mRNA signal with a translation signal.

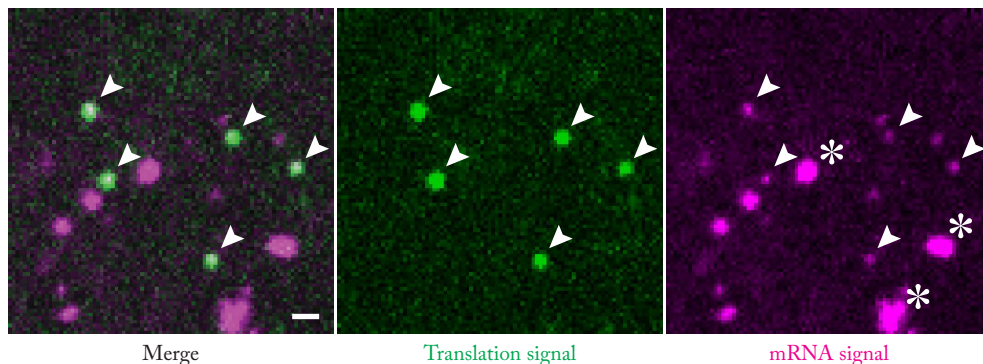




imaging) [2–7]. This method uses multiple copies of a short peptide sequence (e.g. the SunTag [8] or HA frankenbody [9]) that are inserted in the coding sequence of a gene of interest. Upon translation of the gene of interest, the peptide epitopes are synthesized and recognized by a fluorescently-labeled antibody, resulting in the accumulation of a bright fluorescent signal (referred to as the translation signal or channel). To label mRNA molecules independently of translation as well, a repeated stem-loop structure (e.g. PP7 [10] or MS2 [11] binding sites) is introduced in the 3' UTR of the gene of interest. This stem-loop structure is directly bound by a fluorescently-labeled protein (in a different color), resulting in the accumulation of a second fluorescent signal (referred to as the mRNA signal or channel). Together, this method results in an mRNA molecule that is labeled in one color in the absence of translation and in two colors in the presence of translation (Fig. 1b).

Single-molecule translation imaging presents the opportunity to study translation from new angles. Firstly, single-molecule imaging offers high spatial resolution (using confocal microscopy lateral resolutions of ~200 nm can be obtained) and temporal resolution (images can be taken in milliseconds). Secondly, instead of taking a snapshot, translation can be followed in real time, which is a powerful method when studying a sequence of events (e.g. translation initiation, elongation, and termination). Thirdly, by studying mRNA molecules one-by-one, heterogeneity between mRNA molecules can be uncovered that would remain elusive in ensemble approaches. Finally, single-molecule translation imaging allows a quantitative assessment of translation as the brightness of the translation signal directly reports on translational activity. Moreover, with careful calibration the intensity of the translation signal can be converted into the number of ribosomes that translate a single mRNA molecule [3]. Such a quantitative view complements the other unique features of single-molecule imaging and can be used to answer different types of questions. For instance, it allows assessing the frequency a ribosome engages in a form of alternative translation [12] or it enables measuring the effects of an siRNA on translation initiation and elongation [13]. These examples highlight that a quantitative assessment of translation by single-molecule translation imaging is instrumental towards a more complete understanding of the mechanisms of translation.

To quantify translation data obtained with single-molecule translation imaging, it is necessary to follow a large number of mRNA molecules over time. Manual tracking of mRNA molecules and quantification of their translational activity is, however, a laborious and time-consuming undertaking and can result in the introduction of human selection bias. To overcome this problem, many (semi-)automated single-particle tracking algorithms have been developed that minimize manual labor [14,15]. While automated single-particle tracking will greatly reduce the time to



**Fig. 2 | Representative image obtained with single-molecule translation imaging.**

A representative image of a U2OS cell expressing the SunTag antibody (green) and the PP7 coat protein (magenta). The arrowheads indicate translation sites that have a translation and mRNA signal. Asterisks indicate large fluorescent foci in the mRNA channel which are fluorescent lysosomes. Scale bar is 1  $\mu\text{m}$ .

analyze translation of individual mRNA molecules, such an approach can be less reliable than a manual analysis. In the case of single-molecule translation imaging, automated mRNA tracking may be unreliable, since a ‘standard’ image (of the mRNA signal) contains the presence of a high number of background spots (Fig. 2). In addition, the mRNA signal from bona fide mRNA molecules (which can be identified by the co-localization of a translation signal with an mRNA signal) is relatively weak compared to some of the background spots (Fig. 2). This poses two serious challenges for automated single-particle tracking. Firstly, automatic selection of mRNA molecules may result in either high false-positive or false-negative selection rates as mRNA molecules are difficult to distinguish from background spots. Secondly, automated tracking of mRNA molecules is also subject to mistakes as the weak mRNA signal can be lost or mistaken for background signal. Taken together, while automated single-particle tracking diminishes manual labor during analysis, it comes at the cost of a reduced reliability of the analysis.

To overcome the disadvantages of automated single-particle tracking algorithms, we decided to design a new analysis software package that we named TransTrack. In TransTrack, we aimed to combine the speed of automated single-particle tracking algorithms with the preciseness of manual tracking. Above all, the main goal of TransTrack was to develop a user-friendly software that is tailored to the needs of the typical researcher analyzing single-molecule translation images. To this end, we implemented several manual steps in the analysis pipeline of TransTrack. Firstly, while automatic track selection is possible, the user of TransTrack can also manually add or delete mRNA molecules to the track list. Secondly, after track selection the user has several options to correct the automatically-generated tracks that don’t follow an

mRNA molecule correctly. Together, we believe that TransTrack offers an analysis software that allows researchers to reduce their analysis time while maintaining the high quality standards of manual analysis. In the next section we will provide a step-by-step manual that describes how to use TransTrack for the analysis of translation on individual mRNA molecules over time.

### A.3 A step-by-step manual

TransTrack is a software written in MATLAB with the primary goal to track translation sites visualized with live cell microscopy. Translation sites can be tracked over time and intensities in different channels can be measured at the same time. Furthermore, TransTrack aims to combine automatic and manual functions that help the user to speed up the analysis while creating a high quality data-set. Finally, although TransTrack is designed for tracking translation sites, it can also be used as a more general tool where spot tracking and intensity measurements are required.

#### A.3.1 Installing TransTrack

TransTrack 1.0 was built in the MATLAB environment and requires MATLAB version R2018b or later (TransTrack was also tested in MATLAB version R2012b, where it is also functional except for the measure function, section A.3.5).

Take the following two steps (a-b) to install TransTrack.

- a. Download TransTrack from Github (<https://github.com/TanenbaumLab>) and save on a computer (see Supplemental information for a list of all the functions that should be in the TransTrack folder).
- b. Add the saved folder to MATLAB's search paths. This can be done by clicking on **Set Path** (in MATLAB's main window) and selecting the correct folder. Always use the option **Add with Subfolders**, as this will also add the folders that are in the TransTrack folder. Important is that all folders that contain functions that are used by TransTrack are in MATLAB's search paths. Otherwise MATLAB will not be able to locate the specified functions.

#### A.3.2 Starting TransTrack and an initial overview

To start TransTrack, go to the folder where TransTrack is saved and open the **TransTrack.m** function in the MATLAB editor. Next, TransTrack is started by clicking on the **Run** button in the MATLAB editor (when the **TransTrack.m** file is selected).

Menu	Option	Short key	Short description	Section
File	Open	Ctrl + O	This opens a new movie in TransTrack.	A.3.3
	Load – Track Data	Ctrl + L	Load	A.3.10
	Load – ROIs		Load ROIs.	A.3.10
	Save – Track Data	Ctrl + S	Save results from the current analysis.	A.3.9
	Save – ROIs		Save the ROIs that are currently selected.	A.3.9
Image	Display Settings	Ctrl + C	Select brightness and contrast of channels.	A.3.4
	Channels	Ctrl + Z	Select color and visibility of channels.	A.3.4
Analysis	Measure	Ctrl + M	Measure intensity in different channels at a spot selected by the user.	A.3.5
	Photobleach correction – Status		Check which channels have been corrected for photobleaching.	A.3.6
	Photobleach correction – Correction		Activate or inactivate photobleach correction for each channel.	A.3.6
	Single channel tracking	Ctrl + A	Start analysis where tracks can be made in a single user selected channel, and tracks are analyzed.	A.3.7
	Find track	Ctrl + F	Find a specific track in the analysis.	A.3.8

**Table 1 | Overview of TransTrack menu options.**

Overview of all different TransTrack menu options. Here, a short key to the menu option (if there is one), a short description of each option, and the section in which the option will be explained in full detail is provided.

After starting TransTrack, a new window is opened. This is TransTrack's main window and will display the movie that is loaded in TransTrack. All functions in TransTrack are reached through the menu bar at the top of the main window. In the remainder of the manual, the use of all menu options will be described in full detail. The manual is ordered in such a way that it follows the typical workflow of an analysis. To give an overview of all menu options see Table 1. Here you can find: (1) all menu options, (2) potential short keys to reach these options in TransTrack, (3) a short description of the menu option, and (4) a reference to the section where a more detailed explanation is given.

### A.3.3 How to open and navigate through a movie

TransTrack only recognizes .tif files. Furthermore, TransTrack takes each .tif file it encounters as a separate position. Finally, each .tif file should be organized in the following order: frame – channel. For example, a 2-frame and 2-channel movie should be organized as follows: frame 1 - channel 1, frame 1 - channel 2, frame 2 - channel 1, frame 2 - channel 2.

A new movie is opened in TransTrack by selecting **File/Open** (or with shortcut Ctrl + O). This will prompt a popup-menu where the user is asked to give the following information.

- **Number of channels.** Give the amount of channels in the movie.
- **First frame.** Frame in movie from which the user wants to start (if not given, TransTrack will open the movie from the first frame).
- **Last frame.** Frame in movie until the user wants to open (if not given, TransTrack will open the movie until the last frame).

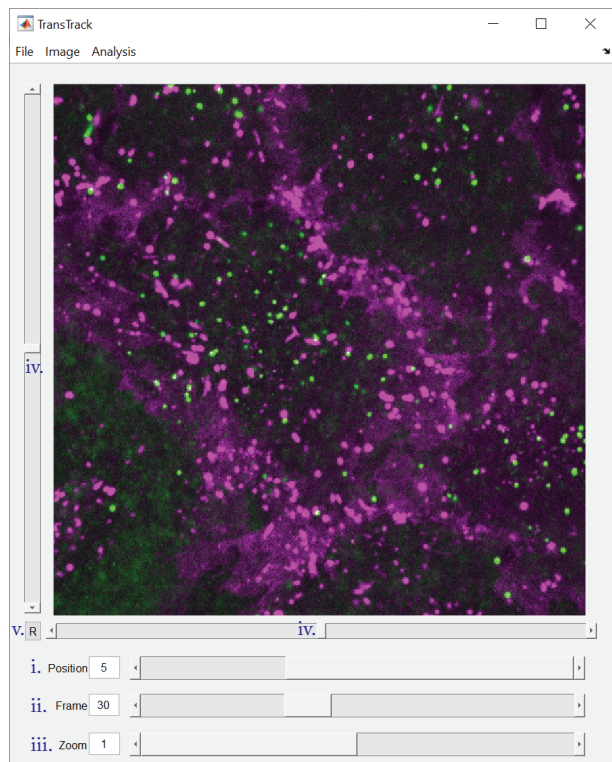
The first input is required to open the movie, the latter two inputs are not required and can be given by the user if only a part of the movie needs to be opened.

Next, a new dialog window is opened where the user can select the folder that contains the actual movie the user wants to open, all .tif files that are in this folder will be opened in TransTrack (note that all .tif files should have similar dimensions). After this, TransTrack will start opening the movie while updating on its progress through a progress bar. Note that all other functions in TransTrack require an image to be opened in TransTrack. Therefore, the other functions will not work as long as there is not an image opened.

Once TransTrack has opened the movie, the first image of the first position will be displayed in TransTrack's main window (Fig. 3). Furthermore, a number of sliding bars appear that allow the user to navigate

**Fig. 3 | Screenshot of TransTrack 'main' window.**

In the main window of TransTrack the user can navigate through the movie by using the position slider (i.), frame slider (ii.), zoom slider (iii.), x and y sliders (iv.) and the reset button (v., which resets the x and y sliders to their original positions).



through the movie. Here, the function of these sliders will shortly be described.

- **Position slider.** With this slider the user can scroll through the different positions. The position can also be controlled by listing a position in the box next to it.
- **Frame slider.** With this slider the user can scroll through the different frames. The frame can also be controlled by listing a frame in the box next to it.
- **Zoom slider.** With this slider the user can zoom into the image. The zoom can also be controlled by listing the amount of zoom in the box next to it. A 'zoom factor' of 1 means that the entire image will be visible. The maximum 'zoom factor' can be 10.
- **x and y sliders.** The x and y sliders are located on the left and lower edge of the image, respectively. With the two sliders the user can change the center of the image left-right (with x slider) and up-down (with y slider). This is useful when zooming in, since the whole image will not be visible anymore once zoomed in.
- **Reset button.** Small button with letter **R**, located in the lower left-right corner of the image. This button automatically centers the image (this re-sets the x and y slider to the middle).

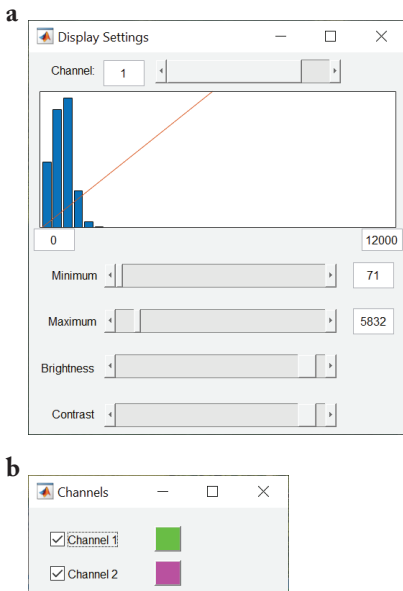
At any point a new movie can be opened in TransTrack by clicking on the **Open** option in the **File** menu. This will replace the previous movie and any analysis done with the previous movie will be deleted.

### A.3.4 How to set image display settings

There are two different options to set the display settings for visualizing the image in TransTrack. First, brightness and contrast can be set by selecting **Image/Display Settings** (or with shortcut Ctrl + C) and the color and visibility of the channels can be set by selecting **Image/Channels** (or with shortcut Ctrl + Z).

After selecting **Image/Display Settings**, a new window will appear (Fig. 4a). Here, the brightness of each channel (select the channel through the sliding bar or the edit box at the top of the window) can be tweaked by changing the minimum, maximum, brightness, and contrast values. Furthermore, a histogram of the intensities in the selected channel is displayed as a visual aid to choose the most optimal settings.

After clicking on **Image/Channels**, a new window will appear (Fig. 4b). Here, a



**Fig. 4 | Screenshot of TransTrack 'image display' windows.**

a, In the display settings window the user can determine the brightness and contrast of each channel. b, In the channel window, the user can select a display color for each channel. Furthermore, the user can toggle on or off the visibility of each channel.

display color can be chosen for each channel by clicking on the colored box next to the channel that the user wants to change. The user can also toggle on or off the visibility of a specific channel by checking or unchecking the checkbox, respectively.

### A.3.5 Measure intensities in movie

The measure function enables the user to quantify the intensities in all channels at a specific user-defined location in the image. The measure function can be started by selecting **Analysis/Measure** (or with shortcut Ctrl + M). Next, the user has to select a point inside the image in TransTrack's main window. TransTrack will determine the following intensities at the location given by the user.

- **Peak intensity.** This is the intensity at the selected location.
- **Mean intensity.** This is the mean intensity at the selected location. This is calculated by taking the mean intensity of a box drawn around the given location (the box size used is the same as in further analyses, see section 5.1).

Note: the mean intensity is NOT background subtracted!

- **Total intensity.** This is the sum intensity of all pixels in the current frame that is displayed.

All three intensities are determined for each channel present in the image. The data will be stored in a table to which any new measurement is added. When the table is closed all measurements are deleted, and a new table (without any previous measurements) will be opened when a new measurement is performed.

### A.3.6 How to perform photobleach correction of movie

TransTrack can perform a photobleach correction in case this is desired by the user. When a photobleach correction is performed, the image displayed in TransTrack's main window will be updated accordingly and any intensity measurements in

downstream analyses will be performed on the photobleach corrected image. The user can always check if a photobleach correction is applied to a channel by clicking on **Analysis/Photobleach correction/Status**. A dialog box will appear, listing for each channel if a photobleach correction has been applied.

To apply or remove the photobleach correction, the user can select **Analysis/Photobleach correction/Correction**. Per channel, the user will be asked to apply a photobleach correction (in case no correction has been applied yet for this channel) or to remove the photobleach correction (in case a correction has already been applied for this channel). It is possible to use a photobleach correction for only a subset of the channels.

Photobleach correction in TransTrack is done as follows: (1) the mean intensity of all pixels for each frame is measured (this is done per position) and (2) a decay curve (Equation A.1) is fitted to each mean intensity time trace.

$$\text{Mean Intensity}(t) = k_1 \cdot e^{-k_2 \cdot t} + k_3 \quad (\text{Equation A.1})$$

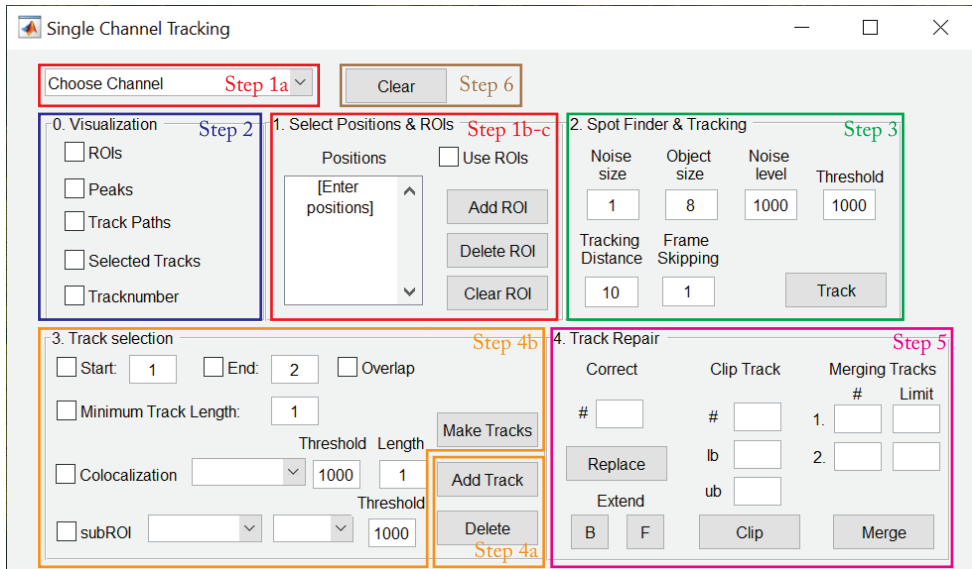
Next, the mean intensity time traces and resulting fits are plotted and shown to the user, who can decide (based on visual inspection) to allow the photobleach correction or to reject it. When the user decides to allow the photobleach correction, all images are corrected per position according to their corresponding fit. Note that each position is photobleach corrected with its own fitted curve, and not with an average of all positions!

### A.3.7 Single channel tracking analysis

The single channel tracking analysis enables the user to track spots in a single channel. To start the single channel tracking analysis the user has to select the **Analysis/Single channel tracking** (or with shortcut Ctrl + A). This will open a new window containing all the steps of the analysis (Fig. 5). The analysis consists of six steps.

- Step 1.* User selects what to analyze (channel, which positions, ROIs per position).
- Step 2.* Visualize tracks that are detected and/or selected by TransTrack.
- Step 3.* Select a threshold to find spots (local maxima) and link them together to make tracks.
- Step 4.* Select tracks either manual or automated (based on different criteria; for example track length).
- Step 5.* Optional: manually curate tracks.
- Step 6.* Clear and close analysis.





**Fig. 5 | Screenshot of TransTrack 'analysis' window.**

In the analysis window the user can perform the single channel tracking analysis. The analysis consists of 6 steps. First, the user selects which channel, positions and (optionally) which ROIs are being analyzed (step 1). Different steps during the remainder of the analysis can be visualized in the 'Visualization' box (step 2). Next, tracks are made through spot detection and linking spots together to form a trajectory (step 3). After tracks are made, they can either be selected manually (step 4a) or tracks can be selected automatically (step 4b). Selected tracks can be manually curated (step 5, optional) and the analysis can be cleared (step 6) to start a new analysis.

During all the steps, the user can set a number of parameters that influence the analysis. Below an extensive description of all the steps and parameters involved is given. Note, that when the analysis is started, all parameters have initial values that could serve as a starting point of an analysis.

### *Step 1.* Select channel, positions and region of interests (ROIs)

Before starting any analysis the user has to choose what to analyze. There are three things to select: (a) the channel used for analysis, (b) the positions that you want to analyze, and (c) ROIs at each position to exclude parts of the image for analysis. While the ROI selection is optional, the analysis channel and the positions used for analysis have to be selected to proceed further with the analysis.

#### a. Choose analysis channel.

The user has to select the channel that is used for the analysis (in which channel it has to find spots and make tracks). This selection is made in the top-left popup menu

which says **Choose Channel**. It is obligatory to select a channel for the analysis to proceed further.

b. Choose positions to analyze.

The positions that the user wants to analyze have to be listed in the **Positions** box. This can be done by typing in the numbers of the positions of interest (separate positions with an enter), or by typing “all” in case you want to analyze all positions. Note, for every analysis it is necessary to give a list of positions, it is not possible to proceed without giving any positions!

c. Select ROIs.

It is possible to only analyze a part of each position by applying one or multiple ROIs to a position. An ROI that you select is only used for that particular position, and it is also possible to use ROIs for some positions, while analyzing the whole image for other positions. An ROI is added to the current position by clicking on the **Add ROI** button. This allows the user to draw an enclosed area in the image, which will be added to the ROI list for that position. There are two options to remove ROIs again. First, a single ROI can be deleted by clicking on the **Delete ROI** button. This allows the user to select an ROI that will be removed (click inside the ROI to remove it). The second option removes all the ROIs at the positions given by the user by clicking on the **Clear ROI** button. The positions where ROIs have to be cleared are given in the **Positions** box.

To use the ROIs during the downstream analysis, the user has to select the tick box called **Use ROIs**. In this case, TransTrack will take for each position the ROI(s) that were selected (if no ROI was selected at a position, it will analyze the whole image), and only tracks in which the entire track remains within an ROI are taken along.

### *Step 2. Visualization of the analysis*

Different steps during the analysis can be visualized. In the visualization panel (Fig. 5, top left) the user can switch on/off the visualization of these steps. The following steps can be visualized.

- **ROIs**. Display the ROIs that are selected by the user at the current position (shows blue outline of the ROI).
- **Peaks**. Display the spots found during step 3 (shows magenta circles centered on the spot).
- **Track paths**. Display the tracks made during step 3 (shows blue lines that

follow the track path).

- **Selected tracks.** Display the tracks that were selected during step 4 (shows green circles centered on the spot; will overlay the magenta circles if both are selected at the same time).
- **Tracknumber.** Display the track numbers of the selected tracks during step 4.

### *Step 3.* Find spots and make tracks

This step is performed when the user clicks on the **Track** button, and involves 3 sub-steps (a-c). All steps are shortly described below. Furthermore, the user can tweak different parameters during these steps and the use of each parameter is explained.

- a. Filtering of the image (low and high bandpass filter).

During this step the image is ‘cleaned’ by using a low and high bandpass filter. The following parameters can be tweaked for the filtering step.

- **Noise size.** The pixel size of noise (this will filter out all small structures up to this size).
- **Object size.** Slightly larger than the pixel size of the object of interest (the filter will filter out large structures).
- **Noise level.** Use as a background subtraction.

- b. Find spots (local maxima) in the filtered image (pixel higher than every other neighbor).

Spots in the filtered image are determined based on that the pixel intensity is higher than the threshold and that the pixel is a local maximum (pixels surrounding it have to have a lower pixel intensity). The following parameter can be tweaked during this step:

- **Threshold.** The threshold value to find a local maximum (in other words: pixel intensity has to be higher).

- c. Link the spots together to make tracks.

During the final step, the spots found in the filtered image are linked together to form tracks. To find the optimal configuration of linked spots, the Hungarian algorithm is used. The following parameters can be tweaked during this step:

- **Tracking distance.** This sets a maximum to the distance a spot can move from

one frame to the next. So, two spots in subsequent frames that are separated by a greater distance than this value, can never be combined in one track.

- **Frame skipping.** Once all the tracks are formed, the algorithm has the possibility to connect two tracks that are not connected in time, but would be in space (the distance between the last frame of the first track and the first frame of the second track is smaller than the **Tracking distance**). So, this gives the possibility to make one track, if several frames are missing. Here, the user can define the amount of frames between two tracks, where it is still possible to connect the two tracks.

#### *Step 4.* Track selection

During this step the user can select the tracks of interest from all the tracks that were made in the previous step. There are two ways of selecting tracks: (a) manual selection and (b) automatic selection.

##### a. Manual selection.

Manual selection of tracks is done by clicking on **Add Track**. Next, the user has to select tracks by clicking (left mouse click!) on the desired tracks in the image. Multiple tracks can be added at once by clicking on multiple tracks, and the user has to press the enter button to stop track selection (this can also be done before clicking on any track if the user accidentally clicked on **Add Track**).

The user can delete a track that was added by clicking on **Delete**. Next, the user has to select the track that needs to be deleted by clicking (left mouse click!) on this track. In case the user accidentally clicked on **Delete**, the user can exit by clicking with a right mouse click.

##### b. Automatic selection

Next to this manual selection of tracks, the user can also choose for an automatic selection by clicking on **Make Tracks**. Initializing this option will remove all current selected tracks at the positions listed in the **Positions** box and select tracks based on the criteria given at those same positions. Note that it is possible to combine an automatic and manual selection, however the user should always start with the automatic selection as all manual selected tracks will be removed once automatic selection is started. Furthermore, the user can also delete tracks that were selected through the automatic track selection by clicking on **Delete**.

The user has a number of criteria he/she can set for the automatic track selection. The four basic criteria are.

- **Start.** the starting frame of analysis, if ticked on all frames before the starting frame will not be analyzed.
- **End.** the last frame of analysis, if ticked on all frames after the end frame will not be analyzed.
- **Overlap.** this option is only used if either **Start**, **End**, or both **Start** and **End** are ticked on. In case **Overlap** is ticked on, all tracks that are partially within the analysis boundaries (as given by the user through **Start** and/or **End**) will be taken along and will be clipped according to the **Start** and/or **End** settings. In case **Overlap** is not ticked on, only the tracks that are completely within the analysis boundaries will be taken along.
- **Minimum Track Length.** The minimum length of a track, if ticked on all tracks that are shorter than this length will be removed.

Next to these four basic criteria there are two more advanced criteria which are: (1) co-localization of a spot (local maximum) in another channel and (2) segmentation of image based on another channel. Both these criteria will be described below.

The user can select tracks based on co-localization with a spot (local maximum) in another channel. If the user ticks on the **Colocalization** option, only tracks that have a co-localization spot (local maximum) in another channel for a certain amount of time will be taken along. The user has to define the following parameters for TransTrack to determine for each track if there is a co-localization spot.

- **Channel.** Choose a channel in which TransTrack has to search for a co-localization spot (this has to be a different channel than the analysis channel).
- **Threshold.** The threshold value for finding spots (local maxima) in the channel selected by the user. For each track, TransTrack will only search in a small area around the track for spots in the 'colocalization channel'. Next, only if a spot is found within 3 pixels of the center of the track it will call that a co-localization spot was present in that frame. This is done for each frame of the track, and in each frame it is indicated if a co-localization spot was present.
- **Length.** The maximum length of the co-localization signal. Based on the threshold set above, TransTrack will determine for each frame if there is a co-localization spot present. Next, it will determine the maximum number of consecutive frames where a co-localization spot was present during the track.

If this number is smaller than the value given by the user (in the **Length** edit box), it will not be regarded as a co-localization event.

In the second advanced criteria, the user can select based on the segmentation of the image performed in another channel than the analysis channel. The segmentation of the image is a simple one, where the image is divided in a high and low intensity region based on a channel given by the user. For example, in case of a DAPI staining, the user can segment the image in a DAPI high (the nuclei) and a DAPI low (the cytoplasm) region. Next, if the user ticks on the **subROIs** option only tracks that are either in the high or in the low region will be taken along. The user has to define the following parameters.

- **Channel.** Choose a channel in which TransTrack has to perform the segmentation of the image.
- **Region.** Choose if tracks in the high or low intensity region should be taken along in the analysis.
- **Threshold.** The threshold value for the segmentation. All pixels with a value lower than the threshold will be put in the intensity low region and all pixels with a value higher than the threshold will be put in the intensity high region. After thresholding a filter is applied to smoothen the segmented image.

### *Step 5. Track curation*

In the final step the user has the option to manually curate tracks that are incorrect. Note that this is an optional step that gives the user the opportunity to improve track quality if this is necessary. Furthermore, this step can also be performed while selecting tracks. For example, the user can select a new track through **Add Track**, inspect if the track is correct, curate it if necessary, and continue with selecting a new track. Only take care that in the case of automatic track selection, all selected tracks will be removed and new ones will be selected and this will also remove all track corrections done up to that point. Therefore, in case of automatic track selection, track repair can only be done after the automatic selection.

There are four different ways of correcting a track, which will all be described below (a-d).

- a. Replace the location of a single frame.

With the replace option the user can change the location of a track in the current frame that is displayed. This option is started by entering the track number of the

track that needs to be replaced (it will find the track in the current frame that is displayed) in the edit box (named #) and by clicking on **Replace**. Next, the user has to select the new coordinates of the track manually by clicking (left-mouse click) in the image at the correct location. In case the user accidentally clicked on **Replace**, the user can exit with a right-mouse click.

b. Extend a track.

With the extend option the user can extend a track manually either at the start of the track or at the end of the track. To select a track, a track number has to be entered (in the same edit box as for the replace option). For extending the start of the track, the user has to click on the **B** ('backward') button and for extending the end of the track the user has to click on the **F** ('forward') button. Next, the frame either before the first frame (in case of clicking on the **B** button) or the frame after the last frame (in case of clicking on the **F** button) will be displayed and the user has to select the location of the spot by clicking with a left-mouse click. The user can extend the track as far as he/she wants by continuously clicking with a left-mouse click at the location of the spot in subsequent frames. The user can stop extension by clicking with a right-mouse click. Furthermore, extension will automatically stop once the first or last frame has been reached.

c. Clip a track.

With the clip option the user can trim a track either at the start or at the end or at both sides at the same time. The user needs to enter a track number (in the edit box named #) and give a lower bound (in the edit box named **lb**) and upper bound (in the edit box named **ub**) that will be used to trim the track. If no lower bound is given, no trimming will occur at the start of the track. Similarly, if no upper bound is given, no trimming will occur at the end of the track. Finally, the trimming of the track is performed by clicking on the **Clip** button.

d. Merge two tracks.

With the merge option the user can merge two tracks together. Track numbers of both tracks have to be given in the edit boxes (named #). It is important that the track which will form the first part of the merged track is listed as first (in the edit box named **1.**) and the track which will form the second part of the merged track is listed as second (in the edit box named **2.**). It is not important if the first track has a higher or lower track number than the second track. Furthermore, (optionally) the user can give limits for which part of the tracks have to be taken in the merged track. When the user selects a limit for the first track, TransTrack will automatically take the first

track from the first frame until this limit. When the user selects a limit for the second track, TransTrack will automatically take the second track from the limit until the last frame of the second track. Therefore, the limit of the first track cannot be higher than the limit of the second track. If no limits are given by the user, TransTrack will take the entire two tracks.

Merging of two tracks is done by clicking on **Merge**. Next, TransTrack will merge the two tracks. However, often either the two tracks don't connect perfectly which results in a gap or the two tracks overlap which results in two conflicting tracks. Both of these problems have to be manually resolved. In the case where the two tracks don't connect perfectly, the user has to manually select the location of the track in the frames with missing information after clicking **Merge**. The displayed image will automatically go to the first frame of the gap and ask the user to click at the correct location. This will continue until all gaps are filled. In the case where the two tracks overlap, the user has to manually select the location of the track in the frames where both tracks are present after clicking **Merge**. Again, the displayed image will automatically go to the first frame with conflicting information and ask the user to click at the correct location for all frames with conflicting information. When all frames with gaps or conflicting information have been resolved, the two tracks will be merged and given the track number of the first track.

### *Step 6. Clear and close analysis*

If the user wants to re-start an analysis or delete an analysis, he/she can either click on the **Clear** button (located in the top) or close the analysis. When the user selects the **Clear** button, only the positions listed in the **Positions** box will be cleared. When the user closes the analysis, all analysis information will be deleted (including ROIs and analysis settings).

### **A.3.8 Find back a track from the analysis**

To find a specific track that was selected in the analysis (during step 4 section A.3.7), the user can select **Analysis/Find track** (or with shortcut Ctrl + F). This will prompt a popup-menu asking the user the following information.

- **Position.** The position of the track of interest.
- **Track number.** The number of the track of interest.

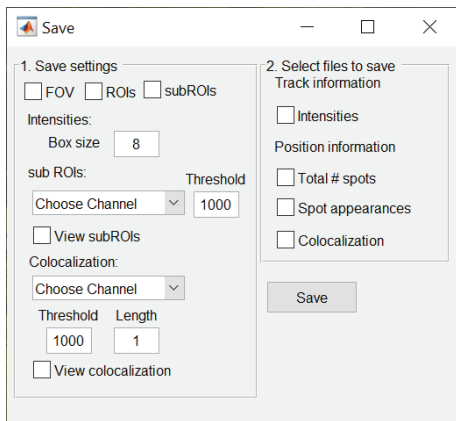
If the track exists, TransTrack will automatically go to the first frame of the track, and a magenta colored circle will be drawn around the track of interest.



### A.3.9 Save ROIs and results from the analysis

Both ROIs that were made as well as results from the single channel tracking analysis (see section A.3.7) can be saved. Both procedures will be described below.

To save ROIs that were made during the analysis, the user has to select **File/Save/ROIs**. A popup-menu will appear asking the user a location for saving and a name for the file. The final output file is a .mat file (which can be opened in MATLAB) and contains information about the outlines of all ROIs at the different positions. This file can also be loaded back into TransTrack if requested by the user (see section A.3.10).



**Fig. 6 | Screenshot of TransTrack 'save' window.**

In the save window the user can save results from the single channel tracking analysis. By choosing between the different options, the user can determine which output data will be saved in the final Excel file.

To save the results from the single channel tracking analysis (see section A.3.7), the user has to select **File/Save/Track data** (or with shortcut Ctrl + S). This will open a new window (Fig. 6), which enables the user to select the exact output that the user wants to extract from the analysis. When the user clicks on the **Save** button (located at the bottom right of the window), a popup-menu will appear asking the user a location for saving and a name for the file. The final output is a .xlsx file and always contains a tab with the metadata used during the analysis and a tab with the general track information. It is important that these two tabs are never modified (contents as well as the title of the tabs). Otherwise, it is not possible anymore to load track data back into TransTrack (see section A.3.10).

The metadata tab in the .xlsx file contains the following information.

- **Row 1.** Shows if a photobleach correction (see section A.3.6) was applied to the different channels when saving the analysis (1 = yes and 0 = no).
- **Rows 2-5.** General info about the movie that was analyzed (number of positions, channels, frames, and dimensions).
- **Row 6.** Which type of analysis was performed (1 = single channel tracking analysis)

- **Rows 7-28.** Settings that were used during the analysis. Corresponds to settings in the analysis window. For all filters, 1 = checkbox was on and 0 = checkbox was off. For subROI selection (**Row 26**) it is reported if tracks in high intensity region (=1) or low intensity region (=2) were taken along (see section A.3.9, step 4b).
- **Rows 29-34.** Settings that were used for saving. Corresponds to settings in the save window.

The general track info tab in the .xlsx file contains the following information.

- **Column 1.** Shows from which position track originates.
- **Column 2.** Shows track number (at each position numbering starts again at 1).
- **Column 3-5.** X, Y, and time coordinates of track.
- **Column 6-end.** Intensities at track coordinates for all different channels. Given are an intensity trace, a background intensity trace (local background of the track), and a background subtracted trace. See below for a detailed description how these intensities are obtained.

Next to the metadata and general track info tab, there are several save options that analyze the tracks in more detail. The following options can be selected: (1) **Intensities**, (2) **Total # spots**, (3) **Spot appearances**, and (4) **Colocalization**. These options are selected with the respective checkboxes in the **Select files to save** panel (Fig. 6).

The **Intensities** option provides a tab in the .xlsx file with all background subtracted intensity traces from one channel (each channel gets an own tab). The traces are ordered by position number and track number. The background subtracted intensities are obtained as follows: first, an intensity trace is made for each track by taking the average pixel intensity of a box drawn around the center of the track. The size of the box can be given by the user at the **Box size** edit box in the **Save settings** panel. Next, a background intensity trace is made for each track by taking the average pixel intensity of boxes (with the same **Box size**) drawn around the central box. Here, only boxes are taken along that have a homogeneous intensity profile (the maximum/minimum pixel value ratio in a box has to be smaller than 4) to exclude boxes that contain other signals that are not true background (e.g. other mRNA molecules). Finally, the background intensity is subtracted from the intensity trace

to get a background subtracted intensity trace. In the intensity tabs for the different channels only the background subtracted intensity trace is listed. The intensity trace and background intensity trace can be found in the general track information.

The **Total # spots** option will provide a tab in the .xlsx file with the amount of tracks present in each frame of the movie. The **Spot appearances** option will provide a tab in the .xlsx file with the amount of tracks that appeared in each frame of the movie (every track is now only counted once, namely when it appears). The **Colocalization** option will provide a tab in the .xlsx file with information about the number of tracks that have a co-localization spot in another channel for a certain amount of time (specified by user). To determine the number of tracks that have a co-localization spot, the user has to select a number of inputs (which are located in the **Save settings** panel at the **Colocalization** option).

- **Choose channel.** Select the channel that TransTrack has to use to look for co-localization spots.
- **Threshold.** The threshold value for finding spots (local maxima) in the channel selected by the user. For each track, TransTrack will only search in a small area around the track for spots in the 'colocalization channel'. Next, only if a spot is found within 3 pixels of the center of the track it will call that a co-localization spot was present in that frame. This is done for each frame of the track, and in each frame it is indicated if a co-localization spot was present.
- **Length.** The minimum length of the co-localization signal. The length of the co-localization signal is determined by taking the maximum number of consecutive frames where a co-localization spot was found during the track.

The results of the co-localization can also be visualized in the image by selecting the **View Colocalization** checkbox. For each track, a green dot (the spot has co-localization with the current settings) or a red dot (the spot has no co-localization with the current settings) will be placed in the center of each track. This visualization can be used as an aid to find the best settings.

For all three options (**Total # spots**, **Spot appearances**, and **Colocalization**) it is required to select which parts of the image the user wants to use to obtain the data. There are two different options: (1) the user wants to determine this for the entire image or (2) the user wants to determine this per ROI (basically: each ROI is treated as a separate position). If any of the three options (**Total # spots**, **Spot appearances**, and **Colocalization**) is selected, it is therefore necessary that the user selects one of the two options (whole image or per ROI) or both. If the user wants to have the

information for the entire image, the user has to select the **FOV** checkbox located at the top of the **Save settings** panel. If the user wants to have the information per ROI, the user has to select the **ROIs** checkbox located at the top of the **Save settings** panel (Fig. 6). Importantly, at least one of the two has to be selected to proceed with saving the data.

Finally, there is an option to split the information (**Total # spots**, **Spot appearances**, and **Colocalization**) in the whole image or per ROI even further. This can be done by selecting the **subROIs** checkbox located at the top of the **Save settings** panel (Fig. 6). When the user selects the **subROIs** checkbox, the image will be segmented in a high and low intensity region based on the settings selected by the user. The user has to select the following settings in the **Save settings** panel at the **subROIs** option.

- **Channel.** Choose a channel in which TransTrack has to perform the segmentation of the image.
- **Threshold.** The threshold value for the segmentation. All pixels with a value lower than the threshold will be put in the intensity low region and all pixels with a value higher than the threshold will be put in the intensity high region. After thresholding a filter is applied to smoothen the segmented image.

The resulting segmentation with the settings selected by the user can be visualized by selecting the **View subROIs** checkbox. In this way the user can choose a good threshold value based on visual inspection.

With the segmentation performed, it is now possible to split the information (**Total # spots**, **Spot appearances**, and **Colocalization**) even further. In Table 2 an example is given how the total number of spots can be quantified in different (sub)regions of the image (whole image, per ROI, high/low intensity region). The same holds true for the spot appearances and number of tracks having a co-localization spot. Table 2 also gives the checkboxes that need to be selected to get the specific tabs and also the final tab names in the .xlsx file.

### A.3.10 Load data from a previous analysis

It is possible to load a previously saved TransTrack analysis session to continue with the analysis or inspect the tracks more closely. Furthermore, it is also possible to load an ROI file to re-use ROIs made and saved previously. Both options will shortly be explained below.

A saved ROI file (as described in section A.3.9) can be loaded into TransTrack by

Output	Checkboxes	Short description
FOV – All	<i>FOV</i>	The total number of spots in each frame for the whole image.
FOV – High	<i>FOV, subROIs</i>	The total number of spots in each frame in the high intensity region of the whole image.
FOV – Low	<i>FOV, subROIs</i>	The total number of spots in each frame in the low intensity region of the whole image.
ROI – All	<i>ROIs</i>	The total number of spots in each frame in each ROI.
ROI – High	<i>ROIs, subROIs</i>	The total number of spots in each frame in the high intensity region of each ROI.
ROI – Low	<i>ROIs, subROIs</i>	The total number of spots in each frame in the low intensity region of each ROI.

**Table 2 | Different options for organization of output data from TransTrack.**

Example of how the total number of spots can be counted in different regions of the image. A different output sheet (left column) will be created in the final Excel file based on which options (middle column) are selected by the user. The user can select one or more of the ‘FOV’, ‘ROIs’, or ‘subROIs’ options at the top of the save window (see Fig. 6) to count the total number of spots in a specific region (right column).

clicking on **File/Load/ROIs**. This will prompt a popup-menu asking the user for the ROI file that needs to be loaded. After selection of the file, the ROIs are loaded into the movie and can be used during further analyses. Note that the ROI file that is loaded should be from a movie with the same number of positions, otherwise it will not be able to load the file. Also, the order of the positions should be the same as at the time the ROI file was saved.

A saved analysis file (as described in section A.3.9) can be loaded into TransTrack by clicking on **File/Load/Track data** (or with shortcut Ctrl + L). A popup-menu will appear where the user has to select the file that needs to be loaded. Loading an analysis file will close any active analysis at this moment, and load the data from the saved file instead. Please note that an analysis file requires to contain two tabs called ‘Metadata’ and ‘General Track info’. These names need to be exact! Furthermore, the metadata file should contain the correct amount of entries (never change the metadata from the saved output!). When all requirements are met, the track data will be loaded and a new analysis window will be started.

## A.4 Discussion

TransTrack is a semi-automated tracking method, aimed to quantify translational activity of single mRNA molecules from single-molecule translation imaging data. Quantifying translation on single mRNA molecules can help to better understand the mechanisms behind translation. For instance, TransTrack has been used to

quantify the amount of alternative translation (e.g. out-of-frame translation) [12], to measure the transition from CBC-bound to eIF4E-bound translation initiation [16], and to assess the effects of siRNAs on translation dynamics [13]. Furthermore, TransTrack can be combined with RiboFitter (see [17] for an overview of this analysis pipeline), to precisely determine the exact initiation times of individual ribosomes on an mRNA molecule. Finally, quantification of translational activity can be used to estimate key translation parameters (such as translation initiation and elongation rates) which in turn can be implemented in stochastic models of translation [18]. Taken together, TransTrack provides a valuable tool in the analysis of single-molecule translation imaging data.

There are many different single-particle tracking methods and multiple of those methods have adapted some hybrid version of automated and manual tracking [14,15]. In this respect, TransTrack is not a unique method to generate single-particle tracks. TransTrack, however, has been specifically designed to analyze single-molecule translation imaging data. Therefore, it is tailored to the requirements and needs of researchers analyzing translation imaging data. Firstly, this is reflected during the analysis in which the user can select the mRNA molecules of choice and curate the tracks which are not followed correctly. Secondly, this is reflected in the output generated by TransTrack. The output proxy's (e.g. intensity, number of mRNA molecules, or their co-localization with other signals) can readily be obtained for the entire data-set, individual cells, or sub-compartments inside a cell (e.g. the nucleus or cytoplasm). Thus, TransTrack's strength lies in the user-friendly design which has been optimized for analysis of single-molecule translation imaging data.

Despite the current disadvantages that come with fully automated single-particle tracking, a labor-free and unbiased analysis method has a strong appeal. To implement a fully automated single-particle tracking method, it will be crucial to improve the reliability of mRNA selection and tracking. There are several approaches that could be taken to improve the reliability of fully automated single-particle tracking methods. Firstly, experimental conditions could be improved to obtain images in which background signals in the mRNA channel are reduced, which would result in better selection and tracking accuracy of mRNA molecules. Secondly, while the mRNA channel contains a lot of background spots, the translation channel hardly contains any background (Fig. 2). Thus, automatic single-particle tracking of the translation signal is presumably more reliable compared to the mRNA signal. Tracking of the translation signal has the disadvantage, though, that non-translating mRNA molecules cannot be tracked. Instead of using traditional single-particle tracking algorithms, it would also be possible to make use of deep learning algorithms, an approach that has recently sparked much interest in the image analysis field [19].

Deep learning has already been applied in image classification, image segmentation, and most notably object tracking. A deep learning approach may fulfill, at last, the promise of a fully automated single-particle tracking method for single-molecule translation imaging data.

## A.5 Methods

TransTrack was designed using MATLAB R2018b. The organization and display of images in TransTrack is inspired by FIJI to simplify the use of TransTrack for the users familiar with FIJI. All code was custom written except for the following functions, which were downloaded from Mathworks file exchange:

- progressbar [20]; version 1.2.0.0 by Steve Hoelzer. Shows progress of ongoing functions. Downloaded from: <https://nl.mathworks.com/matlabcentral/fileexchange/6922-progressbar>.
- SPtrack1.0 [21]; version 1.0.0.0 by saurosaum. Filters image (bpass.m) and finds local maxima (pkfnd.m). For tracking used simpletracker function (see below). Downloaded from: <https://nl.mathworks.com/matlabcentral/fileexchange/15765-sprack1-0>.
- simpletracker [22]; version 1.5.0.0 by Jean-Ives Tinevez. A simple particle tracking algorithm. Downloaded from: <https://nl.mathworks.com/matlabcentral/fileexchange/34040-simple-tracker>.

A list of all other functions can be found in the Supplemental information and the code can be downloaded from: <https://github.com/TanenbaumLab/TransTrack>.

## A.6 Author contributions

SS and MT conceived the project. SS designed TransTrack. SS and MT wrote the manuscript.

## A.7 Acknowledgements

We would like to thank the users of TransTrack for their feedback and testing of TransTrack. Especially, we thank Sanne Boersma, Suzan Ruijtenberg, Tim Hoek, Iris Bally, Bram Verhagen, and Deepak Khuperkar. Several people helped me during the development of TransTrack through discussions and helpful suggestions. For this we would like to thank Eugene Katrukha and Sanne van Velzen.

## A.8 References

- 1 Pichon, X. et al. (2018) A growing toolbox to image gene expression in single cells: sensitive approaches for demanding challenges. *Mol. Cell* 71, 468–480
- 2 Biswas, J. et al. (2019) Fluorescence imaging methods to investigate translation in single cells. *Cold Spring Harb. Perspect. Biol.* 11, a032722

- 3 Yan, X. et al. (2016) Dynamics of translation of single mRNA molecules in vivo. *Cell* 165, 976–989
- 4 Wu, B. et al. (2016) Translation dynamics of single mRNAs in live cells and neurons. *Science* 352, 337–342
- 5 Morisaki, T. et al. (2016) Real-time quantification of single RNA translation dynamics in living cells. *Science* 352, 1425–1429
- 6 Wang, C. et al. (2016) Real-time imaging of translation on single mRNA transcripts in live cells. *Cell* 165, 990–1001
- 7 Pichon, X. et al. (2016) Visualization of single polysomes reveals translation dynamics in living human cells. *J. Cell Biol.* 214, 769–781
- 8 Tanenbaum, M.E. et al. (2014) A protein-tagging system for signal amplification in gene expression and fluorescence imaging. *Cell* 159, 635–646
- 9 Zhao, N. et al. (2019) A genetically encoded probe for imaging nascent and mature HA-tagged proteins in vivo. *Nat. Commun.* 10, 2947
- 10 Chao, J.A. et al. (2008) Structural basis for the coevolution of a viral RNA-protein complex. *Nat. Struct. Mol. Biol.* 15, 103–105
- 11 Bertrand, E. et al. (1998) Localization of ASH1 mRNA particles in living yeast. *Mol. Cell* 2, 437–445
- 12 Boersma, S. et al. (2019) Multi-color single-molecule imaging uncovers extensive heterogeneity in mRNA decoding. *Cell* 178, 458–472
- 13 Ruijtenberg, S. et al. (2020) mRNA structural dynamics shape Argonaute-target interactions. *Nat. Struct. Mol. Biol.* 27, 790–801
- 14 Meijering, E. et al. (2012) Methods for cell and particle tracking. *Methods Enzymol.* 504, 183–200
- 15 Chenouard, N. et al. (2014) Objective comparison of particle tracking methods. *Nat. Methods* 11, 281–290
- 16 Hoek, T.A. et al. (2019) Single-molecule imaging uncovers rules governing nonsense-mediated mRNA decay. *Mol. Cell* 75, 324–339
- 17 Khuperkar, D. et al. (2020) Quantification of mRNA translation in live cells using single-molecule imaging. *Nat. Protoc.* 15, 1371–1398
- 18 Aguilera, L.U. et al. (2019) Computational design and interpretation of single-RNA translation experiments. *PLoS Comput. Biol.* 15, e1007425
- 19 Moen, E. et al. (2019) Deep learning for cellular image analysis. *Nat. Methods* 16, 1233–1246
- 20 Hoelzer, S. progressbar (<https://nl.mathworks.com/matlabcentral/fileexchange/6922-progressbar>). MATLAB Central File Exchange
- 21 saurosum (2019) SPtrack1.0 (<https://nl.mathworks.com/matlabcentral/fileexchange/15765-sptrack1-0>). MATLAB Central File Exchange
- 22 Tinevez, J.-Y. (2019) simpletracker (<https://nl.mathworks.com/matlabcentral/fileexchange/34040-simpletracker>). MATLAB Central File Exchange





## Supplemental information

### Required files for TransTrack

The TransTrack package can be downloaded from <https://github.com/TanenbaumLab/TransTrack>. Here, a complete list of MATLAB guides and functions with short description is provided.

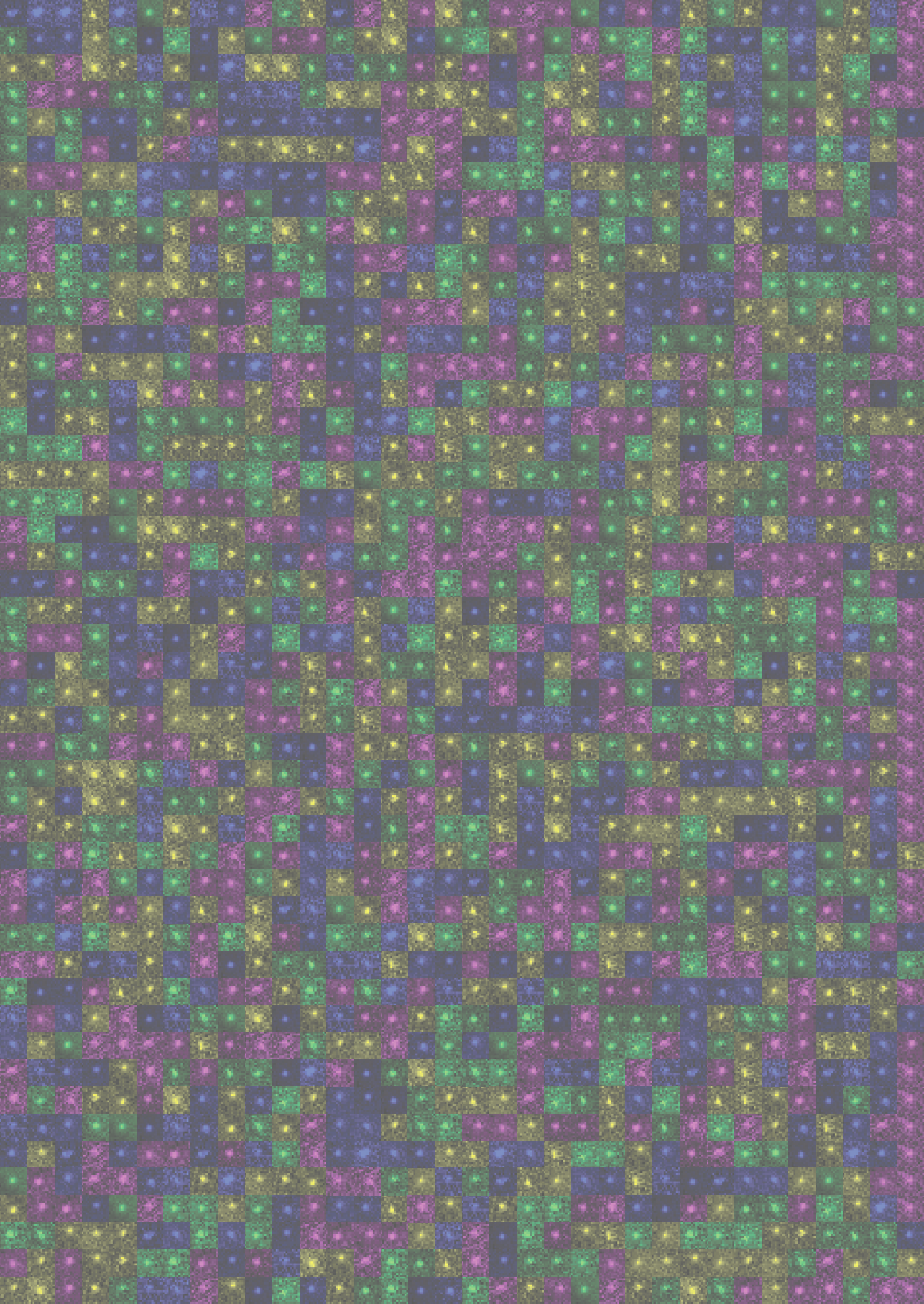
The necessary guide files (both .fig as well as .m files) are as follows.

- **TransTrack.m / TransTrack.fig**. This opens the main window of TransTrack in which the image is displayed.
- **DisplaySettings.m / DisplaySettings.fig**. This opens the window in which brightness and contrast can be regulated (see section A.3.4).
- **Channels.m / Channels.fig**. This opens the window in which channel color display can be regulated (see section A.3.4).
- **SingleChannelTracking.m / SingleChannelTracking.fig**. This opens the window in which the analysis is performed (see section A.3.7).
- **Save.m / Save.fig**. This opens the window in which the current analysis can be saved (see section A.3.9).

The necessary functions are as follows.

- **display.m**. Custom written function that displays the image in the TransTrack main window and with the specific lay-out given by the user (see sections A.3.4 and A.3.7 step 2).
- **intensity\_tracker\_single\_channel.m**. Custom written function that measures the intensity of translation for mRNA tracks (see section A.3.9).
- **make\_tracks.m**. Custom written function that automatically selects tracks based on criteria given by user (see section A.3.7 step 4b).
- **track\_finder.m**. Custom written function that allows user to search a specific track in the image (see section A.3.8).
- **bpass.m / pkfnd.m**. All files that are related to localizing peaks (i.e. local maximum) algorithm version 1.0.0.0 by saurosaum. Filters image (bpass.m) and finds local maxima (pkfnd.m). Downloaded from: <https://nl.mathworks.com/matlabcentral/fileexchange/15765-sptrack1-0>.

- **progressbar.m**. Progressbar algorithm version 1.2.0.0 by Steve Hoelzer. Displays progress. Downloaded from: <https://nl.mathworks.com/matlabcentral/fileexchange/6922-progressbar>.
- **simpletracker.m / hungarianlinker.m / munkres.m**. All files that are related to the simpletracker algorithm version 1.5.0.0 by Jean-Ives Tinevez. A simple particle tracking algorithm. Downloaded from: <https://nl.mathworks.com/matlabcentral/fileexchange/34040-simple-tracker>.





## **Addendum**

Nederlandse samenvatting

About the author

List of publications

Acknowledgements

## Nederlandse samenvatting

De mens is een zogenaamd meercellig of multicellulair organisme. Dit betekent dat het menselijk lichaam is opgebouwd uit een verzameling cellen. Gemiddeld genomen bevat een menselijk lichaam het gigantische aantal van zo'n 20 tot 30 biljoen cellen. Deze cellen zijn niet allemaal hetzelfde: we hebben bijvoorbeeld huidcellen, spiercellen en zenuwcellen. Tegenwoordig wordt er gedacht dat er zelfs meer dan 400 verschillende soort cellen in het menselijk lichaam te vinden zijn. Deze verschillende cellen hebben ook hele verschillende eigenschappen. Zo moeten huidcellen het menselijk lichaam beschermen tegen verschillende externe factoren, terwijl spiercellen moeten samentrekken zodat wij kunnen bewegen. Bijna alle cellen in het menselijk lichaam hebben echter ook één ding gemeen: ze bevatten allemaal hetzelfde DNA. Het verschil tussen deze cellen wordt dus niet verklaard door het DNA dat ze bevatten. In plaats daarvan komt het verschil doordat deze cellen het DNA op een andere manier gebruiken of 'uitlezen'. Het DNA bestaat uit een enorme hoeveelheid nucleotiden (in totaal wel 3.2 miljard!) die samen coderen voor zo'n 30.000 genen. Terwijl een aantal van deze genen nodig zijn voor de basis functies van iedere cel (deze genen worden ook wel huishoudgenen genoemd), zijn er ook een heel aantal genen die alleen nodig zijn voor een specifiek cel type. Door een bepaalde set genen tot expressie te brengen, verkrijgt een cel zijn specifieke eigenschappen en kan het zijn specifieke functie vervullen. Samengenomen noemen we dit *genexpressie* en de regulatie van genexpressie (ook wel genregulatie genoemd) zorgt er dus voor dat een huidcel een huidcel is en een spiercel een spiercel.

Om een gen tot expressie te brengen moet het DNA in de cel worden uitgelezen, zodat RNA moleculen en eiwitten kunnen worden geproduceerd. Tezamen vormen RNA moleculen en eiwitten (waarbij het voornamelijk om de eiwitten draait) de moleculaire machinetjes die de meeste functies in de cel vervullen. We kunnen twee stappen onderscheiden die nodig zijn om RNA moleculen en eiwitten te maken. In de eerste stap wordt het DNA gekopieerd om een RNA molecuul te synthetiseren (dit wordt ook wel transcriptie genoemd). In sommige gevallen kan dit RNA molecuul direct een functie vervullen, in andere gevallen fungeert het RNA molecuul als een tussen product (in deze gevallen wordt het mRNA genoemd of 'messenger' RNA). In deze gevallen wordt het mRNA molecuul in de tweede stap gebruikt om met behulp van ribosomen een eiwit te synthetiseren (dit wordt ook wel translatie genoemd). Samen zorgen de processen van transcriptie en translatie er dus voor dat een gen tot expressie komt. Om de expressie van een gen te reguleren heeft een cel talloze opties tot zijn beschikking. Zo kan een cel de hoeveelheid transcriptie of translatie reguleren en kan de cel bestaande RNA moleculen en eiwitten ook weer snel afbreken. Hoewel er al ontzettend veel onderzoek gedaan is naar de regulatie van genexpressie, zijn

er ook nog steeds veel open vragen. Het beter begrijpen van genregulatie is niet alleen van belang voor een groter inzicht in hoe het menselijk lichaam functioneert, maar ook om te begrijpen wat er misgaat in verschillende ziektes zoals kanker of neurodegeneratieve aandoeningen (bijv. Alzheimer).

Om genexpressie en de regulatie hiervan te bestuderen, hebben onderzoekers veel verschillende methodes tot hun beschikking. In de afgelopen jaren zijn er in hoog tempo ook nieuwe methodes ontwikkeld, die de mogelijkheid geven om genexpressie nog gedetailleerder te kunnen onderzoeken. Eén belangrijke ontwikkeling is dat nieuwe methodes steeds gevoeliger zijn wat de mogelijkheid biedt om enkele cellen of zelfs enkele moleculen te bekijken. Op deze manier kunnen we veel meer informatie verzamelen over de processen die ervoor zorgen dat een gen op de juiste plaats en tijd tot expressie komt. In dit proefschrift hebben we gebruik gemaakt van een aantal van deze nieuwe methodes en met behulp van deze methodes hebben wij drie verschillende processen bestudeerd die betrokken zijn bij de regulatie van genexpressie.

In **hoofdstuk 3** hebben we het functioneren van Argonaute2 (AGO2) in menselijke cellen bestudeerd. AGO2 is een eiwit dat met behulp van een klein stukje RNA (een zogenaemde miRNA of siRNA) op zoek gaat naar specifieke mRNA moleculen. Nadat AGO2 zijn doelwit gevonden heeft, knipt AGO2 het mRNA molecuul in twee stukken waarna het snel verder wordt afgebroken door andere eiwitten. Op deze manier zorgt AGO2 er uiteindelijk voor dat de expressie van een gen drastisch afneemt (zonder mRNA moleculen kunnen er immers geen eiwitten meer worden gemaakt van dit gen). Er is nog veel onbekend over het functioneren van AGO2 in levende cellen, onder meer doordat het niet mogelijk was om heel gedetailleerd naar het functioneren van AGO2 te kijken. Met behulp van een revolutionaire nieuwe techniek slaagden wij erin om het knippen van een enkel mRNA molecuul door AGO2 live te kunnen visualiseren in een menselijke cel. Op deze manier konden wij bestuderen hoe snel AGO2 een mRNA molecuul knipt en welke factoren hierbij een belangrijke rol spelen. Dit heeft verschillende nieuwe inzichten opgeleverd. Het belangrijkste nieuwe inzicht is dat het voor AGO2 vaak niet zo eenvoudig is om het mRNA molecuul te vinden. Een mRNA molecuul blijkt namelijk vaak ontoegankelijk te zijn voor AGO2, omdat het mRNA molecuul een kluwen vormt die eerst ontrafeld dient te worden voordat AGO2 zijn werk kan doen. Wij ontdekten ook dat ribosomen hierbij een belangrijke rol spelen: een ribosoom ontrafeld een mRNA molecuul namelijk tijdens de translatie en dit zorgt ervoor dat het mRNA molecuul kortstondig toegankelijk wordt voor AGO2. Door deze observaties begrijpen we nu beter hoe AGO2 functioneert in een menselijke cel. Daarnaast zou het ook heel goed kunnen dat deze principes niet alleen voor het AGO2 eiwit gelden, maar mogelijk

ook voor andere soorten eiwitten die aan RNA moleculen binden.

In **hoofdstuk 4** hebben we ons gericht op een heel ander proces, namelijk de celcyclus. Tijdens de celcyclus bereidt een cel zich voor op een celdeling, waarbij een cel zich splitst in twee dochter cellen. Celdelingen zijn essentieel tijdens de ontwikkeling (waarbij een eencellig embryo zich ontwikkelt tot een meercellig organisme) en voor onderhoud (huidcellen moeten bijvoorbeeld continu vernieuwd blijven worden door middel van celdelingen). Voor de meeste celdelingen is het noodzakelijk dat al het materiaal (inclusief het DNA) in een cel wordt gedupliceerd en dat dit materiaal vervolgens heel precies wordt verdeeld over de twee dochter cellen. Om dit proces ordentelijk te laten verlopen, heeft de cel de verschillende stappen die hiervoor nodig zijn onderverdeeld in een aantal fases waarbij een cel alleen naar een volgende fase kan gaan als de voorafgaande fase correct is afgerond. Voor iedere stap zijn specifieke eiwitten nodig en het is uiterst belangrijk dat deze eiwitten alleen aanwezig zijn in de juiste fase van de celcyclus (de cel wilt namelijk niet dat een bepaald proces plaats vindt in de verkeerde fase van de celcyclus). Om dit voor elkaar te krijgen heeft de cel zeer uitgebreide regulatie mechanismes opgetuigd, waarbij eiwitten alleen gemaakt worden in de fase wanneer ze nodig zijn en ook direct weer worden afgebroken zodra ze overbodig zijn. In **hoofdstuk 4** hebben wij een nieuw regulatie mechanisme ontdekt dat specifiek een rol speelt wanneer een cel net gedeeld is en een nieuwe celcyclus betreden heeft. Met behulp van ‘RNA sequencing’ hebben wij het volledige transcriptoom (de verzameling van alle RNA moleculen in een cel) in kaart gebracht van cellen die net gedeeld zijn. Door te bekijken hoe het transcriptoom verandert wanneer cellen een nieuwe celcyclus ingaan, ontdekten wij dat de hoeveelheid RNA moleculen van een aantal genen drastisch afneemt direct na een celdeling en dat dit komt doordat deze RNA moleculen zeer snel worden afgebroken door de cel. Verder onderzoek is nodig om te achterhalen op welke manier de cel specifiek deze RNA moleculen afbreekt en wat de functie is van deze ‘golf’ van RNA afbraak.

Tenslotte hebben wij in **hoofdstuk 5** weer een ander aspect van genregulatie bekeken. Hier hebben wij namelijk onderzocht hoe transcriptie initiatie (waarna een RNA molecuul gesynthetiseerd wordt) door de cel gereguleerd wordt. Transcriptie is de eerste stap in genexpressie en is daarom een zeer streng gereguleerde stap, waarbij bepaald wordt welke genen tot expressie komen in de cel. Er zijn talloze factoren betrokken bij de regulatie van transcriptie initiatie, maar in **hoofdstuk 5** hebben wij ons specifiek gefocust op twee belangrijke elementen: promotoren en enhancers. Een promotor is een DNA sequentie die direct naast een gen ligt en die allerlei essentiële factoren voor transcriptie rekruteert. In menselijke cellen is het in veel gevallen echter zo dat de promotor alleen niet genoeg is om transcriptie te initiëren. In deze gevallen is een tweede DNA sequentie nodig die een enhancer (letterlijk: versterker)



wordt genoemd. De enhancer rekruteert, net als de promotor, factoren die essentieel zijn om het proces van transcriptie te kunnen beginnen. Verrassend genoeg liggen enhancer sequenties vaak niet direct naast een promotor. In plaats daarvan kunnen ze soms wel honderd duizenden basen (i.e. 1 base = 1 nucleotide in het DNA) verder op liggen. Hierdoor moeten een promotor en enhancer sequentie eerst dicht bij elkaar komen voordat transcriptie kan plaats vinden. Uit verschillende onderzoeken is ook gebleken dat dit door de cel als een regulatie mechanisme wordt gebruikt: een cel kan een gen ‘aan’ of ‘uit’ zetten door de interactie tussen een promotor en enhancer te controleren.

Er is echter nog veel onbekend over de interacties tussen een promotor en enhancer en voor een groot gedeelte komt dit door het gebrek aan methodes die promotor en enhancer interacties over tijd kunnen volgen. Dit in ogenschouw nemende, hebben wij in **hoofdstuk 5** gewerkt aan het ontwikkelen van een nieuwe methode en analyse om interacties tussen een promotor en enhancer live te kunnen visualiseren en kwantificeren. Wij hebben een ‘gereedschapskist’ gebouwd waarmee we zowel de locatie van een promotor als enhancer kunnen volgen (en hiermee kunnen wij dus zien wanneer zij dicht bij elkaar zijn). Daarnaast kunnen wij ook de transcriptionele activiteit van het gen meten en daarmee kunnen wij de interacties tussen een promotor en enhancer correleren aan de transcriptionele activiteit van het gen. Met deze gereedschapskist hopen wij in de toekomst bij te kunnen dragen aan het beantwoorden van een aantal vragen omtrent promotor en enhancer communicatie.

Samenvattend hebben we in dit proefschrift drie heel verschillende processen bestudeert. Er zijn twee soorten raakvlakken tussen deze drie studies. Allereerst zijn alle processen die wij bestudeerd hebben betrokken bij de regulatie van genexpressie. Ten tweede, hebben wij in de drie studies nieuwe methodes gebruikt, die ervoor zorgden dat we op een hele gedetailleerde manier de verschillende processen konden bestuderen en dit heeft verschillende nieuwe inzichten opgeleverd. Natuurlijk zijn er minstens even veel nieuwe vragen opgeroepen als er zijn beantwoord, en wij hopen dat de bevindingen uit dit onderzoek een inspiratie kunnen vormen voor het opstellen van nieuwe vragen en hypotheses.

### **About the author**

Stijn Sonneveld was born on May 31, 1992 in Zaandam. He attended the Zaanlands Lyceum in Zaandam, from which he graduated in 2010. In 2013, he received his bachelor Medische Natuurwetenschappen cum laude from the Vrije Universiteit in Amsterdam. This was followed by a master Medical Natural Sciences from which he graduated cum laude in 2016. During his master, he did an internship in the group of prof. dr. Gijsje Koenderink at the AMOLF institute in Amsterdam, and he moved for one year to Dresden, Germany, where he did an internship in the group of prof. dr. Stephan Grill. After this second internship, Stijn moved back to the Netherlands and began his PhD research in the group of dr. Marvin Tanenbaum at the Hubrecht Institute for Developmental Biology and Stem Cell Research in Utrecht. The results of this work are described in this thesis.

## List of publications

Boersma, S.\* , Khuperkar, D.\* , Verhagen, B.M.P., **Sonneveld, S.**, Grimm, J.B., Lavis, L.D., and Tanenbaum, M.E. (2019). Multi-color single-molecule imaging uncovers extensive heterogeneity in mRNA decoding. *Cell* 178, 458–472.

Hoek, T.A.\* , Khuperkar, D.\* , Lindeboom, R.G.H., **Sonneveld, S.**, Verhagen, B.M.P., Boersma, S., Vermeulen, M., and Tanenbaum, M.E. (2019). Single-molecule imaging uncovers rules governing nonsense-mediated mRNA decay. *Mol. Cell* 75, 324–339.

Khuperkar, D.\* , Hoek, T.A.\* , **Sonneveld, S.**, Verhagen, B.M.P., Boersma, S., and Tanenbaum, M.E. (2020). Quantification of mRNA translation in live cells using single-molecule imaging. *Nat. Protoc.* 15, 1371–1398.

Ruijtenberg, S.\* , **Sonneveld, S.\*** , Cui, T.J., Logister, I., de Steenwinkel, D., Xiao, Y., MacRae, I.J., Joo, C., and Tanenbaum, M.E. (2020). mRNA structural dynamics shape Argonaute-target interactions. *Nat. Struct. Mol. Biol.* 27, 790–801.

**Sonneveld, S.\*** , Verhagen, B.M.P.\* , and Tanenbaum, M.E. (2020). Heterogeneity in mRNA translation. *Trends Cell Biol.* 30, 606–618.

Krenning, L.\* , **Sonneveld, S.\*** , and Tanenbaum, M.E. (2021). Time-resolved single-cell sequencing identifies multiple waves of mRNA decay during mitotic exit. *BioRxiv*.

*Submitted*

\*Equal contribution



## Acknowledgements

De afgelopen 5 jaar heb ik met ontzettend veel plezier aan het onderzoek gewerkt dat uiteindelijk in dit boekje terecht is gekomen. Dit resultaat zou niet mogelijk zijn geweest zonder de begeleiding en ondersteuning die ik van talloze mensen heb mogen ontvangen. Op deze plek wil ik graag iedereen die direct of indirect een bijdrage heeft geleverd heel erg bedanken.

Marvin, toen ik zo'n 5½ jaar geleden langs was geweest bij jou wist ik zeker: hier wil ik mijn PhD doen. Ik was dan ook erg blij toen ik hoorde dat dit wederzijds was en ik heb er zeker geen spijt van gekregen. Door jouw grote betrokkenheid bij mijn projecten, heb ik ontzettend veel kunnen leren. Besprekingen met jou waren zeer inspirerend en zorgden altijd weer voor vele nieuwe ideeën waar ik dan weer verder aan kon gaan werken. Jouw persoonlijke feedback op tal van verschillende zaken (hoe een experiment aan te pakken of hoe het beste een verhaal op te schrijven) hebben mijn wetenschappelijke vaardigheden erg verbeterd. Ook was het een voorrecht om aanwezig te zijn in die eerste jaren dat jij jouw lab startte. Het was mooi om dit proces van dichtbij te zien en hier ook zelf deel aan te nemen. Ik wil ook graag mijn promotor, Wouter, bedanken. Het DNA looping project was een directe samenwerking met jouw groep en ik wil je graag bedanken voor je begeleiding en enthousiasme. Het project is uiteindelijk een stuk taaier gebleken dan we van te voren gehoopt hadden, maar we blijven nog steeds stappen voorwaarts zetten (en soms ook weer een paar stappen terug). Ik wil ook nog graag mijn beoordelingscommissie bedanken voor het beoordelen van dit proefschrift en de opponenten met wie ik op 3 november verder over dit werk mag discussiëren.

Tijdens mijn projecten heb ik met verschillende mensen mogen samenwerken en zonder deze mensen was ik zeker niet zo ver gekomen. Suzan, niet alleen heb ik ontzettend genoten van onze samenwerking tijdens het siRNA project, ook was jij het die mij wegwijs maakte in het lab. Nog relatief onervaren (cel kweek, hoe gaat dat?), heb ik alle benodigde technieken van jou mogen leren gedurende mijn eerste maanden in het lab. Ook had ik het geluk dat ik mocht aansluiten bij jouw project dat op dat moment al de eerste resultaten begon op te leveren, waardoor ik direct aan de slag kon met leuke biologische vragen. Uiteindelijk bleken de antwoorden op deze vragen toch iets ingewikkelder dan we in eerste instantie dachten en was het ook nodig om extra expertise in te schakelen. Dit heeft een hele leuke internationale samenwerking opgeleverd waarvoor ik graag Chirlmin, Thijs, Ian en Yao wil bedanken. Niet alleen hebben we expertise buiten het lab gezocht, ook hebben we zelf nieuwe technieken geleerd. Ive, heel erg bedankt voor het opzetten van de Northern blots. De resultaten die dit opleverden bleken een essentieel puzzelstukje te vormen. Tenslotte wil ik ook

de studenten, Dion en Loes, bedanken die over de jaren aan dit project hebben mee geholpen.

Ik heb het geluk gehad dat ik niet met één maar met twee post-docs heb mogen samenwerken. Lenno, ik vond het ontzettend leuk dat ik bij jouw project kon aansluiten en dat ik de mogelijkheid kreeg om er allerlei analyses op los te laten. Het was voor mij heel leerzaam om jouw ideeën te horen over de data en op deze manier de analyses verder te verbeteren of aan te vullen. Er is uiteindelijk een heel mooi verhaal uitgekomen vind ik. De weg naar publicatie is helaas wat langer en hobbeler gebleken dan we van te voren misschien gehoopt hadden, maar we zijn er nu echt bijna (toch?). Ook hier wil ik graag de studenten, Ran en Maarten, bedanken die, onder leiding van Lenno, aan dit project hebben bijgedragen.

Niet al mijn projecten voerde ik uit binnen Marvins groep, één project was een directe samenwerking met Wouters groep en in het bijzonder met Erica en Sjoerd. Erica, ik heb met jou tijdens het eerste gedeelte van mijn promotie mogen samenwerken en ik wil je daarvoor heel erg bedanken. We hebben heel wat hobbels moeten overwinnen bij het maken van de knock-ins (we hebben denk ik zo'n beetje alle gekke dingen gezien die konden gebeuren), maar ondanks de tegenslagen was het altijd heel gezellig om met jou samen te werken. Sjoerd, met jou heb ik het tweede gedeelte van mijn promotie samengewerkt en aan jou heb ik nu ook het stokje doorgegeven. Ook jou wil ik bedanken voor alle hulp en voor jouw bijdrage aan dit project. Dit bestond uit heel wat uurtjes klonen bekijken en nieuwe knock-in rondes voorbereiden. Ik vond het ook erg leuk om jou wat van de microscopie te laten zien en ik hoop dat je een leuke en succesvolle PhD verder hebt. Ik wil ook graag nog een aantal andere mensen van Wouters lab, Niels, Peter en Chris, bedanken die aan dit project hebben bijgedragen.

Ik zou graag alle huidige en voormalige leden van het Tanenbaum lab, Tim, Suzan, Lenno, Deepak, Sanne, Ive, Bram, Lucas, Rupa, Max, Iris, Huib, Dhanushika, Micha en Sora, willen bedanken. Ik kon altijd terecht bij jullie voor vragen of hulp. Ook waren jullie altijd erg geïnteresseerd in mijn projecten en heb ik heel veel gehad aan jullie suggesties tijdens de lab besprekingen, in de wandelgangen of bij de koffie automaat. Nog belangrijker dan al dit is dat jullie ook fantastische collega's waren. Er is een ontzettend grote saamhorigheid en we hebben zoveel leuke dingen samen gedaan. De lunch bezoeken aan het pannenkoekenhuis met het traditionele rondtollende ei op het einde, alle 2 euro uitdagingen, de pizza game in de late avond of het tafelvoetballen (tot overlast van sommigen), het is slechts een greep uit de lange lijst. Tezamen zorgde dit ervoor dat ik (bijna) elke dag met veel plezier naar werk ging en dat ik ook gemotiveerd was in tijden dat de resultaten wat tegenvielen.

Ik wil hier ook alle studenten bedanken die we over de jaren heen gehad hebben. In het bijzonder wil ik Elena bedanken, die ik een klein jaar heb mogen begeleiden tijdens haar Master stage. Het was heel erg leuk om zo'n enthousiast iemand te mogen begeleiden en ik hoop dat je er net zoveel van hebt geleerd als ik.

Het Hubrecht instituut is een fantastische locatie geweest om mijn promotieonderzoek uit te voeren. Ik wil graag de verschillende diensten binnen het Hubrecht (van de civiele dienst tot de catering) bedanken. Zonder jullie was het überhaupt niet mogelijk geweest om dit onderzoek uit te voeren. Ook wil ik de FACS faciliteit bedanken (Stefan en Reinier; en Ive in het afgelopen jaar) voor het sorteren van mijn cellen en Litha voor de hulp bij het indienen van mijn proefschrift (ook hier waren toch wat hobbels te nemen). Het Hubrecht is niet alleen een plek met geweldige faciliteiten, het is ook een plek waar ik ontzettend veel leuke dingen heb gedaan en heel veel leuke mensen heb mogen ontmoeten die de afgelopen 5 jaar kleur hebben gegeven. Hier volgt slechts een willekeurige en onvolledige selectie van enkele hoogtepunten van de afgelopen 5 jaar. Ik heb erg genoten van mijn jaar in de personeelsvereniging en dit bleek ook een goede manier te zijn om het Hubrecht beter te leren kennen. Andere hoogtepunten waren de reis naar Mallorca, de wintersport trip naar Winterberg, en natuurlijk de vrijdagmiddag borrels (met als ultiem hoogtepunt onze eigen bierpong borrels, inclusief posters).

Toen ik aan het Hubrecht begon was niet alles aan 'onderzoek doen' nieuw voor mij en dit heeft mij heel erg geholpen tijdens mijn promotie. Hierbij wil ik graag iedereen bedanken die op de een of andere manier heeft bijgedragen aan mijn academische vorming. Ik wil graag mijn stagebegeleiders in AMOLF, Bart, Nicholas en Gijsje, en in Dresden, Dora, Enrico en Stephan, bedanken voor hun supervisie. Hier heb ik mijn eerste wetenschappelijke stappen mogen zetten. Grappig genoeg zou dit hele avontuur uiteindelijk nooit zijn begonnen (of het had in ieder geval een hele andere vorm aangenomen) zonder één persoon. Ik had namelijk nog nooit van het Hubrecht instituut gehoord totdat Teije mij hierover vertelde toen ik aan het eind van mijn stage in Dresden op zoek was naar een promotie plek. Ik wil je daarvoor bedanken, want ik heb een hele mooie tijd hier gehad en ik zou mijn promotie niet ergens anders hebben willen doen.

Tenslotte zou ik ook nog graag familie en vrienden willen bedanken. Jullie waren er altijd voor mij in de afgelopen 5 jaar. Het was erg fijn om (gedeelde) PhD ervaringen te kunnen bespreken of om het gewoon even heel ergens anders over te hebben. Mam, Pap, Anne en Daan, bij jullie kan ik altijd terecht en ik bof ontzettend met jullie.



