In the Footsteps of the Rugged Trail of Transcription

Ruiqi Han

Printing: ProefschriftMaken || www.proefschriftmaken.nl
Lay out: Lieve Stiers
Cover Design: Caiqiu Yang and Ruiqi Han
ISBN 978-94-6380-190-4

In the Footsteps of the Rugged Trail of Transcription

In de voetstappen van het ruige pad van transcriptie

Thesis

to obtain the degree of Doctor from the
Erasmus University Rotterdam
by command of the
rector magnificus

Prof.dr. R.C.M.E. Engels

and in accordance with the decision of the Doctorate Board.

The public defence shall be held on

Thursday 14th of February, 2019 at 13:30 hrs
by
Ruiqi Han

born in Changchun, China

Erasmus University Rotterdam

Doctoral Committee

Promotor: prof.dr. R. Agami

Other members: prof.dr. C.P. Verrijzer
prof.dr. W.T. Zwart
Dr. R.A. Poot

Copromotor: Dr. B. Slododin

TABLE OF CONTENTS

List of Abbreviations

bp	Base pair
DNA	Deoxyribonucleic acid
RNA	Ribonucleic acid
Gro-seq	Global run-on sequencing
RNA Pol II	RNA polymerase II
PAM	protospacer adjacent motif
m6A	N6-methyl-adenosine
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
TE	Translational efficiency
mRNA	Messenger RNA
eRNA	Enhancer RNA
TF	Transcription factor
UTR	Untranslated region
OIS	Oncogene-induced senescence

CHAPTER 1

General introduction

From DNA, RNA to protein

Nature has created numerous different forms of life on the planet, as seen by their appearances, living habits, and reproduction methods. Among all the existing lifeforms and creatures, there is one common trace of life that remains after billions of years of evolution: DNA. The history of our past and the path to our future are largely determined by genetic information, which is considered the DNA sequence of bases along a nucleic-acid chain (Berg et al., 2002, chapter 5). In human cells, there are two copies of each chromosome, a result of the fusion of sperm and egg cells. To interpret its information, the double helix DNA is transcribed into a single-strand RNA. The series of events occur inside the nucleus of every cell. Those RNA codes for specific proteins, namely mRNA, are then exported to the cytoplasm to be translated into proteins. The cellular and physiological functions of our body are highly dependent on the proper production and regulation of proteins.

Transcription regulation

The human genome contains about 20,000 protein-coding genes and even more non-coding genes (Djebali et al., 2012) through which the transcription takes place. The process of transcription initiating at the sequences is termed ‘promoters’. These include the transcription start site, TATA box, and transcription regulatory regions as the core elements (Lee and Young, 2000). The strength of the promoters partially governs the transcriptional potential. The transcription is further regarded as a combinatorial interaction of many different processes, including the recruitment of chromatin remodelers, polymerases, acetyltransferases, methyltransferases, etc (Coulon et al., 2013). The mystery of these collaborative actions remains mostly unknown considering the dynamics of transcription.

Each human cell comprises more than six billion nucleotides packed into a tiny nucleus of just a few μm (Gillooly et al., 2015). The contents of the nucleus require highly organized structures to maintain viable cellular functions. The cells handle the long stretches of DNA by wrapping them around the chromatin, or more specifically, the nucleosomes. Each nucleosome includes two pairs of the four core histone proteins, H2A, H2B, H3, and H4, with two meters of DNA compacted inside (Li and Reinberg, 2011). The physical barrier of histones blocks the accessibility of transcription factors. In addition, this also prevents the RNA polymerases from initiating transcription. This suggests how exactly DNA is wrapped around the histones is tightly controlled. These histone molecules harbor various post-translation modifications (methylation, acetylation, phosphorylation,

ubiquitination, etc.) that govern the compaction and accessibility of DNA (Lawrence et al., 2016) (see also Figure 1a).

Typically, the types of chromatin are divided into heterochromatin and euchromatin. Heterochromatin is denser and transcription-inactive, whereas euchromatin has a more relaxed structure for active transcription (Lelli et al., 2012). Within the gene-rich euchromatin, the transcription of coding genes is largely coordinated by transcription factors (TF) and RNA polymerase II (Pol II) machinery (Figure 1b). The promoter region acts as a dock for sequence-specific TFs, which determine numerous cellular functions, including differentiation, stress-response, and proliferation. Moreover, a single TF regulate different genes in a distinct cellular context, suggesting the regulatory network of transcription is very dynamic (Lambert et al., 2018). The recruitment of different TFs and cofactors relays the genetic information to different transcription initiation factors. Upon the stepwise assembly of the transcription initiation complex (TFIIB, TFIIF, TFIIE, TFIIH, RNA Pol II, etc.) (Kadonaga, 2004), RNA Pol II is released to elongate along the DNA template (Figure 1c).

While RNA Pol II elongates along the DNA, the splicing of the pre-mRNA happens co-transcriptionally and requires multiple spliceosomes to remove introns from the nascent transcript (Herzel et al., 2017) (Figure 1d). A mammalian spliceosome consists of U1, U2, U4, U5, and U6 small nuclear RNPs (snRNPs), together with a large set of supplementary factors (Stark and Lührmann, 2006). Assembly of the spliceosome is associated with the C-terminal domain (CTD) of RNA Pol II. The phosphorylation state of the CTD (RNA Pol II0) enhances splicing by facilitating the binding of spliceosomes (Hocine et al., 2010). The interaction between the transcription machinery and the spliceosome suggests a link between the availability of the transcript exposed to the spliceosomes and the transcription rate. Abnormal transcription elongation rate affects the inclusion or exclusion of exons, many of which are found in tumors (Fong et al., 2014). Hence, an optimal process for transcription elongation is needed for appropriate pre-mRNA splicing. Before the mature mRNA reaches the cytoplasm for translation, it needs to be efficiently exported (Figure 1e). SR proteins are affiliated with the pre-mRNA during splicing and help to guide the mature mRNA into the cytoplasm further (Moore and Proudfoot, 2009). Furthermore, the THO/TREX complex is coupled with the mRNA co-transcriptionally, which facilitates efficient export (Sträßer et al., 2002).

As discussed, transcription initiation, elongation, splicing, and export are tightly coupled together, where they collectively influence the outcome of transcription.

The inherited genetic information is not only deciphered into mRNA but is also escorted to the cytoplasm through various factors.

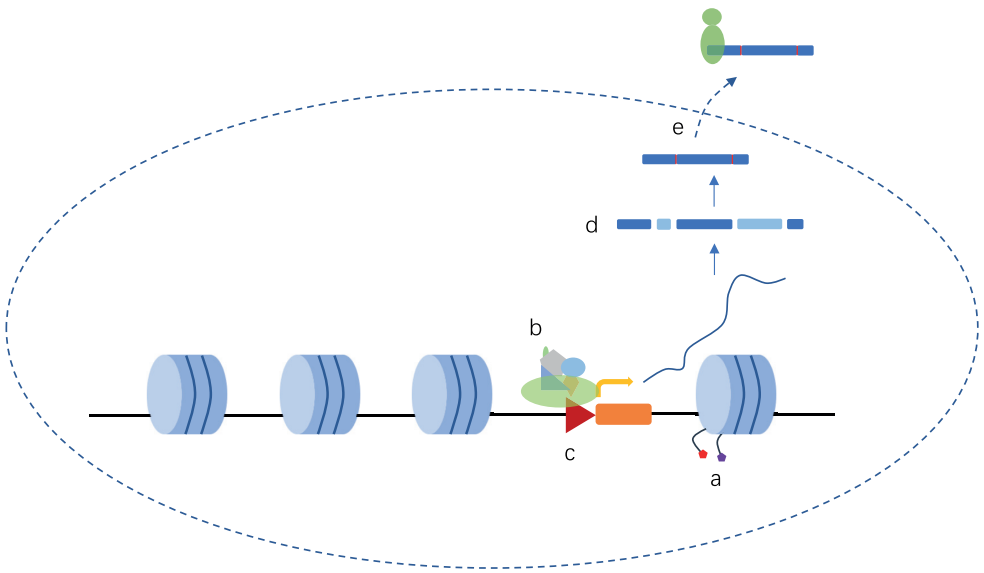


Figure 1

Sketch of eukaryotic transcription regulation. Inside the nucleus of eukaryotic cells, DNA is wrapped around histone molecules, which dictates the accessibility of DNA sequences to various transcription factors. Upon histone modifications (e.g., methylation, acetylation) depending on cellular context, the dynamics of chromatin structures altered, resulting in an active or inhibitory transcription state (a). In the case of active transcription, transcription factors scan for the binding platforms at the promoter regions of a gene (b), which eventually recruit RNA polymerase to initiate the process of transcription (c). Primary messenger RNA (mRNA) products are produced with both exons and introns. These immediate products are then spliced to remove the non-coding regions of introns, which could lead to alternative splicing, generating mRNA variants (d). Once the mature mRNA is manufactured within the nucleus, these molecules require efficient export into the cytoplasm for further translation into proteins (e).

Nature's selection for variations, epigenetics

As human beings, 99% of our genetic information is shared between us. Still, our appearances, physiological characteristics, and living habits vary significantly depending on where we live, how we eat, and what we do. Interestingly, even twins with identical genetic background differ in many aspects. Such differences have been explained in the study of epigenetics - a type of gene regulatory mechanism that functions without altering the DNA sequence. DNA methylation is one of

the first studied epigenetic regulations. Among all the different biochemical modifications, cytosine is considered a dynamic nucleotide, which is frequently found methylated at the fifth carbon (5mC). Such methylations exist primarily in the form of CpG dinucleotides, of which 60-80% are found methylated in the human genome (Lister et al., 2009). DNA-methyl transferases (DNMT) is the family of proteins that catalyze the process. The distribution of these CpG sequences is biased toward the promoter regions of coding genes. Studies have shown its essential role in the epigenetic regulation of development, tumorigenesis, and genomic imprinting (Li et al., 1993; Linhart et al., 2007; Okano et al., 1999).

As mentioned earlier, histone modification is another fold of epigenetic regulation related to transcriptional output. Almost all forms of modifications are associated with transcription, suggesting the complex dynamics of transcription control via conformation changes of nucleosomes (Kouzarides, 2007). Furthermore, specific histone modifications are valuable markers for identifying functional DNA elements. For instance, H3K4me3 is associated with active transcribing promoter regions, while H3K4me1 and H3k27ac indicate typical enhancer elements (Calo and Wysocka, 2013). These modifications on DNA sequences shed light on how epigenetic information is coded in the nucleus.

Likewise, epigenetic regulation touches further into the cytoplasm in which the courier of the genetic information is interpreted. mRNA undergoes extensive regulations on its rate of synthesis and decay; however, in the last two decades, our knowledge about the RNA biology has expanded enormously. MicroRNA (miRNA) is a family of non-coding RNA of ~22 nucleotides long. In the nucleus, miRNA is transcribed by RNA Pol II, processed by RNase III proteins, Drosha and Dicer, where it is finally exported by Exportin 5. The short transcript works by targeting the 3' untranslated region (UTR) of mRNA via its complementary sequence at the 5' end (Bartel, 2009). Upon target recognition, miRNA is assembled into an RNA-induced silencing complex (RISC), inducing translation inhibition and mRNA decay (Ha and Kim, 2014). A single miRNA could target hundreds of putative mRNA strands, affecting their regular functions (Baek et al., 2008).

Like DNA methylation, RNA also sustains different biochemical modifications, including N6-methyladenosine (m6A), N1-methyladenosine (m1A), N6, 2'-O-dimethyladenosine (m6Am), etc. As the most abundant RNA modification, m6A can be found at every 700-800 nucleotides, accounting for 0.2-0.6% of all adenosines in mammalian RNA populations (Roundtree et al., 2017). The development of new methods allows transcriptome-wide quantification of m6A content across different organisms (Dominissini et al., 2012; Meyer et al., 2012),

and even at single nucleotide resolution (Linder et al., 2015). The biogenesis of m6A involves different sets of proteins, including writers (METLL3/14, WTAP), erasers (FTO, ALKBH5), and readers (YTHDF1/2, HNRNPC). Interruptions of these key factors have been reported to influence the level of m6A and alter its cellular functions (Batista et al., 2014; Jia et al., 2011; Liu et al., 2014). Given the localization of m6A effectors in the nucleus and cytoplasm, the role of m6A has been widely researched. METTL3/14, FTO, WTAP are localized at nuclear speckles, and control alternative splicing (Ping et al., 2014; Zhao et al., 2014). Also, cellular mRNA stability could be regulated through m6A modification via m6A reader-assisted RNA decay (Wang et al., 2013; Wang et al., 2014). More recently, m6A profiling indicates the enrichment of m6A modifications within exons and around stop codons (Dominissini et al., 2012), suggesting it could exhibit regulation on translation. m6A readers recognize both m6A in the 5'UTR and 3'UTR, promoting cap-independent and cap-dependent translation (Meyer et al., 2015; Wang et al., 2015). Conversely, as discussed later in this thesis, we have identified an inhibitory role of m6A on translational efficiency through coupled deposition of m6A with RNA Pol II (Slobodin et al., 2017).

Emerging power of CRISPR-Cas9

The field of gene editing has arisen in the past five years, indicating a bright future for disease treatment, embryo adjustment, crop improvement, etc. Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR), originated from a type of adaptive immune system used by bacteria against foreign viruses. It further incorporates DNA sequences from the invading organism into its own genome for a memorable immune response (Fineran and Charpentier, 2012; Wiedenheft et al., 2012). The invading DNA is processed into the CRISPR repeat array and transcribed into CRISPR RNA (crRNA) as photospacer sequences, serving as a guide for the transactivating CRISPR RNA (tracrRNA) and Cas9 nuclease to cleave the target DNA (Deltcheva et al., 2011). To minimize self-cleavage of the integrated photospacer sequences, the cleavage only happens on the invading DNA through which the photospacer is next to a photospacer adjacent motif (PAM). In cells, cleavage of the DNA creates double-strand breaks, allowing error-prone DNA repairs (e.g., non-homologous end joining, NHEJ) to introduce deletions or insertions. As a result, this disrupts the coding sequences of the target gene (see also Figure 2). Significant efforts have been made to generate a simple toolbox for biological studies, resulting in a series of genome-wide genetic screen studies in human and mouse cells (Koike-Yusa et al., 2014; Shalem et al., 2014; Wang et al., 2014).

Alternative applications have been developed to harness gene expression owing to the tight interaction of Cas9 nuclease on its target DNA. CRISPR interference (CRISPRi) was developed to stimulate or inhibit the transcription of a specific gene, using a modified dead Cas9 (dCas9) that has already lost its nuclease activity. By tethering a transcription activator (e.g., VP64) or a suppressor (KRAB) to dCas9, researchers have successfully manipulated the transcriptional profiles of different genes (Maeder et al., 2013; Mali et al., 2013; Qi et al., 2013). Furthermore, when fusing dCas9 with DNA methylation proteins (DNMT3A), researchers have successfully deposited methylation at the target loci and altered gene expression

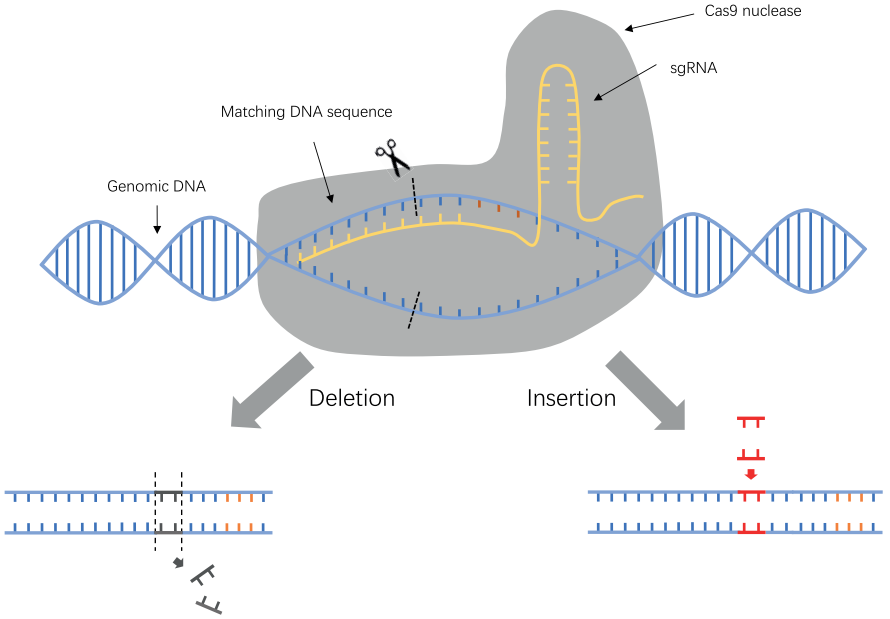


Figure 2

Schematic view of how CRISPR-Cas9 modifies DNA sequences. A complex of crRNA, tracrRNA, and Cas9 nuclease (upper panel in yellow, grey) is formed to search for the complementary sequences on the host genome (upper panel in blue). Once bound to the target sequence adjacent to the PAM sequence (upper panel in red), Cas9 cleaves three nucleotides upstream of the PAM sequence, causing double-strand break. Mutations at the target loci are introduced via NHEJ, leading to deletions or insertions.

(Amabile et al., 2016; Liu et al., 2016). Lastly, using CRISPR to target transcription factor-binding sites at enhancer elements, we have achieved programmable

enhancer control in human cells and identified novel regulatory elements during senescence (Han et al., 2018; Korkmaz et al., 2016).

The complex life of AP1

Activating protein 1 (AP1) was one of the first identified transcription factors, whose discovery widened our knowledge on how transcription factors activate the transcription of genes bearing AP1 binding sites, during cell proliferation and transformation (Angel and Karin, 1991). The AP1 family comprises several groups of dimeric proteins with structurally related leucine zipper domains: JUN, FOS, ATF and MAF family members(Shaulian and Karin, 2002). Functional DNA binding requires the dimerization of AP1 proteins. Although JUN proteins could form homo- and heterodimers, FOS proteins can only bind to JUN proteins to assemble a tight conformation (Hess, 2004). It has recently been suggested that AP1 could serve as co-factors in modulating chromatin dynamics. In macrophages, it binds to C/EBP factors to organize enhancer activities and dictate cell identities (Heinz et al., 2010); and in fibroblasts, AP1 plays a pivotal role in displacing nucleosomes, granting more accessibility to other TFs (Vierbuchen et al., 2017). Therefore, asides from promoting transcription at genes, the activities of AP1 at non-coding regions (e.g., enhancers) might reveal novel regulatory networks of the genome.

Cellular senescence

The phenotypes of senescence have been well-known for centuries. Humans have continuously witnessed the cycle of life and death. However, the scientific description of senescence was only proposed five decades ago by Hayflick and Moorhead (Hayflick and Moorhead, 1961). They discovered that non-tumorigenic cells cultured *in vitro* have a limited proliferation span, in contrast to the infinite proliferation of cancer cells. The emergence of cellular senescence is due to the shortening of the telomeres located at the end of each chromosome. Given the substantial interests in understanding the hidden mechanisms behind cellular senescence, extensive researche have uncovered a large part of the puzzle. Various stimuli activate p53, the guardian of cellular proliferation, and its downstream cyclin-dependent kinase inhibitors p16 (CDKN2A) and p21 (CDKN1A). Activation of the CDK inhibitors causes cell cycle arrest, together with a series of senescence markers. These include the degradation of the nuclear lamina, heterochromatinization of E2F targets, and senescence-associated secretory phenotype (SASP) (He and Sharpless, 2017). Once the stimuli persist, the cells

eventually enter an irreversible growth arrest (Stein et al., 1999). Although the outcome of cellular senescence is similar to different stimuli, the responsible factors can vary. According to the different types of stimuli, cellular senescence has been grouped into multiple subtypes. 1) **Replicative senescence** refers to the condition of multiple cell division, resulting in reduced proliferation capacity (Hayflick, 1965; Hayflick and Moorhead, 1961). 2) **DNA-damage induced senescence** describes the situation in which the cells sense an overload of irreparable DNA damage, leading to the execution of cells by either senescence or apoptosis (Hernandez-Segura et al., 2018). 3) **Oncogene-induced senescence (OIS)** is a type of senescence with excessive activation of oncogenes (e.g., RAS, BRAF), normally adopted from genomic mutations (Sharpless and Sherr, 2015).

The occurrence of OIS *in vivo* has been debated for years considering it was first discovered *in vitro* (Serrano et al., 1997). Some have found that mouse embryonic fibroblast cells (MEF) with ectopic expression of RAS become immortal instead of triggering OIS, which suggests that the overexpression of RAS was not adequate to initiate OIS in MEF (Trotman et al., 2003). In contrast, several studies proposed that mutations of different oncogenes (KRAS, BRAF) trigger OIS in human and mouse tumor models (Courtois-Cox et al., 2006; Michaloglou et al., 2005). The function of OIS during tumor development remains to be elucidated, yet it provides some insights on potential therapies.

The scope of this thesis

This thesis explores multiple aspects of the transcription regulatory network with regards to human cells - From the initiation of transcription to the efficiency of translation. **Chapter 2** attempts to establish a connection between the transcription and translation processes. By employing multiple *in vitro* experiments, a positive correlation between transcription and translation was first discovered, where it was solely controlled via transcription rate. The study leads to the identification of one of the first known functions of m6A at coding regions. While seemingly counterintuitive to the current recognition about m6A on translation, we augment the role of m6A on translation via the discovery of its co-transcriptional deposition. **Chapter 3** reviews the connection between the transcription steps of splicing, export, decay, and translation of mRNA, reinforcing the hypothesis that transcription is not an independent event, and is instead linked to the entire life cycle of mRNA. **Chapter 4** investigates the functional role of AP1 in enhancer regions during OIS. The study results in the identification of a novel enhancer with an AP1 binding motif regulating senescence through its

target gene FOXF1. Although extensively characterized, AP1 and FOXF1 are not reported to act as regulators of senescence. A new trans-regulatory network of genes to counterbalance the effect of oncogene activation was uncovered. In **Chapter 5**, a general discussion about the current view of the field is conducted. Finally, some outlooks are raised to potentially generate a better understanding of transcriptional regulation.

Reference

Amabile, A., Migliara, A., Capasso, P., Biffi, M., Cittaro, D., Naldini, L., and Lombardo, A. (2016). Inheritable Silencing of Endogenous Genes by Hit-and-Run Targeted Epigenetic Editing. Cell 167, 219–232.e14.

Angel, P., and Karin, M. (1991). The role of Jun, Fos and the AP-1 complex in cell-proliferation and transformation. BBA - Rev. Cancer 1072, 129–157.

Baek, D., Villén, J., Shin, C., Camargo, F.D., Gygi, S.P., and Bartel, D.P. (2008). The impact of microRNAs on protein output. Nature 455, 64–71.

Bartel, D.P. (2009). MicroRNAs: Target Recognition and Regulatory Functions. Cell 136, 215–233.

Batista, P.J., Molinie, B., Wang, J., Qu, K., Zhang, J., Li, L., Bouley, D.M., Lujan, E., Haddad, B., Daneshvar, K., et al. (2014). M6A RNA modification controls cell fate transition in mammalian embryonic stem cells. Cell Stem Cell 15, 707–719.

Berg, J.M., Tymoczko, J.L., and Stryer, L. (2002). Biochemistry. W H Free. New York., 320–323.

Bregman, A., Avraham-Kelbert, M., Barkai, O., Duek, L., Guterman, A., and Choder, M. (2011). Promoter elements regulate cytoplasmic mRNA decay. Cell 147, 1473–1483.

Calo, E., and Wysocka, J. (2013). Modification of Enhancer Chromatin: What, How, and Why? Mol. Cell 49, 825–837.

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

Courtois-Cox, S., Genther Williams, S.M., Reczek, E.E., Johnson, B.W., McGillicuddy, L.T., Johannessen, C.M., Hollstein, P.E., MacCollin, M., and Cichowski, K. (2006). A negative feedback signaling network underlies oncogene-induced senescence. Cancer Cell 10, 459–472.

Deltcheva, E., Chylinski, K., Sharma, C.M., Gonzales, K., Chao, Y., Pirzada, Z.A., Eckert, M.R., Vogel, J., and Charpentier, E. (2011). CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III. Nature 471, 602–607.

Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., et al. (2012). Landscape of transcription in human cells. Nature 489, 101–108.

Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M., Ungar, L., Osenberg, S., Cesarkas, K., Jacob-Hirsch, J., Amariglio, N., Kupiec, M., et al. (2012). Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. Nature 485, 201–206.

Fineran, P.C., and Charpentier, E. (2012). Memory of viral infections by CRISPR-Cas adaptive immune systems: Acquisition of new information. Virology 434, 202–209.

Fong, N., Kim, H., Zhou, Y., Ji, X., Qiu, J., Saldi, T., Diener, K., Jones, K., Fu, X.D., and Bentley, D.L. (2014). Pre-mRNA splicing is facilitated by an optimal RNA polymerase II elongation rate. Genes Dev. 28, 2663–2676.

Gillooly, J.F., Hein, A., and Damiani, R. (2015). Nuclear DNA content varies with cell size across human cell types. Cold Spring Harb. Perspect. Biol. 7, 1–27.

Ha, M., and Kim, V.N. (2014). Regulation of microRNA biogenesis. Nat. Rev. Mol. Cell Biol. 15, 509–524.

Han, R., Li, L., Ugalde, A.P., Tal, A., Manber, Z., Barbera, E.P., Chiara, V. Della, Elkon, R., and Agami, R. (2018). Functional CRISPR screen identifies AP1-associated enhancer regulating FOXF1 to modulate oncogene-induced senescence. Genome Biol. 19, 118.

Harel-Sharvit, L., Eldad, N., Haimovich, G., Barkai, O., Duek, L., and Choder, M. (2010). RNA polymerase II subunits link transcription and mRNA decay to translation. Cell 143, 552–563.

Hayflick, L. (1965). The limited in vitro lifetime of human diploid cell strains. Exp. Cell Res. 37, 614–636.

Hayflick, L., and Moorhead, P.S. (1961). The serial cultivation of human diploid cell strains. Exp. Cell Res. 25, 585–621.

He, S., and Sharpless, N.E. (2017). Senescence in Health and Disease. Cell 169, 1000–1011.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. Mol. Cell 38, 576–589.

Hernandez-Segura, A., Nehme, J., and Demaria, M. (2018). Hallmarks of Cellular Senescence. Trends Cell Biol. *28*, 436–453.

Herzel, L., Ottoz, D.S.M., Alpert, T., and Neugebauer, K.M. (2017). Splicing and transcription touch base: Co-transcriptional spliceosome assembly and function. Nat. Rev. Mol. Cell Biol. *18*, 637–650.

Hess, J. (2004). AP-1 subunits: quarrel and harmony among siblings. J. Cell Sci. *117*, 5965–5973.

Hocine, S., Singer, R.H., and Grünwald, D. (2010). RNA processing and export. Cold Spring Harb. Perspect. Biol. *2*.

Jia, G., Fu, Y., Zhao, X., Dai, Q., Zheng, G., Yang, Y., Yi, C., Lindahl, T., Pan, T., Yang, Y.G., et al. (2011). N6-Methyladenosine in nuclear RNA is a major substrate of the obesity-associated FTO. Nat. Chem. Biol. *7*, 885–887.

Kadonaga, J.T. (2004). Regulation of RNA Polymerase II Transcription by Sequence-Specific DNA Binding Factors. Cell *116*, 247–257.

Koike-Yusa, H., Li, Y., Tan, E.P., Velasco-Herrera, M.D.C., and Yusa, K. (2014). Genome-wide recessive genetic screening in mammalian cells with a lentiviral CRISPR-guide RNA library. Nat. Biotechnol. *32*, 267–273.

Korkmaz, G., Lopes, R., Ugalde, A.P., Nevedomskaya, E., Han, R., Myacheva, K., Zwart, W., Elkon, R., and Agami, R. (2016). Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9. Nat. Biotechnol. *34*, 1–10.

Kouzarides, T. (2007). Chromatin Modifications and Their Function. Cell *128*, 693–705.

Lambert, S.A., Jolma, A., Campitelli, L.F., Das, P.K., Yin, Y., Albu, M., Chen, X., Taipale, J., Hughes, T.R., and Weirauch, M.T. (2018). The Human Transcription Factors. Cell *172*, 650–665.

Lawrence, M., Daujat, S., and Schneider, R. (2016). Lateral Thinking: How Histone Modifications Regulate Gene Expression. Trends Genet. *32*, 42–56.

Lee, T.I., and Young, R.A. (2000). Transcription of Eukaryotic Protein-Coding Genes. Annu. Rev. Genet. *34*, 77–137.

Lelli, K.M., Slattery, M., and Mann, R.S. (2012). Disentangling the Many Layers of Eukaryotic Transcriptional Regulation. Annu. Rev. Genet. *46*, 43–68.

Li, G., and Reinberg, D. (2011). Chromatin higher-order structures and gene regulation. Curr. Opin. Genet. Dev. *21*, 175–186.

Li, E., Beard, C., and Jaenisch, R. (1993). Role for DNA methylation in genomic imprinting. Nature *366*, 362–365.

Linder, B., Grozhik, A. V., Olarerin-George, A.O., Meydan, C., Mason, C.E., and Jaffrey, S.R. (2015). Single-nucleotide-resolution mapping of m6A and m6Am throughout the transcriptome. Nat. Methods *12*, 767–772.

Linhart, H.G., Lin, H., Yamada, Y., Moran, E., Steine, E.J., Gokhale, S., Lo, G., Cantu, E., Ehrich, M., He, T., et al. (2007). Dnmt3b promotes tumorigenesis in vivo by gene-specific de novo methylation and transcriptional silencing. Genes Dev. *21*, 3110–3122.

Lister, R., Pelizzola, M., Dowen, R.H., Hawkins, R.D., Hon, G., Tonti-Filippini, J., Nery, J.R., Lee, L., Ye, Z., Ngo, Q.M., et al. (2009). Human DNA methylomes at base resolution show widespread epigenomic differences. Nature *462*, 315–322.

Liu, J., Yue, Y., Han, D., Wang, X., Fu, Y., Zhang, L., Jia, G., Yu, M., Lu, Z., Deng, X., et al. (2014). A METTL3-METTL14 complex mediates mammalian nuclear RNA N6-adenosine methylation. Nat. Chem. Biol. *10*, 93–95.

Liu, X.S., Wu, H., Ji, X., Stelzer, Y., Wu, X., Czauderna, S., Shu, J., Dadon, D., Young, R.A., and Jaenisch, R. (2016). Editing DNA Methylation in the Mammalian Genome. Cell *167*, 233–247.e17.

Maeder, M.L., Linder, S.J., Cascio, V.M., Fu, Y., Ho, Q.H., and Joung, J.K. (2013). CRISPR RNA-guided activation of endogenous human genes. Nat. Methods *10*, 977–979.

Mali, P., Aach, J., Stranges, P.B., Esvelt, K.M., Moosburner, M., Kosuri, S., Yang, L., and Church, G.M. (2013). CAS9 transcriptional activators for target specificity screening and paired nickases for cooperative genome engineering. Nat. Biotechnol. *31*, 833–838.

Meyer, K.D., Saletore, Y., Zumbo, P., Elemento, O., Mason, C.E., and Jaffrey, S.R. (2012). Comprehensive analysis of mRNA methylation reveals enrichment in 3??? UTRs and near stop codons. Cell *149*, 1635–1646.

Meyer, K.D., Patil, D.P., Zhou, J., Zinoviev, A., and Skabkin, M.A. (2015). 5′ UTR m 6 A Promotes Cap-Independent Translation. Cell.

Michaloglou, C., Vredeveld, L.C.W., Soengas, M.S., Denoyelle, C., Kuilman, T., Van Der Horst, C.M.A.M., Majoor, D.M., Shay, J.W., Mooi, W.J., and Peeper, D.S. (2005). BRAFE600-associated senescence-like cell cycle arrest of human naevi. Nature *436*, 720–724.

Moore, M.J., and Proudfoot, N.J. (2009). Pre-mRNA Processing Reaches Back toTranscription and Ahead to Translation. Cell *136*, 688–700.

Okano, M., Bell, D.W., Haber, D.A., and Li, E. (1999). DNA methyltransferases Dnmt3a and Dnmt3b are essential for de novo methylation and mammalian development. Cell *99*, 247–257.

Ping, X.L., Sun, B.F., Wang, L., Xiao, W., Yang, X., Wang, W.J., Adhikari, S., Shi, Y., Lv, Y., Chen, Y.S., et al. (2014). Mammalian WTAP is a regulatory subunit of the RNA N6-methyladenosine methyltransferase. Cell Res. *24*, 177–189.

Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., and Lim, W.A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell *152*, 1173–1183.

Roundtree, I.A., Evans, M.E., Pan, T., and He, C. (2017). Dynamic RNA Modifications in Gene Expression Regulation. Cell *169*, 1187–1200.

Serrano, M., Lin, A.W., McCurrach, M.E., Beach, D., and Lowe, S.W. (1997). Oncogenic ras provokes premature cell senescence associated with accumulation of p53 and p16(INK4a). Cell *88*, 593–602.

Shalem, O., Sanjana, N.E., Hartenian, E., Shi, X., Scott, D.A., Mikkelsen, T.S., Heckl, D., Ebert, B.L., Root, D.E., Doench, J.G., et al. (2014). Genome-scale CRISPR-Cas9 knockout screening in human cells. Science (80-.). *343*, 84–87.

Sharpless, N.E., and Sherr, C.J. (2015). Forging a signature of in vivo senescence. Nat. Rev. Cancer *15*, 397–408.

Shaulian, E., and Karin, M. (2002). AP-1 as a regulator of cell life and death. Nat Cell Biol *4*, E131-6.

Slobodin, B., Han, R., Calderone, V., Vrielink, J.A.F.O., Loayza-Puch, F., Elkon, R., and Agami, R. (2017). Transcription Impacts the Efficiency of mRNA Translation via Co-transcriptional N6-adenosine Methylation. Cell *169*, 326–337.e12.

Stark, H., and Lührmann, R. (2006). CRYO-ELECTRON MICROSCOPY OF SPLICEOSOMAL COMPONENTS. Annu. Rev. Biophys. Biomol. Struct. *35*, 435–457.

Stein, G.H., Drullinger, L.F., Soulard, A., and Dulić, V. (1999). Differential Roles for Cyclin-Dependent Kinase Inhibitors p21 and p16 in the Mechanisms of Senescence and Differentiation in Human Fibroblasts. Mol. Cell. Biol. *19*, 2109–2117.

Sträßer, K., Masuda, S., Mason, P., Pfannstiel, J., Oppizzi, M., Rodriguez-Navarro, S., Rondón, A.G., Aguilera, A., Struhl, K., Reed, R., et al. (2002). TREX is a conserved complex coupling transcription with messenger RNA export. Nature *417*, 304–308.

Trotman, L.C., Niki, M., Dotan, Z.A., Koutcher, J.A., Di Cristofano, A., Xiao, A., Khoo, A.S., Roy-Burman, P., Greenberg, N.M., Van Dyke, T., et al. (2003). Pten dose dictates cancer progression in the prostate. PLoS Biol. *1*.

Vierbuchen, T., Ling, E., Cowley, C.J., Couch, C.H., Wang, X., Harmin, D.A., Roberts, C.W.M.M., and Greenberg, M.E. (2017). AP-1 Transcription Factors and the BAF Complex Mediate Signal-Dependent Enhancer Selection. Mol. Cell *68*, 1067–1082.e12.

Wang, T., Wei, J.J., Sabatini, D.M., and Lander, E.S. (2014a). Genetic screens in human cells using the CRISPR-Cas9 system. Science *343*, 80–84.

Wang, X., Lu, Z., Gomez, A., Hon, G.C., Yue, Y., Han, D., Fu, Y., Parisien, M., Dai, Q., Jia, G., et al. (2013). N6-methyladenosine-dependent regulation of messenger RNA stability. Nature *505*, 117–120.

Wang, X., Zhao, B.S., Roundtree, I.A., Lu, Z., Han, D., Ma, H., Weng, X., Chen, K., Shi, H., and He, C. (2015). N^6-methyladenosine modulates messenger RNA translation efficiency. Cell *161*, 1388–1399.

Wang, Y., Li, Y., Toth, J.I., Petroski, M.D., Zhang, Z., and Zhao, J.C. (2014b). N6-methyladenosine modification destabilizes developmental regulators in embryonic stem cells. Nat. Cell Biol. *16*, 191–198.

Wiedenheft, B., Sternberg, S.H., and Doudna, J.A. (2012). RNA-guided genetic silencing systems in bacteria and archaea. Nature *482*, 331–338.

Zhao, X., Yang, Y., Sun, B.F., Shi, Y., Yang, X., Xiao, W., Hao, Y.J., Ping, X.L., Chen, Y.S., Wang, W.J., et al. (2014). FTO-dependent demethylation of N6-methyladenosine regulates mRNA splicing and is required for adipogenesis. Cell Res. *24*, 1403–1419.

Zid, B.M., and O'Shea, E.K. (2014). Promoter sequences direct cytoplasmic localization and translation of mRNAs during starvation in yeast. Nature *514*, 117–121.

CHAPTER 2

Transcription impacts the efficacy of mRNA translation via co-transcriptional N6-adenosine methylation

Boris Slobodin[1,4,5], Ruiqi Han[1,4], Vittorio Calderone[1], Joachim A.F. Oude Vrielink[1], Fabricio Loayza-Puch[1], Ran Elkon[3,5], and Reuven Agami[1,2,5,6]

[1]Division of Oncogenomics, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands.
[2]Department of Genetics, Erasmus University Medical Center, Wytemaweg 80, 3015 CN Rotterdam, The Netherlands.
[3]Department of Human Molecular Genetics and Biochemistry, Sackler School of Medicine, Tel Aviv University, Tel Aviv 69978, Israel.
[4]Co-first authors
[5]Co-corresponding authors
[6]Lead contact
Correspondence: boris.slobodin@gmail.com; ranel@tauex.tau.ac.il; r.agami@nki.nl

Adapted from Cell (2017), 169: 326-337.

Summary

Transcription and translation are two main pillars of gene expression. Due to the different timings, spots of action and mechanisms of regulation, these processes are mainly regarded as distinct and generally uncoupled, despite serving a common purpose. Here we sought for a possible connection between transcription and translation. Employing an unbiased screen of multiple human promoters, we identified a positive effect of TATA box on translation and a general coupling between mRNA expression and translational efficacy. Using CRISPR-Cas9-mediated approach, genome-wide analyses and *in vitro* experiments, we show that the rate of transcription regulates the efficacy of translation. Furthermore, we demonstrate that m^{6}A modification of mRNAs is co-transcriptional and depends upon the dynamics of the transcribing RNAPII. Suboptimal transcription rates lead to elevated m^{6}A content, which may result in reduced translation. This study uncovers a general and widespread link between transcription and translation that is governed by epigenetic modification of mRNAs.

Keywords

Transcription, translation efficacy, N6-adenosine methylation, m^{6}A, TATA, RNAPII, gene regulation.

Introduction

Transcription of genome-encoded information into mRNA and translation of mRNA into a functional protein are the main layers of gene expression. Due to the existential need to adjust gene expression to both intracellular requirements and extracellular stimuli, both processes are subject to regulation at multiple levels. Transcription is a highly controlled process that is extensively regulated at the levels of initiation, elongation, and termination. Recent studies in eukaryotes linked transcription to other levels of mRNA regulation, such as alternative splicing (Dujardin et al., 2014), polyadenylation (Oktaba et al., 2015), localization and translation (Zid and O'Shea, 2014), and degradation (Dori-Bachash et al., 2012). While splicing and polyadenylation are thought to be co-transcriptional, and therefore could be directly affected by the RNAPII dynamics, the effect on translation and degradation, which have distinct spatial and temporal dynamics, is more complicated to perceive. A recently formulated model explains the "imprinting" role of transcription by co-transcriptional recruitment of "coordinator" proteins (Haimovich et al., 2013), which accompany the synthesized transcript and are capable of regulating its future fate.

Translation of mRNAs is controlled mainly via initiation (Sonenberg and Hinnebusch, 2009) and elongation (Richter and Coller, 2015). Although several recent studies suggested certain levels of dependency between transcription and translation (Elfakess and Dikstein, 2008, Harel-Sharvit et al., 2010, Tamarkin-Ben-Harush et al., 2014, Zid and O'Shea, 2014), it is not very clear whether these are limited to certain subgroups of mRNAs or represent a general link. In general, transcription and translation are still regarded as mutually independent processes, characterized by different timings, cellular locations, functional complexes and mechanisms of action.

N6-methyladenosine (m^{6}A) is considered to be one of the most abundant RNA modifications, detected in thousands of human transcripts (Dominissini et al., 2012, Meyer et al., 2012). Several recent studies connected m^{6}A to the regulation of splicing (Xiao et al., 2016), translation (Meyer et al., 2015, Wang et al., 2015) and degradation (Wang et al., 2014). Overall, a growing body of evidence suggests that m^{6}A plays an important role in multiple levels of mRNA regulation.

In this study, we tested a hypothesis suggesting a direct flow of information from transcription to translation. Combining an unbiased screen for examination of the effect of human promoters on mRNA translation and genome-wide analyses, we identified a positive correlation between mRNA expression and translation

efficacy (TE) and found that rate of transcription positively affects TE. Moreover, we observed that transcriptional dynamics are reflected in the relative deposition of m^{6}A on mRNAs that affects translation. This study establishes a general and robust link between transcription and translation of mRNAs, and provides a mechanistic insight regarding the way transcription epigenetically "imprints" mRNA molecules.

Results

A reporter vector system to examine transcription-translation relationship

To examine the relationship between transcription and translation, we set to determine the effect of different human promoters on the translation of a reporter gene (Renilla luciferase, Rluc). For this purpose, we defined promoters as 0.5-2.5Kb long regions characterized by high H3K4Me3 and low H3K4Me1 epigenetic marks upstream of transcriptional start sites (TSSs), supported by RNA-seq data of MCF7 cells (Loayza-Puch et al., 2013). To make our screen versatile and diverse, we cloned promoters from genes connected to stress response, autophagy, ER metabolism, metastasis, as well as multiple transcription factors. While cloning the promoter sequences, we avoided regions extending downstream the respective TSSs as these might result in the inclusion of additional 5' untranslated regions (5'UTRs) into the reporter transcripts, potentially complicating the interpretation of results. As each of the cloned promoters drives the expression of the same reporter gene, we affiliated every promoter to a unique 10-nt barcode (cloned in the 3'UTR of Rluc gene) to follow its expression in a pool of *Rluc* mRNAs (Figure S1A). Following these guidelines, we cloned 135 human promoters (Table S1) to create a library named Pro-Lib, where most of the promoters were associated with two or more different barcodes to provide higher experimental confidence. As expected, cloned promoter regions substantially induced Rluc expression (Figure S1B). For normalization of expression, Pro-Lib included an additional reporter gene, Firefly luciferase (Fluc) used as inner control (Figure S1A). Last, we employed FRT recombination system and used competent Flp-In MCF7 cells to stably integrate a single copy of a Pro-Lib vector per cell in the identical genetic locus in order to avoid any possible influence of different chromatin neighborhoods on the reporter gene expression.

After establishing the library in a stable population, we performed a polysomal profiling experiment using sucrose gradients, a classical method to separate mRNAs according to the amount of bound ribosomes. Since we examined only the barcoded *Rluc* mRNAs, their relative segregation in the gradient indicates

ribosome density and translation efficiency (TE) estimation. We named the whole procedure barcoded polysomal profiling, or BPP (Figure 1A). Relative enrichment of Pro-Lib barcodes in the various fractions of the gradient showed that most *Rluc* mRNAs are localized in the initial polysomal fractions (i.e., 9-12; Figure 1B). Control total RNA segregation showed a characteristic pattern of polysomes and EDTA-sensitivity, in line with the known dependence of polysomes on the availability of Mg^{2+} ions (Figures S1C and S1D).

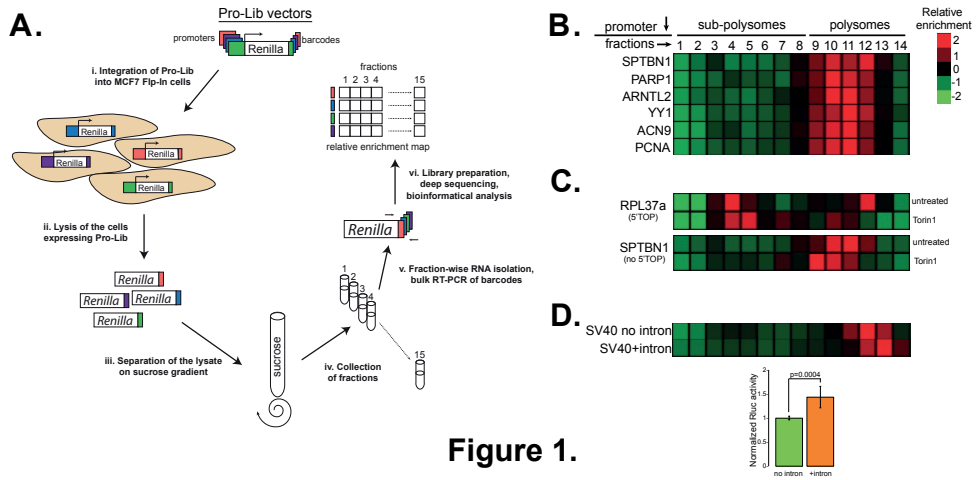


Figure 1. A screen for examination of relationship between transcription and translation.

(A) Schematics of the barcoded polysomal profiling (BPP) approach. (B) A typical segregation of Rluc mRNAs. (C) Control Rluc transcript with 5'TOP sequence exhibits rapid shift to non-translating fractions upon inhibition of mTORC1. (D) Control Rluc transcript shifts to denser fractions following splicing. The bar diagram below represents normalized relative Rluc protein expression assessed by luciferase assay in two separate clones. See also Figure S1.

Next, we tested whether BPP can detect changes in TE. For this purpose, we employed *Rluc* mRNA containing 5'-terminal oligopyrimidine tracts (5'TOP) derived from RPL37a gene. Translation of mRNAs possessing 5'TOP is highly dependent on mTOR activity, resulting in a rapid translational arrest following mTOR inhibition, compared to other mRNAs (Thoreen et al., 2012). Indeed, inhibition of mTORC1 resulted in a global moderate shift of the *Rluc* mRNAs to fractions

9-11 (e.g., SPTBN1 promoter), while a 5'TOP-containing transcript was depleted from the translated fractions of the gradient (Figure 1C). Interestingly, we also identified several additional *Rluc* transcripts exhibiting similar hypersensitivity towards the inhibition of mTORC1 (Figure S1E). Further analysis of the sequences adjacent to their TSSs (Figure S1F) identified stretches of pyrimidines that could serve as 5'TOP signals, providing a probable explanation for their dramatic response. To further test detection capabilities of BPP, we examined the effect of splicing on TE. Indeed, we observed that spliced *Rluc* mRNA shifts to heavier polysomal fractions and yields more protein (Figure 1D), as expected from the known positive effect of splicing on translation (Nott et al., 2004). Altogether, our control experiments demonstrate that BPP is capable of examining multiple *Rluc* mRNAs in bulk and detecting changes in TE of individual transcripts.

TATA box confers higher translational efficacy

Inspecting the segregation of the barcoded *Rluc* mRNAs on sucrose gradients, we identified 12 promoters that caused a shift of the reporter transcript toward higher ribosomal occupancy fractions in at least two independent experiments. Intriguingly, four of these promoters contained a TATA-box element (Figure 2A) and four others had TA-rich sequences that could potentially serve as non-canonical TATA boxes (Figure S2A), indicating that this promoter element could positively influence translation. To test this possibility, we supplied several TATA-less promoters with an artificial TATA element (consisting of the TATA sequence followed by the short downstream sequence derived from human ACTB gene). In all cases, this manipulation resulted in *Rluc* transcripts occupying denser fractions of the gradient (i.e., 11-13; Figures 2B and S2B), supporting the previous observations. Taking the ASNSD1 promoter as a model, we observed a positive effect of TATA addition on TE under various conditions (Figure S2C), thus indicating a robust phenomenon.

Since artificial introduction of TATA may alter the TSS, possibly impacting translational capacity (Rojas-Duran and Gilbert, 2012), we investigated in details the 5'UTRs produced from the TATA-containing and TATA-less promoter pairs. By northern blotting we observed that most of the tested promoters resulted in reporter transcripts of a similar length, with a noticeable enhancement of mRNA levels in TATA-containing promoters (Figure 2C). To establish precisely the 5'-ends of the transcripts, we performed 5'RACE analyses, which showed a very narrow peak ~25nt downstream the inserted TATA element (Figure S3A,B). To test whether

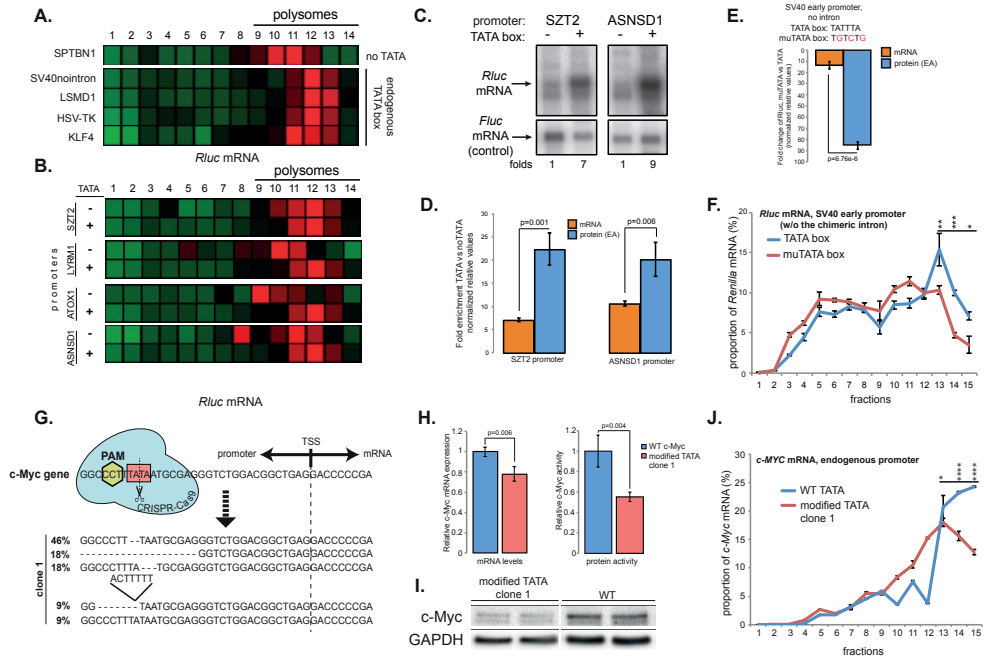


Figure 2. Presence of the TATA element in promoters enhances TE.

(A) Pro-Lib vectors encoding for TATA-containing promoters result in *Rluc* mRNAs shifted to the denser fractions of the gradient compared to other transcripts (e.g., under SPTBN1 promoter). (B) Pro-Lib vectors supplied with an artificial TATA yield *Rluc* mRNAs that are shifted to the denser fractions of the gradient compared to the parental TATA-less promoters. (C) Northern blot analysis of *Rluc* mRNAs; quantification reflects relative *Rluc*/*Fluc* ratio. (D) Relative levels of Rluc mRNAs and proteins produced by promoters with or without TATA element; note the super-induction of the protein expression. Data are represented as mean±SEM, n=4. (E) Relative levels of Rluc mRNAs and proteins produced by SV40 promoter upon mutagenesis of TATA; data are represented as mean±SEM, n=3. (F) Mutagenesis of SV40 promoter-derived TATA results in less efficient translation of *Rluc* mRNA as detected by BPP; data are represented as mean±SEM of a representative gradient, n=2. (G) Schematics of the CRISPR-Cas9-mediated mutagenesis of the endogenous TATA (in pink square) of c-Myc gene. Below: characterization of the different c-Myc alleles of one of the isolated clones, including relative abundances. (H) Mutagenesis of the c-Myc TATA results in lowered mRNA levels (left chart) and further reduced c-Myc activity (right chart); data are represented as mean±SEM of n=3. (I) Western blot analysis of c-Myc protein in the TATA-mutated clone and wt MCF7 cells; see Figure S2I for the uncropped blot. (J) Polysomal profilings of *c-Myc* mRNAs isolated from the clone with mutated TATA and wt MCF7 cells; data are represented as mean±SEM of a characteristic gradient; n=3. See also Figure S2G-K for the characterization of an additional clone.

these alterations in the 5'UTRs could explain the observed changes in translation, we *in vitro* synthesized *Rluc* transcripts bearing the different 5'UTRs and examined their relative TE (Figure S3C). As the 5'UTR generated after the insertion of TATA sequence did not confer higher TE, we conclude that the observed positive effect

of TATA on translation is unlikely to stem from the differences in the 5'UTRs of *Rluc* mRNAs.

So far we inferred TE from ribosome occupancy measurements reflected by migration within sucrose gradients. Next, we tested TE changes by measuring separately the levels of the reporter mRNA and protein in the promoter pairs. As expected, we observed significantly more mRNA produced from TATA-containing promoters (7-9 folds, Figure 2D), consistent with the levels measured by northern blotting (Figure 2C). In contrast, the increase in the level of protein activity was significantly higher (>20 folds), supporting the connection between the presence of TATA in promoter and enhanced protein production. To further test the role of the TATA element in translation, we mutagenized it within the SV40 promoter. Remarkably, loss of TATA reduced the reporter mRNA expression by ~14-fold and protein activity by ~80-folds (Figure 2E), suggesting reduced TE. Indeed, mRNA produced by a TATA-mutated promoter was enriched in the lighter polysomal fractions (Figure 2F), further supporting this indication. We observed similar results upon mutagenesis of TATA in other promoters (Figure S2D,E), suggesting that this effect is not restricted to any particular promoter.

Next we tested whether the positive effect of TATA on TE applies also to the endogenous mammalian gene expression. For this purpose, we chose c-Myc, a ubiquitously expressed gene with a single active TATA-positive promoter in MCF7 cells (Figure S2F) and used CRISPR-Cas9 system to alter its TATA sequence (Figure 2G). This strategy yielded isolated cell clones with disrupted c-Myc TATA on most of their alleles (Figure 2G and Figure S2G). Examination of c-Myc expression in these clones revealed a reduction of ~25% in *c-Myc* mRNA and ~50% in both protein activity and expression (Figure 2H,I and S2I,J). Moreover, polysomal profiling of *c-Myc* transcripts showed reduced ribosomal occupancy upon mutating the TATA element (Figure 2J and S2H), while a control transcript displayed very similar profiles in both clones (Figure S2K), indicating a c-Myc–specific effect. Thus, we conclude that the presence of the TATA element in a promoter enhances the efficiency of mRNA translation and note the validity of this observation to multiple promoters and genes.

General association, but not causal link, between mRNA levels and translational efficiency

In all the cases we studied, presence of TATA in a promoter stimulated both mRNA expression levels and TE. To examine if TE positively correlates with mRNA

expression levels, we clustered the Pro-Lib BPP data to separate the transcripts of our library into two relative groups: one with lower TE (Figure S4A,B) and another with higher TE (Figure S4C,D), and compared the expression levels between these two groups. Indeed, we found that transcripts with higher TE tend to be more abundant (Figure 3A). To test whether this coupling is TATA-dependent, we repeated this analysis while omitting the promoters with artificially added TATA elements. Notably, this analysis yielded similar albeit less significant results (Figure 3B), thus suggesting a positive correlation between the abundance of *Rluc* transcripts and their TE, with no dependency on TATA. To further test this observation, we employed cells expressing an inducible version of the barcoded reporter gene (TRex-Rluc), in which the levels of *Rluc* mRNA are stimulated ~17-fold after induction (Figure 3C). Importantly, induction of this gene yielded a significantly greater enhancement of protein activity (~60-fold, Figure 3C), suggesting translation boost. Indeed, subsequent BPP analysis revealed a strong increase in the ribosome occupancy of the induced reporter mRNA (Figure 3D and Figure S5A), supporting the notion that increased expression results in higher TE. Notably, we performed 5'RACE analysis of both induced and non-induced TRex-Rluc mRNAs and found the scattering of TSSs to be very similar (Figure S3D). Altogether, we conclude that the observed link between the expression levels of mRNAs and their TE is not restricted to the TATA element and might therefore represent a more general phenomenon.

To test if genome-wide data support global relationship between mRNA level and TE, we analyzed pairs of RNA-seq and Ribo-seq (ribosomal footprinting) datasets from multiple human cell lines. Intriguingly, we observed a positive global correlation between mRNA expression level and TE, which was rather weak but consistent and statistically significant (Figure 3E). Moreover, this correlation between mRNA levels and TE was also apparent in various stress conditions (Figure S5B,C). These results indicate that also in mammalian genomes, there is a positive correlation between expression levels of mRNAs and their TE.

To examine if mRNA levels could directly regulate TE, we transfected different amounts of *in-vitro* transcribed and purified *Rluc* mRNA into MCF7 cells and monitored the activity of the produced Rluc protein. We anticipated that if levels of mRNA positively stimulated TE, transfecting increasing amounts of

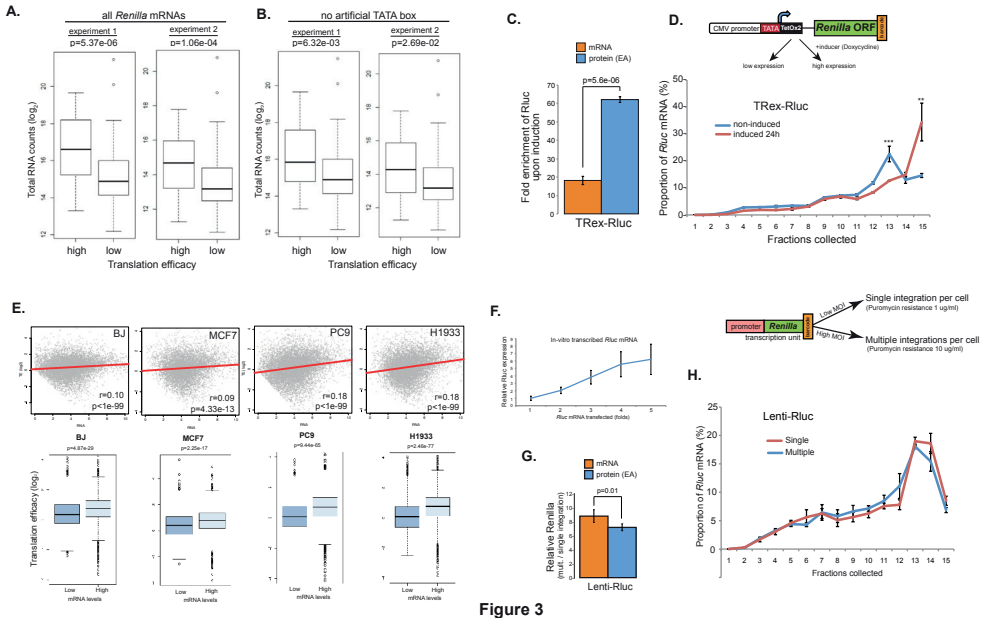


Figure 3

Figure 3. Levels of mRNAs positively correlate with TE but do not dictate it.

(A) Reporter mRNAs from the Pro-Lib screen were separated into groups with relatively high or low TE (derived from the polysomal profiles, see Figure S4) and compared with their expression levels (estimated from read counts observed for each vector over all fractions). p-values calculated using Wilcoxon's test. (B) Same comparison as described in (A) was performed after exclusion of mRNAs transcribed from promoters with artificial TATA element. (C) Levels of mRNA and protein resulting from the induced TRex-Rluc gene were measured and plotted relatively to the non-induced condition; data are represented as mean±SEM of n=3. (D) BPP examination of either induced or non-induced TRex-Rluc mRNAs; data are represented as mean±SEM of a characteristic gradient; n>5, see also Figure S5A. (E) Upper panels: paired RNA-seq and Ribo-seq datasets from different human cell lines were examined for relationship between mRNA expression level and TE (calculated as the (log_2) ratio between densities of ribosome footprint and RNA-seq reads). Lower panels: comparisons between the 10% of genes with lowest and highest expression levels are presented; p-values calculated using Wilcoxon's test. (F) MCF7 cells were transfected with fold-wise amounts of *in vitro* transcribed (using HeLa nuclear extract) *Rluc* mRNA followed by measurement of RLuc activity after 18 hours; data are represented as mean±SEM of n=3. (G) Levels of Rluc mRNA and protein were compared between two populations of MCF7 cells expressing near single or multiple integrated copies of Lenti-Rluc unit. Data are represented as mean±SEM of n=3. (H) BPP analysis of *Rluc* transcripts described in (G); data are represented as mean±SEM of a characteristic gradient, n=3.

mRNA would result in an exponential increase of protein production. However, we observed a clear linear dependency between the two parameters (Figure 3F), suggesting indifference of the protein production rates to mRNA abundance. To

test this conclusion further, we employed lentiviral-mediated stable integration of a transcriptional unit including promoter and Rluc gene and generated two stable populations with either nearly single or multiple integrations of the same transcriptional unit per cell, resulting in ~9-fold difference in *Rluc* mRNA levels (Figure 3G). Also here, increase in protein activity was similar to the increase in mRNA levels, indicating comparable TEs. Supporting this conclusion, *Rluc* mRNAs derived from both cell populations exhibited similar segregation patterns in sucrose gradients (Figure 3H). Altogether, these results suggest that while mRNA expression levels positively correlate with TE in multiple cellular models, mRNA abundance does not directly regulate it.

Rate of transcription positively affects translation efficiency

Next we searched for the causal origin of the observed correlation between mRNA levels and TE. We considered the rate of transcription to be the most likely candidate since it directly regulates the expression levels of mRNAs. To test this possibility, we correlated transcription rates (as estimated by Genome Run-On sequencing, GRO-seq, (Core et al., 2008)) and TE (as determined by Ribo-seq and RNA-seq data) in BJ and MCF7 cell lines. In both cell types we observed a significant positive association between the rates of transcription and TE of genes that do not possess upstream ORFs (Figure 4A and Figure S6A,B). This correlation is notable as this analysis integrates distinct datasets independently obtained from three genomic techniques in two different cellular systems.

To further investigate the relationship between rates of transcription and TE, we employed Camptothecin (CPT), a chemical compound that inhibits topoisomerase I and, in mild concentrations, slows down the progression of transcribing RNA polymerase II (RNAPII) (Dujardin et al., 2014). Indeed, when CPT was applied in parallel to the induction of TRex-Rluc gene, we observed a reduction of ~60% of *Rluc* mRNA, but a more prominent reduction of Rluc activity of ~80% (Figure 4B). This difference could be explained by the reduced ribosome occupancy, as reflected by polysomal profiling analyses (Figure 4C and Figure S5D). Thus, impediment of RNAPII progression does not only reduce the transcriptional rate of TRex-Rluc gene but also attenuates its TE. To further test the role of RNAPII dynamics on translation, we assessed TE of multiple genes in CPT-treated cells. Overall, CPT treatment resulted in TE reduction of 691 genes by more than 1.5-fold in two independent experiments (Table S2). As highly transcribed genes showed higher TE (Figure 4A), we speculated that CPT treatment would cause more pronounced

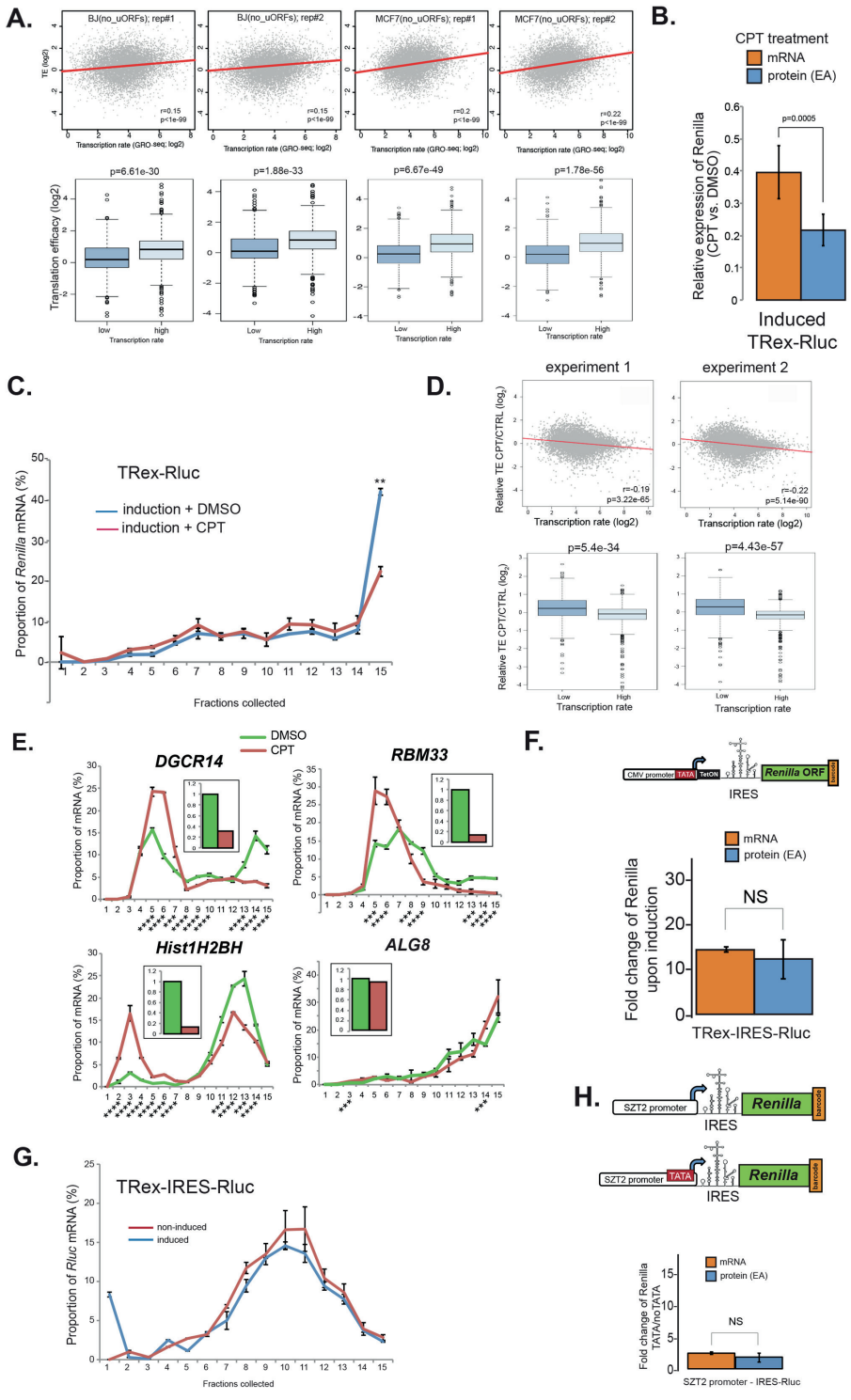


Figure 4. TE is affected by the rate of transcription.

(A) Upper panels: positive genome-wide correlations between translational efficacies and rates of transcription in BJ and MCF7 cells. Lower panels: direct comparisons between 10% of genes with lowest and highest transcription rates. GRO-seq data were from (Korkmaz et al., 2016, Leveille et al., 2015); Ribo-seq and RNA-seq data were from (Loayza-Puch et al., 2013); p-values were calculated using Wilcoxon's test. **B.** Expression levels of Rluc mRNA and protein were examined following parallel induction of the TRex-Rluc gene and CPT treatment for 7 hours; data are represented as mean±SEM of n=3. **C.** Lysates of the two populations described in (B) were subjected to BPP procedure; data are represented as mean±SEM of a characteristic gradient, see also Figure S5D; n=3. **D.** Upper panels: genome-wide correlations between rates of transcription and TE changes after treatment with CPT. Lower panels: direct comparison of the effect on TE between the 10% of most highly and lowly transcribed genes; p-values were calculated using Wilcoxon's test. **E.** Polysomal profilings of mRNAs after CPT treatment (red) show reduced TE compared to untreated cells (green). Columns represent the relative mRNA levels as detected by qRT-PCR; ALG8 was used as a control gene. Data are presented as mean±SEM of three technical measurements of a characteristic graident; n=3. See also Figure S6C. **F.** Cells with barcoded TRex-IRES-Rluc cassette (schematics) were induced for 18 hours and subjected for examination of *Rluc* mRNA and protein levels relative to the non-induced cells; data are presented as mean±SEM of n=3. **G.** Same cells as in (F) were treated in a similar way and subjected to BPP procedure. Data are presented as mean±SEM of three measurements of a characteristic gradient; n=3. **H.** Pro-Lib vectors with SZT2 promoters, with or without the TATA element, were supplemented with IRES-encoding sequence as shown on the schematics. Cells expressing these constructs were subjected to quantification of mRNA and protein levels; data are presented as mean±SEM of n=3.

repressive effect on the translation of these genes. Indeed, intersection with GRO-seq data demonstrated that impediment of RNAPII caused a significant reduction in TE of genes with relatively high transcription rate (Figure 4D). We further examined the polysomal segregation of several mRNAs, suggested by the genome-wide experiments, and validated that these mRNAs indeed displayed reduced TE coupled with reduced expression, compared to *ALG8* control mRNA (Figure 4E). To eliminate the possibility that the transcripts exhibiting translational shifts are truncated due to the CPT treatment, we re-examined these and other mRNAs using primers annealing to 3'UTRs only. Also in this case, the shift of mRNAs was apparent (Figure S6C). Taken together, we conclude that transcription rate is an important positive determinant of translation efficiency.

Transcription-dependent translation regulation requires canonical translation initiation

To further examine the mechanism by which transcription affects TE, we introduced an internal ribosome entry site (IRES) sequence into the inducible

TRex-Rluc cassette (TRex-IRES-Rluc gene, see schematics in Figure 4F) to bypass the canonical 5'cap-mediated initiation of translation. Interestingly, induction of TRex-IRES-Rluc gene resulted in Rluc activity that mirrored the changes in the mRNA levels (i.e.,~12-15 fold induction at both levels, Figure 4F), which contrasted the previously observed induction of the Rluc protein activity (>60-fold, Figure 3C). Moreover, BPP analyses revealed similar segregation of both induced and non-induced *IRES-Rluc* mRNAs (Figure 4G), further suggesting lack of translational induction. Furthermore, introduction of IRES into constructs containing the artificial TATA element (e.g., SZT2 promoter) abolished the previously observed enhancement of TE (Figure 4H, compare with Figure 2D) and resembled the effect of TATA mutagenesis (Figure S2D). Thus, we conclude that canonical translation initiation is required for transcription-dependent enhancement of translation.

m^{6}A modification mediates transcription-responsive translation

Next we looked for a molecular event that could explain the link between transcription and translation. Since changes in transcription affected translation of multiple genes (Table S2), we reasoned that this molecular event should be robust and widespread. N6-methyladenosine (m^{6}A) is an abundant nuclear dynamic mRNA modification identified in thousands of human transcripts (Dominissini et al., 2012, Meyer et al., 2012), recently implicated in translational regulation (Wang et al., 2015). Therefore, we asked whether m^{6}A could bridge transcription and translation.

First, we examined whether m^{6}A levels could be affected by different rates of transcription. We therefore performed immuno-precipitations of m^{6}A-modified RNAs (MeRIP) from cells with either induced or non-induced TRex-Rluc gene, and compared the relative recoveries of the different *Rluc* transcripts. Strikingly, we observed very significant differences between the two *Rluc* populations; the non-induced mRNA was isolated much more efficiently in all tested mRNA regions (Figure 5A). Importantly, a control mRNA (AHNAK) showed similar recovery in both samples, suggesting a specific enrichment of the non-induced *Rluc* transcripts by MeRIP. We also assessed the methylation levels of *Rluc* mRNAs of the two populations expressing different copy number of Lenti-Rluc gene (Figure 3G) and found them to be very similar (Figure S6D), suggesting that the differences observed in MeRIP are unlikely to stem from unequal mRNA expression levels. Furthermore, we observed differences in the post-MeRIP recovery of *Rluc* mRNAs transcribed from either TATA-containing or TATA-less promoters (Figure S6E), suggesting that promoters with intact TATA elements yield mRNAs with relatively lower m^{6}A

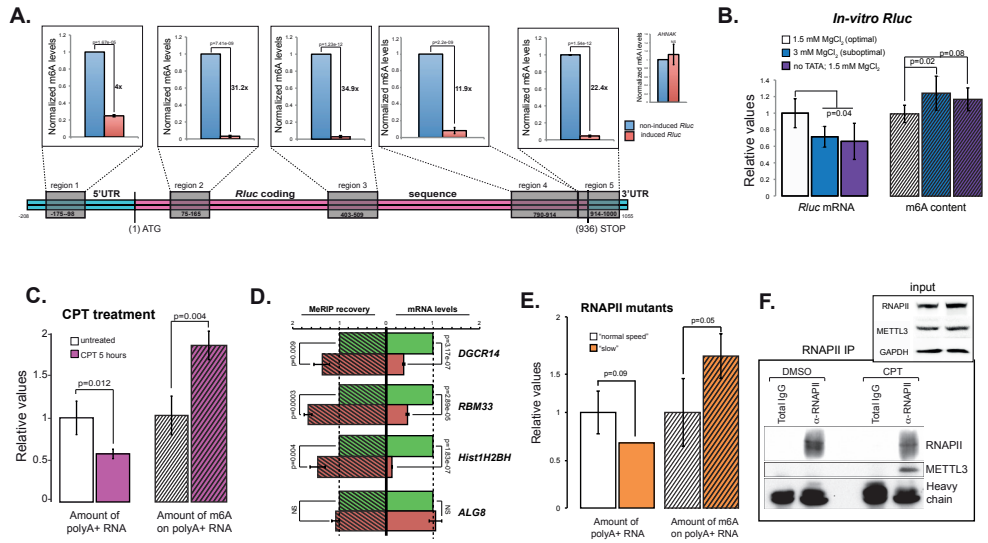


Figure 5. Enhancement of m^{6}A modification of mRNAs upon attenuated transcription.

(A) Cells bearing the TRex-Rluc gene were induced for 18 hours and subjected to the MeRIP procedure, together with untreated cells. Recovery efficiencies of regions spanning the *Rluc* mRNA were examined, normalized to the input levels and compared; *AHNAK* was used as a control methylated mRNA. Data are represented as mean±SEM of n=3. (B) *Rluc* mRNAs were transcribed *in vitro* using HeLa extract as detailed, resolved on agarose gels and quantified (left chart, n=3) and further subjected to the relative comparison of their m^{6}A contents (right chart, n=2). (C) Total RNA isolated from MCF7 cells treated with CPT for 5 hours was enriched for polyA$^+$ population and quantified (left chart, n=3) and further subjected to the relative comparison of their m^{6}A contents (right chart, n=3). (D) mRNAs presented in Figure 4E were tested for recovery after MeRIP procedure (striped columns) relatively to their RNA levels (plain columns) using qRT-PCR. Data are represented as mean±SEM of n=2. See also Figure S6D,E. (E) Relative levels of polyA+ RNA populations (plain columns) in cells expressing different RNAPII mutants, and their relative m^{6}A levels (striped columns) were examined; n=2. (F) MCF7 cells were treated for 5 hours and subjected to immuno-precipitation of RNAPII, followed by detection of METTL3 protein using Western blot. A characteristic blot is presented, n=4.

content. Next, we *in-vitro* transcribed Rluc gene in m^{6}A-compatible nuclear HeLa extract (Shimba et al., 1995) using conditions that reduce rate of transcription (e.g., elevated $MgCl_2$ (Wildeman et al., 1984), or a template with mutated TATA box), and tested the different transcripts for m^{6}A content. Indeed, both changes resulted in the relatively lower levels of *Rluc* mRNA, which were associated with higher deposition of m^{6}A (Figure 5B). Importantly, both the UTRs and polyA tails of the different transcripts were similar (Figure S3E,F). Furthermore, treatment with CPT reduced mRNA levels in parallel to the relative enhancement of their m^{6}A content (Figure 5C), suggesting a negative correlation between transcription elongation

and m⁶A deposition. Additionally, we performed MeRIP analysis of genes that exhibited reduction of both mRNA levels and TE after CPT treatment (Figure 4E). These genes indeed displayed lower levels of mRNA coupled with enhanced m⁶A content (Figure 5D). To further address the link between the rate of transcription and m⁶A deposition in an additional independent system, we interrogated α-amanitin resistant RNAPII mutants exhibiting either normal (N792D) or slow (C4/R749H) rates of elongation (Fong et al., 2014). Similarly to the previous observations, we found that cells expressing the "slow" RNAPII mutant contained relatively lower levels of mRNA, associated with elevated methylation content (Figure 5E). Thus, various experimental approaches indicate that suboptimal rate of transcription results in higher m⁶A deposition on mRNAs.

This conclusion raised the possibility that m⁶A modification of mRNAs occurs co-transcriptionally and therefore we anticipated a physical interaction between RNAPII and the methyltransferase complex (MTC). To test this, we immuno-precipitated RNAPII from either normally growing cells or following CPT treatment and probed the eluate for METTL3, a catalytic core of MTC (Wang et al., 2016). Remarkably, while in untreated conditions we observed no visible interaction, upon CPT treatment, METTL3 was efficiently co-precipitated (Figure 5F), suggesting co-transcriptional m⁶A modification and enhancement of the interaction between RNAPII and MTC upon impediment of RNAPII elongation dynamics.

Next, we tested whether elevated levels of m⁶A could negatively affect translation. We first examined the translation capacity of the *in vitro* transcribed *Rluc* mRNAs in the conditions of reduced transcription and enhanced m⁶A (as presented in Figure 5B). In line with our hypothesis, conditions associated with attenuated transcription resulted in lower efficacy of protein production compared with the normal transcription conditions (Figure 6A). Next, we examined the global effects of CPT treatment and expression of the "slow" RNAPII mutant on polysomes. Figures 6B,C show that these two treatments considerably reduced the RNA content of the polysomal fractions, suggesting a general reduction of translation and providing further evidence for the positive feedback between transcription and translation.

To directly examine the role of m⁶A in translation, we knocked-down two genes in the m⁶A pathway (METTL14 and YTFDH1, Figure S6F) and examined the effect on the expression of *Rluc* mRNA. Interestingly, knocking down either factor resulted in a significant boost in the expression of the non-induced Rluc, while the effect on the induced gene was moderate (Figure 6D). Importantly, while global polysome segregation analysis indicated only a slight effect of each knockdown on

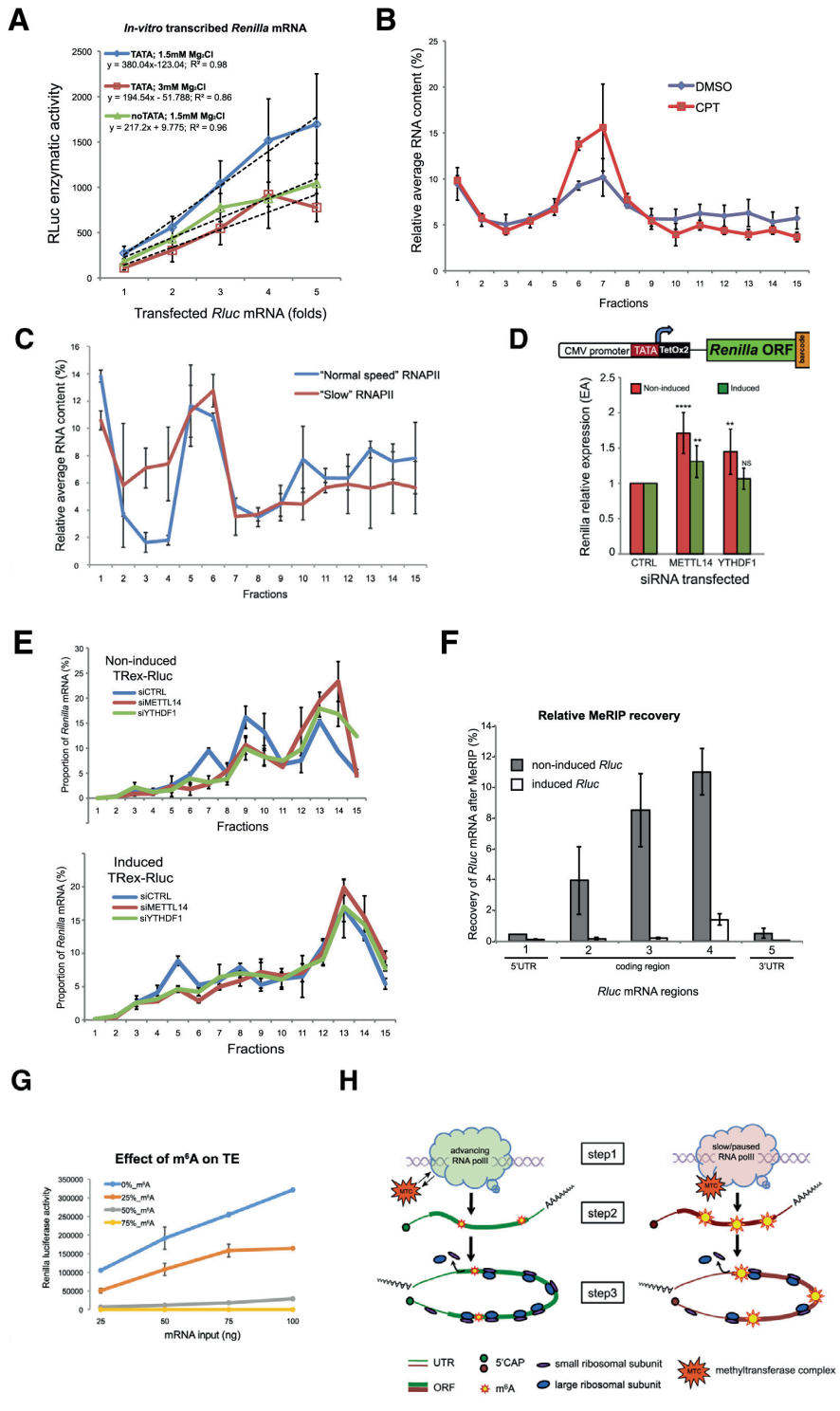


Figure 6. Effect of the m6A modifications on translation.

(A) Rluc mRNAs were transcribed invitro using HeLa extract in the indicated conditions, and transfected into MCF7 cells in fold-wise amounts; RLuc enzymatic activity was measured after 24 hours. The trend lines are represented by the dashed lines; n=3. (B) MCF7 cells were treated with CPT for 5 hours and subjected to polysomal profiling with subsequent measurement of the average levels of total RNA in the collected fractions. Data are presented as mean±SEM, n=3. (C) Polysomal segregation of total RNA from cells expressing "normal speed" or "slow" RNAPII; data are presented as mean±SEM of n=3. See also Figure S6H. (D) TRex-Rluc gene was induced for 18 hours in MCF7 cells transfected with the detailed siRNAs, or left uninduced. Expression levels of RLuc protein were assessed; data are represented as mean±SEM of n=3. (E) Cells bearing the TRex-Rluc gene were transfected with the indicated siRNAs, grown for 56 hours and subjected to BPP. Upper panel: non-induced Rluc; lower panel: Rluc expression was induced for 24 hours prior to harvesting; data are represented as mean±SEM of three measurements of a characteristic gradient; n=3. (F) Results of the MeRIP experiment presented in Figure 5A were normalized to calculate the total relative recovery of Rluc mRNAs by the different regions. Error bars represent mean±SEM of n=3. (G) Rluc mRNAs were in vitro transcribed using T7 polymerase in the presence of the indicated proportions of m6A nucleoside and in vitro translated using indicated amounts of mRNA as input. Data are represented as mean±SEM of n=4. (H) Model of the transcription-dependent regulation of TE. Slow or paused RNAPII results in the enhanced interaction with MTC (step 1), which leads to enhanced deposition of m6A on mRNAs (step 2), negatively affecting translation efficiency (step 3).

general translation (Figure S6G), possibly due to compensatory mechanisms, BPP assay verified a considerable increase in the TE of the non-induced *Rluc* mRNA in these conditions (Figure 6E). Since methylation of coding regions (CDRs) has been recently suggested to attenuate translation (Choi et al., 2016), we re-examined the experiment presented in Figure 5A, in order to compare the relative changes in methylation of the different *Rluc* transcript regions. Remarkably, the highest gain of methylation upon transcriptional attenuation was observed throughout the CDR of *Rluc* (up to ~11% near the end of the non-induced mRNA CDS), whereas both UTRs (i.e., 5' and 3') displayed the lowest levels (<1%, Figure 6F). This result therefore suggests a plausible explanation for the strong inhibitory effect of transcriptional repression on TE that we observed (Figure 3D). Lastly, we sought to directly examine the effect of m^6A levels on protein production, and produced *Rluc* mRNAs *in vitro* with different proportions of methylated adenosines. Strikingly, increased incorporation of m^6A led to a progressive attenuation of TE of the *Rluc* (Figure 6H), suggesting an overall inhibitory effect of m^6A modification on translation. Altogether, these results demonstrate that covalent m^6A modification of mRNAs depends on the dynamics of the transcribing RNAPII, and tends to negatively affect translation efficacy. Therefore, we conclude that m^6A links transcription and translation.

Discussion

In this study, we describe a direct and general flow of information from transcription to translation in human cells. We show that genes possessing strong transcriptional activity give rise to mRNAs with greater capacity to produce proteins. We propose that this link is mediated, at least in part, by the co-transcriptional N6-methylation of adenosines in mRNA - the m^6A modification. Suboptimal transcription results in a relatively higher m^6A content of transcripts and this, in turn, tend to reduce translation efficacy (Figure 6H).

Initially, we employed Pro-Lib, a library consisting of multiple human promoters that drive the transcription of a single reporter Rluc gene, and combined it with barcoded polysomal profiling procedure to examine translation efficacy (TE) of the reporter mRNAs. This approach allowed for several advantages. First, recombination-mediated single-copy genomic integration resulted in a balanced expression of the reporter constructs. Second, all vectors were integrated into the same genetic locus, minimizing the probability of an artificial effect of chromatin neighborhood on transcription. Third, establishing a pooled stable reporter population simplified the maintenance and treatments and increased experimental accuracy. Lastly, deep sequencing of the barcodes enabled a quantitative coverage of nearly all reporter transcripts in the polysomal fractions within a single experiment.

Using Pro-Lib, we observed that promoters possessing either confirmed or suspected TATA-box element (TATA) yield transcripts characterized by a higher capacity to accommodate ribosomes and produce proteins. This indicated that the presence of this element in a promoter not only boosts transcription, but also enhances mRNA translation, suggesting a positive effect of TATA box promoter element on TE. TATA elements activate transcription (Xiao et al., 1995), and a recent genome-wide bioinformatics analysis suggested that mammalian genes possessing TATA in their promoter tend to be translated more efficiently (Tamarkin-Ben-Harush et al., 2014), which was attributed to their characteristics (e.g., gene length, probability of upstream ORF, etc.). In this study, we support these observations and conclude, in addition, that a mere presence of intact TATA element in a promoter potentiates mRNA translation (Figures 2B, S2A).

The TATA element activates transcription (Xiao et al., 1995), resulting in elevated levels of mRNA. Indeed, we identified a strong correlation between mRNA levels and translation in the Pro-Lib screen and further showed that this link is not unique to TATA, holding true for the global mammalian gene expression. We note that

the correlations between mRNA levels and TE are rather weak, probably reflecting multiple additional levels of regulation, but statistically significant. Interestingly, this correlation was noticed in both animal cells and yeast (Schwanhausser et al., 2011, Weinberg et al., 2016), likely indicating an evolutionarily conserved phenomenon. Importantly, while mRNA levels could be linked to TE via mechanisms that couple mRNA stability and translation (Radhakrishnan et al., 2016), we found that mRNA levels *per se* do not directly affect translation (Figures 3F-H). Instead, we observed a global positive effect of transcription on TE. Impediment of transcription either via CPT treatment or in the *in vitro* assays, demonstrated a role for transcriptional elongation in the determination of translational efficacy. As transcription initiation and elongation are related processes (Marbach-Bar et al., 2013), future studies are necessary in order to precisely appreciate the role of transcription initiation in the regulation of translation. Regarding mRNA translation, our observations suggest its responsiveness to transcriptional changes only when the initiation of translation involves 5'cap. When this canonical initiation was bypassed by the introduction of IRES, the translation machinery failed to respond to transcriptional fluctuations (Figures 4F-H), indicating that this initial step, which is considered to be rate limiting and highly controlled, might exert additional aspects of translational regulation.

Here we suggest that m^{6}A, at least partially, mediates the communication between transcription and translation. Our results indicate that attenuated transcription leads to enhanced methylation of mRNAs. We present here biochemical evidence for a physical interaction between METTL3 enzymatic subunit of the methylation complex (MTC) and retarded RNAPII (Figure 5F). These results suggest that m^{6}A modification is, at least in part, co-transcriptional and indicate that the affinity between MTC and RNAPII depends upon the dynamics of the latter. While it remains unclear how exactly this interaction is induced, we propose that slow progression or frequent pausing of RNAPII increase the probability of MTC engagement. The global enrichment of modified nucleotides at the start of mRNAs and around STOP codons (Meyer et al., 2012, Dominissini et al., 2012) supports this hypothesis, as these are the locations of probable RNAPII pausing (Jonkers and Lis, 2015). Since a considerable proportion of genes experience slow or paused transcription (Core et al., 2008), it would be important to examine, in future studies, whether these genes are preferentially methylated. We note that together with m^{6}A, additional factors might bridge transcription and translation, e.g., (Rojas-Duran and Gilbert, 2012, Tamarkin-Ben-Harush et al., 2017).

Globally, we observed that enhanced mRNA methylation results in lower translation capacity. Interestingly, several recent studies indeed linked m^{6}A to

translation, but reported mainly a stimulatory role of this modification (Meyer et al., 2015, Wang et al., 2015, Zhou et al., 2015). We anticipate that the precise effect of m^{6}A on translation depends on the location of the modified nucleotide within the transcript. Remarkably, most of the mentioned studies observed enhanced methylation within UTRs of transcripts. In contrast, methylation within CDRs reduced translation (Qi et al., 2016), probably due to the attenuated elongation (Hoernes et al., 2016, Choi et al., 2016). Given that most of the methylated residues are located within CDRs, we would anticipate an overall negative effect of m^{6}A methylation on translation, which we indeed observed upon random incorporation of modified adenosines (Figure 6G).

Our study suggests that m^{6}A modification of mRNAs is co-transcriptional and is used to epigenetically imprint mRNAs and control their future activity and fate. This way the dynamics of the transcription process could be "recorded" in the methylation profile of a given mRNA, which, in turn, could further impact multiple steps of mRNA biology.

Contributions

B.S. and R.A. planned and designed the experiments, analyzed the results and wrote the manuscript; B.S., R.H. and J.A.F.O.V. cloned Pro-Lib; B.S., R.H. F.L. and V.C. performed the experiments; R.E. performed all bioinformatical tests. All authors read and approved the manuscript.

Acknowledgements

We thank Pieter van Breugel, Alejandro Piñeiro Ugalde, Behzad Moumbeini and Nicolas Léveillé for the technical assistance and fruitful discussions and greatly appreciate the help of Ron Kerkhoven and the NKI genomics core facility in sequencing. We thank David Bentley for kindly providing us the cell lines expressing α-amanitin resistant versions of RNAPII. R.E. is a Faculty Fellow of the Edmond J. Safra Center for Bioinformatics at Tel Aviv University. This work was supported by funds from the enhReg ERC-AdV and NWO (NGI 93512001/2012) programs to R.A.

Conflict of interests

The authors declare no conflict of interests.

References

Byrne, B. J., Davis, M. S., Yamaguchi, J., Bergsma, D. J. & Subramanian, K. N. (1983). Definition of the simian virus 40 early promoter region and demonstration of a host range bias in the enhancement effect of the simian virus 40 72-base-pair repeat. Proc Natl Acad Sci U S A, 80, 721-5.

Choi, J., Ieong, K. W., Demirci, H., Chen, J., Petrov, A., Prabhakar, A., O'leary, S. E., Dominissini, D., Rechavi, G., Soltis, S. M., et al. (2016). N(6)-methyladenosine in mRNA disrupts tRNA selection and translation-elongation dynamics. *Nat Struct Mol Biol,* 23**,** 110-5.

Core, L. J., Waterfall, J. J. & Lis, J. T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. *Science,* 322**,** 1845-8.

Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M., Ungar, L., Osenberg, S., Cesarkas, K., Jacob-Hirsch, J., Amariglio, N., Kupiec, M., et al. (2012). Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. *Nature,* 485**,** 201-6.

Dori-Bachash, M., Shalem, O., Manor, Y. S., Pilpel, Y. & Tirosh, I. (2012). Widespread promoter-mediated coordination of transcription and mRNA degradation. *Genome Biol,* 13**,** R114.

Dujardin, G., Lafaille, C., De La Mata, M., Marasco, L. E., Munoz, M. J., Le Jossic-Corcos, C., Corcos, L. & Kornblihtt, A. R. (2014). How slow RNA polymerase II elongation favors alternative exon skipping. *Mol Cell,* 54**,** 683-90.

Elfakess, R. & Dikstein, R. (2008). A translation initiation element specific to mRNAs with very short 5'UTR that also regulates transcription. *PLoS One,* 3**,** e3094.

Fong, N., Kim, H., Zhou, Y., Ji, X., Qiu, J., Saldi, T., Diener, K., Jones, K., Fu, X. D. & Bentley, D. L. (2014). Pre-mRNA splicing is facilitated by an optimal RNA polymerase II elongation rate. *Genes Dev,* 28**,** 2663-76.

Haimovich, G., Choder, M., Singer, R. H. & Trcek, T. (2013). The fate of the messenger is pre-determined: a new model for regulation of gene expression. *Biochim Biophys Acta,* 1829**,** 643-53.

Harel-Sharvit, L., Eldad, N., Haimovich, G., Barkai, O., Duek, L. & Choder, M. (2010). RNA polymerase II subunits link transcription and mRNA decay to translation. Cell, 143, 552-63.

Hoernes, T. P., Clementi, N., Faserl, K., Glasner, H., Breuker, K., Lindner, H., Huttenhofer, A. & Erlacher, M. D. (2016). Nucleotide modifications within bacterial messenger RNAs regulate their translation and are able to rewire the genetic code. *Nucleic Acids Res,* 44**,** 852-62.

Ingolia, N. T., S. Ghaemmaghami, J. R. Newman and J. S. Weissman (2009). "Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling." Science 324(5924): 218-23.

Jonkers, I. & Lis, J. T. (2015). Getting up to speed with transcription elongation by RNA polymerase II. *Nat Rev Mol Cell Biol,* 16**,** 167-77.

Korkmaz, G., Lopes, R., Ugalde, A. P., Nevedomskaya, E., Han, R., Myacheva, K., Zwart, W., Elkon, R. & Agami, R. (2016). Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9. *Nat Biotechnol,* 34**,** 192-8.

Langmead, B., C. Trapnell, M. Pop and S. L. Salzberg (2009). "Ultrafast and memory-efficient alignment of short DNA sequences to the human genome." Genome Biol 10(3): R25.

Leveille, N., Melo, C. A., Rooijers, K., Diaz-Lagares, A., Melo, S. A., Korkmaz, G., Lopes, R., Akbari Moqadam, F., Maia, A. R., Wijchers, P. J., et al. (2015). Genome-wide profiling of p53-regulated enhancer RNAs uncovers a subset of enhancers controlled by a lncRNA. *Nat Commun,* 6**,** 6520.

Loayza-Puch, F., Drost, J., Rooijers, K., Lopes, R., Elkon, R. & Agami, R. (2013). p53 induces transcriptional and translational programs to suppress cell proliferation and growth. *Genome Biol,* 14**,** R32.

Loayza-Puch, F., K. Rooijers, L. C. Buil, J. Zijlstra, J. F. Oude Vrielink, R. Lopes, A. P. Ugalde, P. van Breugel, I. Hofland, J. Wesseling, O. van Tellingen, A. Bex and R. Agami (2016). "Tumour-specific proline vulnerability uncovered by differential ribosome codon reading." Nature 530(7591): 490-4.

Marbach-Bar, N., Ben-Noon, A., Ashkenazi, S., Tamarkin-Ben Harush, A., Avnit-Sagi, T., Walker, M. D. & Dikstein, R. (2013). Disparity between microRNA levels and promoter strength is associated with initiation rate and Pol II pausing. *Nat Commun,* 4**,** 2118.

Meyer, K. D., Patil, D. P., Zhou, J., Zinoviev, A., Skabkin, M. A., Elemento, O., Pestova, T. V., Qian, S. B. & Jaffrey, S. R. (2015). 5' UTR m(6)A Promotes Cap-Independent Translation. *Cell,* 163**,** 999-1010.

Meyer, K. D., Saletore, Y., Zumbo, P., Elemento, O., Mason, C. E. & Jaffrey, S. R. (2012). Comprehensive analysis of mRNA methylation reveals enrichment in 3' UTRs and near stop codons. *Cell,* 149**,** 1635-46.

Nagel, R., C. le Sage, B. Diosdado, M. van der Waal, J. A. Oude Vrielink, A. Bolijn, G. A. Meijer and R. Agami (2008). "Regulation of the adenomatous polyposis coli gene by the miR-135 family in colorectal cancer." Cancer Res 68(14): 5795-802.

Nott, A., Le Hir, H. & Moore, M. J. (2004). Splicing enhances translation in mammalian cells: an additional function of the exon junction complex. *Genes Dev,* 18**,** 210-22.

Oktaba, K., Zhang, W., Lotz, T. S., Jun, D. J., Lemke, S. B., Ng, S. P., Esposito, E., Levine, M. & Hilgers, V. (2015). ELAV links paused Pol II to alternative polyadenylation in the Drosophila nervous system. *Mol Cell,* 57**,** 341-8.

Qi, S. T., Ma, J. Y., Wang, Z. B., Guo, L., Hou, Y. & Sun, Q. Y. (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.

Radhakrishnan, A., Chen, Y. H., Martin, S., Alhusaini, N., Green, R. & Coller, J. (2016). The DEAD-Box Protein Dhh1p Couples mRNA Decay and Translation by Monitoring Codon Optimality. *Cell,* 167**,** 122-132 e9.

Richter, J. D. & Coller, J. (2015). Pausing on Polyribosomes: Make Way for Elongation in Translational Control. *Cell,* 163**,** 292-300.

Rojas-Duran, M. F. & Gilbert, W. V. (2012). Alternative transcription start site selection leads to large differences in translation activity in yeast. *RNA,* 18**,** 2299-305.

Schwanhausser, B., Busse, D., Li, N., Dittmar, G., Schuchhardt, J., Wolf, J., Chen, W. & Selbach, M. (2011). Global quantification of mammalian gene expression control. *Nature,* 473**,** 337-42.

Shimba, S., Bokar, J. A., Rottman, F. & Reddy, R. (1995). Accurate and efficient N-6-adenosine methylation in spliceosomal U6 small nuclear RNA by HeLa cell extract in vitro. *Nucleic Acids Res,* 23**,** 2421-6.

Sonenberg, N. & Hinnebusch, A. G. (2009). Regulation of translation initiation in eukaryotes: mechanisms and biological targets. *Cell,* 136**,** 731-45.

Tamarkin-Ben-Harush, A., Schechtman, E. & Dikstein, R. (2014). Co-occurrence of transcription and translation gene regulatory features underlies coordinated mRNA and protein synthesis. *BMC Genomics,* 15**,** 688.

Tamarkin-Ben-Harush, A., Vasseur, J. J., Debart, F., Ulitsky, I. & Dikstein, R. (2017). Cap-proximal nucleotides via differential eIF4E binding and alternative promoter usage mediate translational response to energy stress. *Elife,* 6.

Thoreen, C. C., Chantranupong, L., Keys, H. R., Wang, T., Gray, N. S. & Sabatini, D. M. (2012). A unifying model for mTORC1-mediated regulation of mRNA translation. *Nature,* 485**,** 109-13.

Ulitsky, I., A. Maron-Katz, S. Shavit, D. Sagir, C. Linhart, R. Elkon, A. Tanay, R. Sharan, Y. Shiloh and R. Shamir (2010). “Expander: from expression microarrays to networks and functions.” Nat Protoc 5(2): 303-22.

Wang, X., Feng, J., Xue, Y., Guan, Z., Zhang, D., Liu, Z., Gong, Z., Wang, Q., Huang, J., Tang, C., et al. (2016). Structural basis of N6-adenosine methylation by the METTL3-METTL14 complex. *Nature,* 534**,** 575-578.

Wang, X., Lu, Z., Gomez, A., Hon, G. C., Yue, Y., Han, D., Fu, Y., Parisien, M., Dai, Q., Jia, G., et al. (2014). N6-methyladenosine-dependent regulation of messenger RNA stability. *Nature,* 505**,** 117-20.

Wang, X., Zhao, B. S., Roundtree, I. A., Lu, Z., Han, D., Ma, H., Weng, X., Chen, K., Shi, H. & He, C. (2015). N(6)-methyladenosine Modulates Messenger RNA Translation Efficiency. *Cell,* 161**,** 1388-99.

Weinberg, D. E., Shah, P., Eichhorn, S. W., Hussmann, J. A., Plotkin, J. B. & Bartel, D. P. (2016). Improved Ribosome-Footprint and mRNA Measurements Provide Insights into Dynamics and Regulation of Yeast Translation. *Cell Rep,* 14**,** 1787-99.

Wildeman, A. G., Sassone-Corsi, P., Grundstrom, T., Zenke, M. & Chambon, P. (1984). Stimulation of in vitro transcription from the SV40 early promoter by the enhancer involves a specific trans-acting factor. *EMBO J,* 3**,** 3129-33.

Xiao, H., Friesen, J. D. & Lis, J. T. (1995). Recruiting TATA-binding protein to a promoter: transcriptional activation without an upstream activator. *Mol Cell Biol,* 15**,** 5757-61.

Xiao, W., Adhikari, S., Dahal, U., Chen, Y. S., Hao, Y. J., Sun, B. F., Sun, H. Y., Li, A., Ping, X. L., Lai, W. Y., et al. (2016). Nuclear m(6)A Reader YTHDC1 Regulates mRNA Splicing. *Mol Cell,* 61**,** 507-19.

Zhou, J., Wan, J., Gao, X., Zhang, X., Jaffrey, S. R. & Qian, S. B. (2015). Dynamic m(6)A mRNA methylation directs translational control of heat shock response. *Nature,* 526**,** 591-4.

Zid, B. M. & O’shea, E. K. (2014). Promoter sequences direct cytoplasmic localization and translation of mRNAs during starvation in yeast. *Nature,* 514**,** 117-21.

STAR Methods

CONTACT FOR REAGENT AND RESOURCE SHARING

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Reuven Agami (r.agami@nki.nl).

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Human breast adenocarcinoma MCF7 cells (female) were purchased from ATCC and authenticated by detecting the deficiency of caspase-3 using Western blotting. These cells were grown in DMEM media (Gibco) supplemented with 10% bovine serum and penicillin/streptomycin (Gibco) in 5%CO_2-buffered incubators at 37°C. Cells were split twice per week and kept in culture for up to 8 weeks. To create cells compatible with FLP recombination system (ThermoFisher scientific), MCF7 cells were transfected with the pFRT/*lac*Zeo and grown in the presence of Zeocin (400μg/ml) until stable colonies appeared. These cells were named MCF7/FRT and used for integration of Pro-Lib. To integrate Pro-Lib, these cells were transfected with Pro-Lib vectors along with pOG44 vector (ThermoFisher scientific) in quantities of 1.6μg+3μg respectively, per 35-mm well. After 48h, the cells were collected, re-plated onto 10-cm dishes, and selected for three weeks in the presence of Hygromycin (0.1 mg/ml f.c.). After appearance of the stable colonies, the cells were collected, frozen and used for further experiments. To create cells compatible with T-REx™ inducible system (Invitrogen), MCF7/FRT cells were transfected with pcDNA6/TR vector encoding for the tetracycline-sensitive repressor and separate clones were selected for their resistance to Blasticidin S (5mg/ml). The induction efficiency was examined in separate clones by transient transfection with pcDNA4/TO-Rluc construct, incubation with Doxycycline (Sigma, 1μg/ml final conc.) for 17 hours and measurement of Renilla protein activity. One suitable clone was chosen for further experiments and the cells were named MCF7/FRT/TR. These cells were stably transfected with pcDNA5/FRT/TO plasmid encoding for barcoded Renilla ORF (TRex-Rluc) in the presence of pOG44 vector (ThermoFisher scientific) and grown for three weeks in the presence of Hygromycin (0.1 mg/ml f.c.). Individual colonies were isolated and tested by treatment with Doxycycline (Sigma, 1μg/ml final conc.). 293 cells bearing integrated alpha-amanitin resistant mutants of Rpb1("wild-type" N792D, and "slow" C4/R749H), were a kind gift of D. Bentley. They were maintained in DMEM supplemented with 10% FBS, 1% penicillin/streptomycin, 200μg/mL hygromycin, and 6.5μg/mL blasticidin. To stimulate the expression of Rpb1 mutants, cells were induced with 2.0 μg/mL doxycycline for 16 hours and treated with 2.5 μg/mL α-amanitin for additional 42 hours.

METHOD DETAILS

Manipulations in cells. To induce autophagy, MCF7 cells were incubated in EBSS (Gibco) medium for 3 hours. To inhibit mTORC1, Torin1 was added to the medium for 2 hours (250 nM; Tocris Bioscience, Bristol, UK). Cells were irradiated at 10Gy dose and then incubated for additional 18 hours at normal conditions. To arrest cells in mitosis, Taxol (Sigma-Aldrich, 1μM f.c. for 17 hours) was added to the growth medium. FuGENE6 (Roche) was used for transfection of DNA plasmids, DharmaFECT1 (Dharmacone) for transfection of siRNAs and in-vitro transcribed mRNAs, according to the manufacturers' instructions. To impede the dynamics of transcribing RNA polII, cells were treated with Camptothecin (CPT, Sigma) to the final conc. of 6μM for the indicated times.

DNA cloning. Psi-Check2 vector with cloned FRT sequence and hygromycin-resistance gene served as the basic vector for the construction of Pro-Lib. To clone human promoters upstream of the Renilla ORF in Pro-Lib, putative promoter regions were cloned between BglII and NheI sites of the basic Pro-Lib vector (Figure S1A) after addition of random barcodes and removal of the SV40 promoter. All barcodes were sequenced as well as the proximal parts of the cloned promoter regions to confirm their identity. All other cloning procedures were performed using standard procedures; DNA was extracted from gel using Agarose Gel DNA Extraction Kit (Roche); all constructs were sequenced. The list of the plasmids used in the study together with their details is available in Table S3.

Luciferase assay. Luciferase assay was measured using Dual-Luciferase Reporter assay kit (Promega). Cultured cells were lysed with passive lysis buffer for 15 min at room temperature. Renilla (Rluc) and Firefly (Fluc) enzymatic activities were assayed with the substrates supplied with the kit using Centro XS^3 LB960 machine (Berthold technologies). Typically, expression of Fluc was used to normalize the Rluc expression. Myc activity was measured using Myc-reporter constructs (pGL3-M4 and pGL3-M4-mut, (Nagel et al. 2008). MCF7 cells ($1x10^5$ cells) were plated in 12-well plates and transfected with 200ng of the reporter constructs using Fugene-6 (Promega). Luciferase activity was measured 36 hours after transfection.

Barcoded polysomal profiling. Sucrose gradients for separation of polysomes were usually prepared by gentle sequential addition of 2.2ml of the different sucrose solutions (e.i., 47, 37, 27, 17 and 7% in Tris-HCl pH=7.5 (f.c. 20mM), $MgCl_2$ (f.c. 10mM) and KCl (f.c. 100mM), supplemented with 2mM DTT, Ribosafe RNAse inhibitor (Bioline, 1μl/ml) and CHX (100μg/ml) into a 12-ml tube (Beckman, 9/16 x 3 ½ in.) and left overnight at 4°C to achieve continuous gradient prior to the

centrifugation. Cells growing in subconfluent conditions (typically ≈80-85% confluency) were washed once with ice-cold PBS, collected into 15-ml tubes and incubated with PBS/cycloheximide (CHX, 100µg/ml) for 5 min on ice. After sedimentation (400xg for 3min), the cells were lysed in lysis buffer (20 mM Tris-HCl (pH 7.5), 10 mM $MgCl_2$, 100 mM KCl, 1% NP40) supplemented with EDTA-free cOmplete Protease Inhibitors (Roche), 2mM DTT, Ribosafe RNAse inhibitor (Bioline, 1µl/ml) and CHX (100µg/ml) for 10min on ice. To remove cell aggregations, the lysates were passed three times through a 25G needle. The lysates were centrifuged 1300xg for 10 min at 4°C and the supernatants were transferred into new tubes. From the cleared lysates, 500µl were loaded on top of each gradient, mounted on SW41TI rotor and centrifuged at 36000rpm for 2 hours at 4°C. Following the centrifugation, each gradient was split into 15 equal fractions of 760µl, which were subjected to RNA isolation using TRIsure (Bioline) according to the manufacturer's instructions. RNA was precipitated from 750µl of the collected material, using 2µl of GlycoBlue (Ambion) and 750µl of isopropanol, washed once with 70% of ethanol and reconstituted in 25µl of water. Typically, 7µl of total RNA were taken as a template for reverse transcription reactions using Tetro cDNA synthesis kit (Bioline) and oligo dT primers. Barcodes of the Pro-Lib transcripts detected by qRT-PCR using specific primers. To construct libraries ready for deep sequencing, barcodes were PCR amplificed using indexed primers bearing P5, P7 and Illuseq sequences, while each collected fraction was associated with a unique index. The PCR products were cleaned using QIAquick PCR purification kit (Qiagen), size selected on E-Gel SizeSelect 2% gels (Invitrogen) and sequenced using Illumina Hi-seq 2000 or 2500.

RNA manipulations. RNA was isolated from cultured cells using TRI-sure reagent (Bioline) according to the manual, using Glycoblue (Ambion) as a carrier. RNA was typically reconstituted in 30µl of sterile nuclease-free water (Gibco) and stored at -20°C. To analyze RNA on agarose gels, a portion of isolated RNA was mixed with equal volume of 2xRNA-loading dye (Thermo Fisher Scientific), supplemented with 20ng/ml Ethidium bromide, heated at 70°C for 5min and loaded on 1.5% agarose gel. After separation at 100mV for 30min, gels were visualized and documented. Reverse transcription was done using Tetro cDNA synthesis kit (Bioline) according to the manual. Typically, 7µl of total RNA was taken for a single RT reaction. Real time PCR experiments were performed using SensiFAST CYBR (Bioline) reagent and LightCycler 480II (Roche) using standard conditions.

***In vitro* synthesis of mRNA.** To synthesize *Renilla* mRNA using nuclear extract, we used HeLaScribe Nuclear Extract in vitro Transcription System (Promega) according to the manufacturer's protocol. Briefly, DNA templates including promoter, ORF,

UTRs and polyA site were PCR-amplified from the appropriate pGEM-T vectors, isolated on agarose gels and cleaned. For a single reaction, 200ng of a clean template were transcribed with 7µl of nuclear extract, in a buffer containing 1.5mM $MgCl_2$ in a siliconized RNAse-free 1.5-ml eppendorf tube (Ambion) at 30°C for 1.5 hours. The reaction was stopped by addition of TRIsure compound (Bioline) and RNA was isolated, analysed on agarose gels, quantified using NanoDrop, compared using ImageJ and stored at -20°C. Transfection was done in 96-wells plates using Dharmafect1 agent (Thermo Scientific), according to the manufacturer's instructions. Approximately 10ng of recovered mRNA were transfected as x1-amount. Rluc activity was measured ≈24 hours following transfection. For m^6A measurements, reconstituted synthetic RNA was furthermore cleaned using RNeasy kit (Qiagen). To synthesize *Renilla* mRNA using T7 RNA polymerase, we used HiScribe T7 High Yield RNA Synthesis kit (NEB) according to the manufacturer's instruction. Briefly, PCR-produced DNA fragments encoding for *Renilla* ORF were transcribed in presence of Ribo m7G Cap Analog (Promega, 8mM f.c.) at 37°C for 2 hours followed by DNAse treatment for 15 min and isolation of RNA. To produce transcripts with methylated content, N6-Methyladenosine-5'-Triphosphate (Trilink, 100mM stock conc.) was added together with ATP (supplied in the kit) at the indicated proportions. After recovery, the RNAs were polyadenylated using *E.Coli* Poly(A) polymerase (NEB) and purified. After recovery and analysis on gels, RNAs were transfected into living cells as described above or subjected to *in-vitro* translation.

***In vitro* translation of mRNAs**. To translate mRNAs encoding for Rluc protein, rabbit reticulocytes system, nuclease treated (Promega) was employed, according to the manufacturer's protocol. All reactions were done in volume of 25µl at 30°C for 10 min. To quantify the produced Rluc protein, aliquots of 5µl were taken for luciferase measurements using Dual-Luciferase Reporter assay kit (Promega).

Northern blotting. Total RNA was extracted from growing MCF7 cells as described and enriched for mRNA using Oligotex kit (Qiagen). Equal volumes of reconstituted RNA were separated on 1.2% agarose gel supplemented with 0.66% (vol/vol) formaldehyde and transferred onto Amersham Hybond-N^+ membrane (GE Healthcare). RNA probes targeting *Renilla* or *Firefly* mRNAs were produced by *in-vitro* transcription (T7 high yield kit, NEB) of PCR fragments encoding for sequences complementary to these transcripts (see Table S4 for used oligos) using $^{32}P(\alpha)$-UTP. Following cleaning of the probes using MicroSpin G-50 columns (GE Healthcare), hybridization was performed in ULTRAhyb Hybridization buffer (Ambion) rotating at 68°C for 4 hours. The membrane was washed twice with 2xSSC buffer supplemented with 0.1%SDS and twice with 0.2xSSC buffer supplemented

with 0.1%SDS. The signal was detected by FujiFilm FLA-3000 PhosphorImager and analyzed using ImageJ.

Measurement of m6A content. To quantify the m^{6}A content of RNA, EpiQuik m6A RNA Methylation Quantification Kit (Epigentek) was used. Briefly, equal volumes of RNA solution (4-8µl) were processed according to the manufacturer's instructions, aside with the negative, positive and standards controls.

5'RACE analysis. Generally, 5'RACE experiments followed the procedure established for 5' RACE System for Rapid Amplification of cDNA Ends, v2.0 (Invitrogen). Briefly, total RNA was extracted with TRIsure (Bioline) and subsequently poly-A selected with Oligotex Direct mRNA mini kit (Qiagen). For each 5' RACE assay, 500ng of poly-A selected mRNA were used for cDNA synthesis with the following primer: CCCTTCTCCTTGAATG. Synthesized cDNA were purified with QIAquick PCR purification kit (Qiagen). Then the cDNA were dC-tailed with TdT for 10min at 37°C and a first round PCR was performed with a gene specific primer (GSP2), AGTTTCCGCATGATCTTGCTTG, and Abridged Anchor Primer (AAP). A nested PCR was performed with a nested gene specific primer (GSP1), GTTGATGAAGGAGTCCAGCACGTT, and Abridged Universal Amplification primer (AUAP). For determination of the sequences of 5' RACE PCR products, the nested PCR products were ligated into pGEM-T vector (Promega) and transfected into competent bacteria. At least 20 different bacterial colonies from each ligation were subsequently analyzed with Sanger sequencing.

3'RACE analysis.

The protocol was adapted from 3'RACE system for rapid amplification of cDNA ends (Invitrogen). Briefly, total RNA was poly-A selected and reverse transcribed with oligo(dT)-containing adapter primer (AP) GGCCACGCGTCGACTAGTACTTTTTTTTTTTTTTTT. Synthesized cDNA was treated with Rnase H to remove residual RNA templates and PCR-amplified with Rluc specific primer CTGAGAGTGTCGTGGACGTGA and AAP. A nested PCR was performed with a nested Rluc primer GCCTAAGATGTTCATCGAGTCC and AAP.

Sequencing library preparation of 5' and 3'RACE
PCR products from 5' and 3' RACE preparations were PCR-amplified with Illumina sequencing compatible primers (p5 primer, 5'RACE-p7 primer, and 3'RACE-p7 primer). PCR products were then purified with CleanPCR magnetic beads (CleanNA), and pooled with equal molar, ran on Illumina Miseq platform.

Poly(A) tail analysis.

Determination of poly(A) tail-length was performed with USB® Poly(A) tail-length assay kit (Affymetrix) following the manufactor's protocol. In short, in-vitro transcribed RNA was G/I tailed, and reverse-transcribed with specific adapter primer (sequence not provided). The poly G/I tailed cDNA was then amplified with a Rluc specific primer (GCGTGCTGAAGAACGAGCAGTAA) and a provided universal PCR reverse primer. Finally, PCR products were resolved on a 2.5% agarose gel with ethidium bromide.

Lentiviral infection. To generate viruses with the Renilla-expressing plasmid, HEK293T cells were plated 24h prior transfection. For virus production, 3.5µg of pVSV-G, 5µg of pMDL-RRE and 2.5µg of pRSV-REV was transfected with 10µg of the expression plasmid pLenti-puro-Rluc. Lentivirus-containing supernatant was harvested 48h after the transfection, filtered through a 0.45µm membrane (Milipore Steriflip HV/PVDF) and kept at -80°C. Subsequently, to infect MCF7 cells, 1x10^6 cells were plated in 10-cm dish one day prior infection. For low MOI infection, cells were infected with 20µL of the virus-containing supernatant; for high MOI infection, cells were infected with 900µL (45-fold) of the supernatant. Cells were refreshed with DMEM 18h after infection and subsequently recovered for 24h. Afterwards, low MOI-infected MCF7 cells were selected with 1 mg/mL puromycin and high MOI-infected MCF7 cells were selected with various concentrations of puromycin (1ug/mL, 5ug/mL, 10ug/mL and 25ug/mL) for 1 week.

CRISPR/Cas9-mediated mutagenesis. For the mutagenesis of c-Myc TATA box sequence, lentivirus containing sgRNAs targeting the TATA box sequence were generated and subsequently used for infecting MCF7 cells as described above. After puromycin (2µg/ml) selection, cells were plated in a 96-well plate following serial dilutions in order to achieve single cell clones of the Cas9-generated mutants. Several single cell clones were collected, expanded and sequenced. To sequence the indel mutations within the c-Myc locus, genomic DNA was extracted from the grown colonies using DNeasy Blood and Tissue Kit (Qiagen), PCR amplified using c-Myc promoter-specific primers, sub-cloned into pGEM-T vector (Promega) and sequenced.

Western blotting. MCF7 cells were washed with ice-cold PBS and scraped off the plate. Cells were centrifuged at 1000xg for 10min at 4°C. Subsequently, cell pellets were lysed, by adding an appropriate amount of ice-cold lysis buffer (20mM Tris-HCl at pH 7.5, 150mM NaCl, 1.8mM $MgCl_2$ and 0.5% NP40) supplemented with cOmplete protease inhibitor cocktail (Roche) and 1mM DTT. Cell lysates were

incubated on ice for 20min with occasional vortexing followed by centrifugation at 15000xg for 15min at 4°C. Protein concentrations were determined using Pierce BCA protein assay kit (Thermo Scientific). Proteins were separated on SDS-PAGE gels and transferred onto Nitrocellulose membrane (Bio-rad, 0.2µM), which was blocked in 5% Blotting-grade blocker and probed with the indicated antibodies. All antibodies were diluted in PBS-0.25% Tween solution. For detection, blots were reacted with the SuperSigna West Dura Extended Duration Substrate (ThermoFisher). Images were captured with Chemidoc XRS+ (Biorad) and analyzed with Image Lab software; quantifications were done using ImageJ.

m^{6}A RNA immuno-precipitation (MeRIP). Immunoprecipitation of m6A was adapted from the protocol of EpiMark® N6-Methyladenosine Enrichment Kit (New England Biolabs). Total RNA was isolated as mentioned above and enriched for mRNA using Oligotex mRNA kit (Qiagen) following manufacturer's instruction. The isolated mRNA was fragmented with RNA fragmentation reagent (Ambion) for 5min, 94°C and purified through ethanol precipitation. To bind antibody to the beads, protein G beads (Invitrogen) were pre-incubated with 1µL of anti-m^{6}Aantibody (Neb, Cat. E1610S) in IPP buffer (150 mM NaCl, 10 mM Tris-HCl, pH 7.5, 0.1% NP-40)at 4°C. Subsequently, the beads were washed twice in IPP buffer and incubated with RNA for 3h with head-to-tail rotation at 4°C (10% of the material were kept as input control). Afterward, beads were washed twice in IPP buffer, twice in low salt buffer (50 mM NaCl, 10 mM Tris-HCl, pH 7.5, 0.1% NP-40), and twice in high salt buffer (500 mM NaCl, 10 mM Tris-HCl, pH 7.5, 0.1% NP-40). RNA was eluted with 30µL buffer RLT (Qiagen, Cat. 79216) for 5min and subsquently purified through ethanol precipitation. For quantification of precipitation, both input samples and IP eluates were examined by qRT-PCR as described above.

RNAPII immuno-precipitation. MCF7 cells growing on 15-cm dishes were treated with CPT (Sigma, 6µM f.c.) or DMSO (as a control) for 5 hours. After collection on ice, the cells were incubated in 8ml of PBS supplemented with 1% formaldehyde for 10 min at room temperature in constant rotation. After quenching with glycine (0.125M f.c.) for 5 min, the cells were spun, washed once with PBS, flash-frozen in liquid nitrogen and kept at -80°C. After thawing in 1ml of lysis buffer (50 mM HEPES-NaOH-pH7.5, 150 mM NaCl, 5 mM $MgCl_2$, 0.1% NP40) supplemented with EDTA-free cOmplete Protease Inhibitors and PhosSTOP (Roche), the lysate was divided into three equal 2-ml eppendorf tubes and sonicated at 4°C using Bioraptor (Diagenode) for 10 cycles (each cycle including 30sec of high-power sonication followed by 30sec pause). The lysates were centrifuged for 10 min. at 1800xg at 4°C to remove unbroken cells. Equal volumes of supernatants were added to 20 µl of Dynabeads-Protein G suspension (Life Technologies) pre-coupled with 5µg of

anti-RNA polymerase II antibodies (Millipore, clone CTD4H8) or normal mouse IgG (Santa Cruz) in PBS and 1% BSA for 2 hours at room temperature. The IP reaction was carried out for 2 hours at 4°C with constant rotating. After that, the beads were washed once with the regular lysis buffer and three times with stringent lysis buffer (as detailed above, but with 300 mM NaCl). The last washing was done with PBS, and the beads were transferred into new 1.5-ml tubes and the bound proteins were eluted and reverse-crosslinked by incubation with 50µl of 1x Laemmli sample buffer at 100°C for 10min. After collection, the eluates were resolved on 8% SDS-PAGE and probed with the indicated antibodies.

Ribosomal profiling. Ribosome Profiling (RP) was performed as previously described (Loayza-Puch et al. 2016). Briefly, CHX (100µg/ml) was added to the medium of growing cells and incubated for 5min at 37°C. Approximately 30x10^6 cells were collected in ice-cold PBS and lysed in lysis buffer (20 mM Tris-HCl pH 7.8, 10 mM $MgCl_2$, 100 mM KCl, 1% Triton X-100, 2 mM DTT, 100µg/ml CHX, cOmplete protease inhibitors (Roche). Lysates were passed through a 26G needle (BD bioscience) for further homogenization, and centrifuged at 4°C 1300xg for 10 minutes. Supernatants were treated with 2.5U/µl of RNase I (Ambion) for 45 min at room temperature in gentle constant rotation. Lysates were loaded onto a linear sucrose gradient (7%-47%, as detailed above), and fractionated by ultracentrifugation, using a SW-41Ti rotor at 36000 rpm (221632.5g) during 2 hours. The sucrose gradient was divided into 14 fractions of 830µl. The monosome-enriched sucrose fractions (7 to 10) were collected and treated with proteinase K (PCR grade, Roche), in presence of 1% SDS. The so-released ribosome protected fragments (RPFs) were purified using TRIsure reagent (Bioline), following the manufacturer's instructions. RPFs between 30 and 33 nucleotides in length were size-selected in 10% acrylamide gel and isolated. The 3' ends of the RPFs were dephosphorylated by treatment with T4 polynucleotide kinase (New England Biolabs) for 6 hours at 37°C in MES-NaOH buffer (100mM MES-NaOH, pH 5.5, 10mM $MgCl_2$, 10mM β-mercaptoethanol, 300mM NaCl). 3' adaptor (RA3) was ligated using T4 RNA ligase 1 (New England Biolabs) for 3.5 h at 37°C, in absence of ATP. Ligation products were size selected in 10% acrylamide gel, and the 5' ends were phosphorylated by treatment with T4 polynucleotide kinase for 30 minutes at 37°C, in presence of 1mM ATP. 5' adaptor (RA5) was ligated using T4 RNA ligase 1 (New England Biolabs) during 2.5h at 37°C, and the ligation products were selected in 10% acrylamide gel. Ribosomal RNA depletion was performed by biotin-streptavidin affinity purification using biotinylated ribosomal RNA probes and streptavidin dynabeads. Retro-transcription of the ligation products into cDNA was performed using Super Script III reverse-transcriptase (Invitrogen) following the manufacturer's instructions, and the primer RTP. PCR amplification

was performed using the forward primer RP1, and the reverse primer RPI that contained a hexanucleotide index used to multiplex different samples during next generation sequencing (NGS). PCR products were size selected by E-Gel electrophoresis (Invitrogen), and submitted to NGS using HiSeq 2500 System (Illumina). The sequence of 3' and 5' adapters, RTP, RP1 and different RPI primers is available in Table S5.

RNA-seq

Total RNA Isolation. Total RNA was extracted using TRIzol reagent (15596-018, Ambion life technologies) according to the manufactures protocol. Briefly, 0.2x volumes of chloroform (Chloroform stab./Amylene, Biosolve) was added to the Trizol homogenate and the tube(s) (Falcon, 15mL) were shaken vigorously. The tube(s) were incubated for 2-3 minutes at room temperature and centrifuged (Hettich, rotanta 46 RS) for 1 hour at 4 °C. Approximately 70% of the upper aqueous phase was transferred to a clean 15 mL tube and 0.5x volume of isopropanol (33539, Sigma-Aldrich,) was added. The tube(s) were incubated overnight at -20 °C and centrifuged for 30 minutes at 4 ° C. The supernatant was removed and the pellet was washed twice with 80% ethanol (32221-2.5L, Sigma-Aldrich). The total RNA pellet was air-dried for 8 minutes and dissolved in an appropriate volume of nuclease free water (Ambion) and quantified using Nanodrop UV-VIS Spectrophotometer. The total RNA was further purified using the MinElute Cleanup Kit (74204, Qiagen) according to the manufactures instructions. Quality and quantity of the total RNA was assessed by the 2100 Bioanalyzer using a Nano chip (Agilent, Santa Clara, CA). Total RNA samples having RIN>8 were subjected to library generation.

TruSeq Stranded mRNA sample preparation. Strand-specific libraries were generated using the TruSeq Stranded mRNA sample preparation kit (Illumina Inc., San Diego, RS-122-2101/2) according to the manufacturer's instructions (Illumina, Part # 15031047 Rev. E). Briefly, polyadenylated RNA from 1000ng intact total RNA was purified using oligo-dT beads. Following purification the RNA was fragmented, random primed and reverse transcribed using SuperScript II Reverse Transcriptase (Invitrogen) with the addition of Actinomycin D. Second strand synthesis was performed using Polymerase I and RNaseH with replacement of dTTP for dUTP. The generated cDNA fragments were 3' end adenylated and ligated to Illumina Paired-end sequencing adapters and subsequently amplified by 12 cycles of PCR. The libraries were analyzed on a 2100 Bioanalyzer using a 7500 chip (Agilent, Santa Clara, CA), diluted and pooled equimolar into a 10nM multiplex sequencing pool.

Sequencing. The libraries were sequenced with 65 base single reads on a HiSeq2500 using V4 chemistry (Illumina Inc., San Diego).

Quantification and statistical analysis

While mentioned, *n* represents the number of biological repeats of the same experiment. Data are represented as mean±SEM. P-values presented in Figures 3A,E; 4A,D; S5B,C; S6A,B were calculated using Wilcoxon's test. P-values mentioned in other figures were calculated using t-test (one-tail distribution, homoscedastic variance). In all figures, *=p<0.05, **=p<0.01, ***=p<0.005, ****=p<0.001.

Polysomal data analysis and preparation of heatmaps. Polysomal fractions and promoter vectors were identified by their unique indexes and barcodes, respectively. Expression count matrix was derived by counting the frequency of each barcode in each fraction. To avoid fluctuations due to low coverage, only vectors covered by at least 10,000 counts (summation over all fractions) were included in the analyses. For heatmap display, vector counts (rows) were standardized to mean=0 and SD=1. Heatmaps were generated using the EXPANDER package (Ulitsky et al. 2010). Order of promoter vectors in the heatmap was determined by hierarchical clustering of the rows.

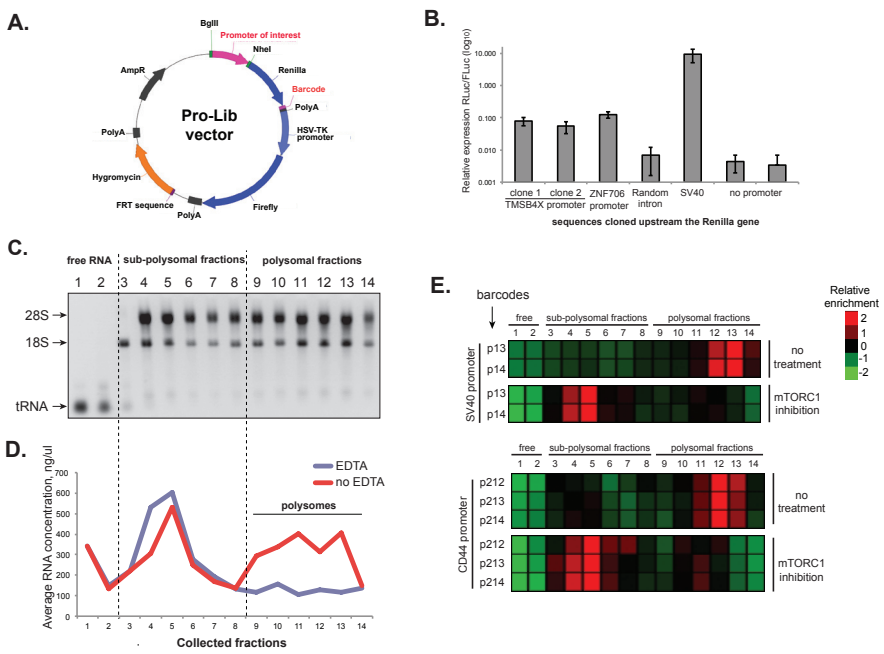
RNA-seq and Ribo-seq analysis. Sequenced reads were aligned to a reference set of human curated protein-coding transcripts (plus the five human rRNA transcripts) using Bowtie (Langmead et al. 2009). This reference set of transcripts was based on Ensembl's gene annotations (release 69). For genes with multiple isoforms, the one with longest coding DNA sequence (CDS) region and, in case not unique, the one with longest UTRs among the ones with the longest CDS, was selected to represent the gene. Only uniquely mapped reads were used in subsequent analyses. RNA expression levels and ribosome occupancy were estimated by calculating reads per kilobase of mRNA per million reads (RPKM) per transcript, taking into account either all reads that map to the transcript (for estimation of RNA levels using RNA-Seq data) or only those mapping to its CDS (for estimation of ribosome occupancy). In estimation of ribosome occupancy in CDS, 5' ends of reads were offset 12 nucleotides to the 3' direction to match the P-site location of ribosome (Ingolia et al. 2009). Translation efficiency (TE) was estimated by the (log_2) ratio between ribosome occupancy and mRNA level. Only genes with at least 20 reads in RNA-seq and Ribo-seq samples were included in analyses.

GRO-seq analysis. Sequenced reads were aligned to the human genome (hg19) using bowtie2, and number of reads mapping to annotated genes (Ensembl v69) were counted by HTseq (Anders et al. 2015). Quantile normalization was then applied to allow comparisons between different samples. Only genes covered by at least 20 reads were included in the analysis.

Data and software availability

The deep sequencing datasets generated in this study have been deposited in the GEO DB under accession number GSE96643.

Supplemental data



Supplemental Figure 1

Figure S1.

Setup of the promoter library (Pro-Lib) and barcoded polysomal profiling (BPP), Related to Figure 1. (A) Schematics of the Pro-Lib vector. The reporter Rluc gene is flanked upstream by sequences encoding for putative promoters of interest (magenta colored) and downstream by unique 10-nt barcodes (also magenta colored). Transcription of the control Fluc gene is driven by HSV-TK promoter in all vectors. (B) Several Pro-Lib vectors with the indicated regions cloned upstream of Rluc gene were integrated into MCF7/FRT cells, grown separately and subjected to dual Rluc/Fluc assay. Note the uniform expression of the tested constructs that is substantially higher than the background signal (assessed by expression of vectors with no promoter or with random intron sequence cloned instead of a promoter) and milder than the expression of the strong SV40 promoter. Data are represented as mean±SEM of n=3. (C) Total RNA isolated from fractions of a typical BPP experiment were separated on 1.5% agarose gels and documented. Note the characteristic segregation of RNA as well as the relative enrichment of ribosomal RNAs in the polysomal fractions. (D) Total RNA concentrations in the different fractions were measured following lysis of cells in the presence or absence of EDTA and polysomal profiling procedure. Note the dramatic drop of the RNA content in the fractions 9-13 upon presence of EDTA in the lysis buffer. (E) BPP experiment identifies two *Rluc* transcripts (driven by SV40 and CD44 promoters) exhibiting super-sensitivity to the inhibition of mTORC1. Note the multiple barcodes for each promoter showing identical behavior. (F) The transcripts mentioned in (E) contain putative 5'TOP sequences (marked in blue). The TSS of the SV40 promoter was reported previously (Byrne et al., 1983), while the TSS of the CD44-driven construct was determined by 5'-RACE analysis in this study.

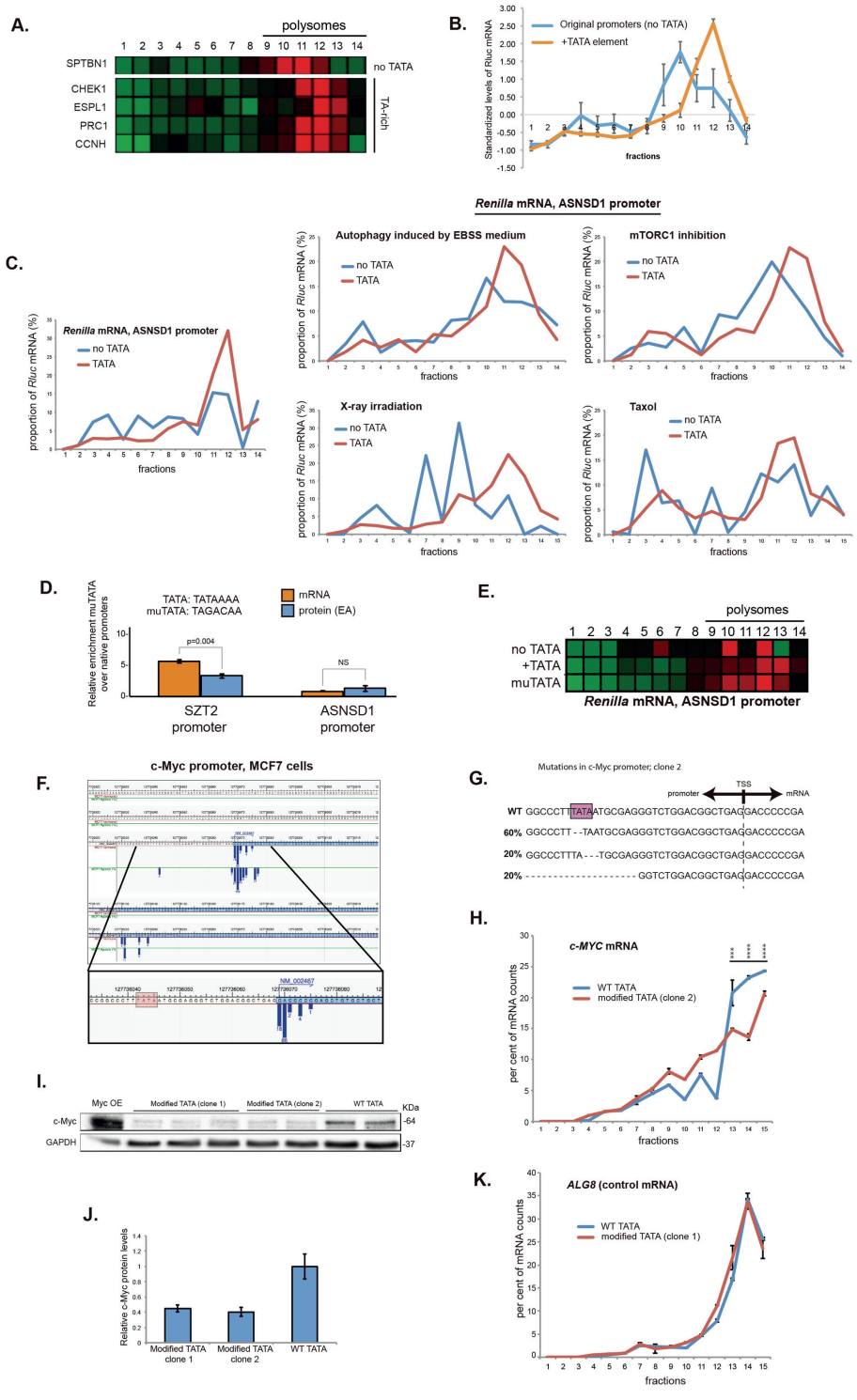
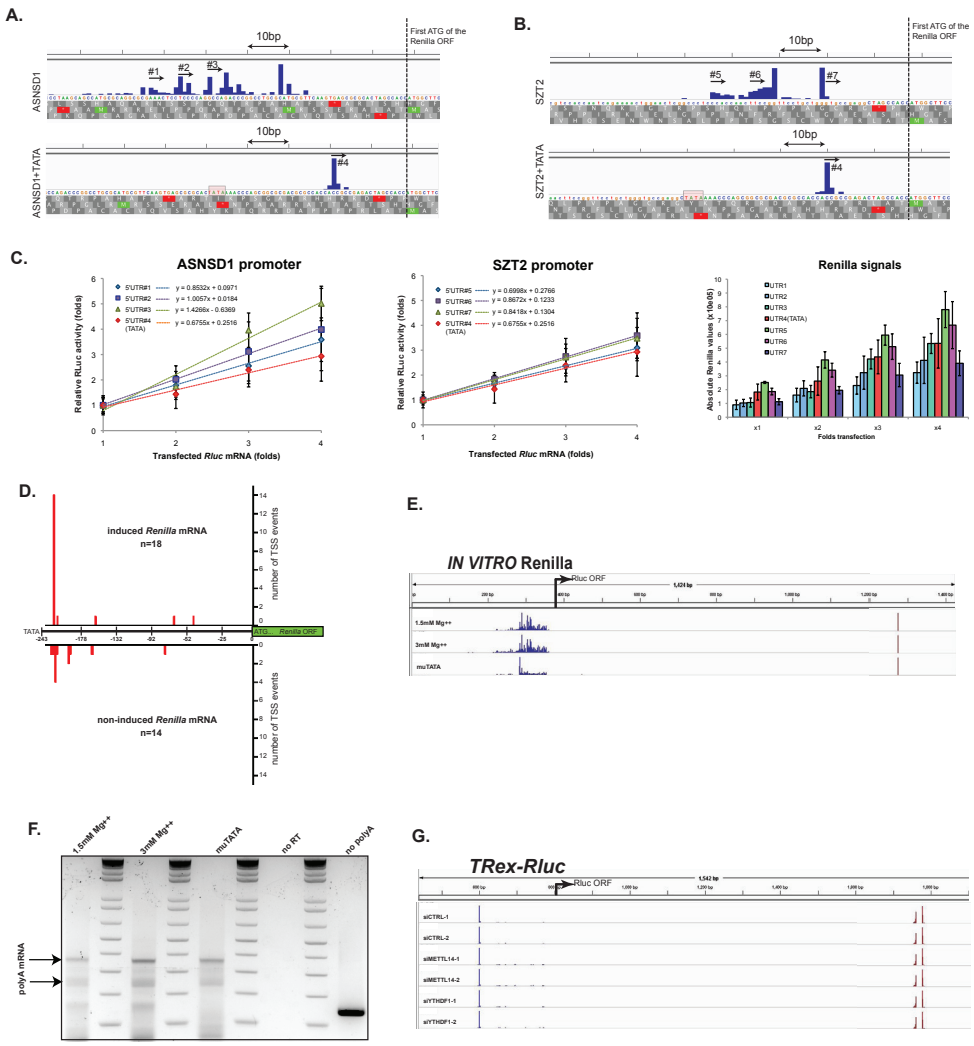


Figure S2.

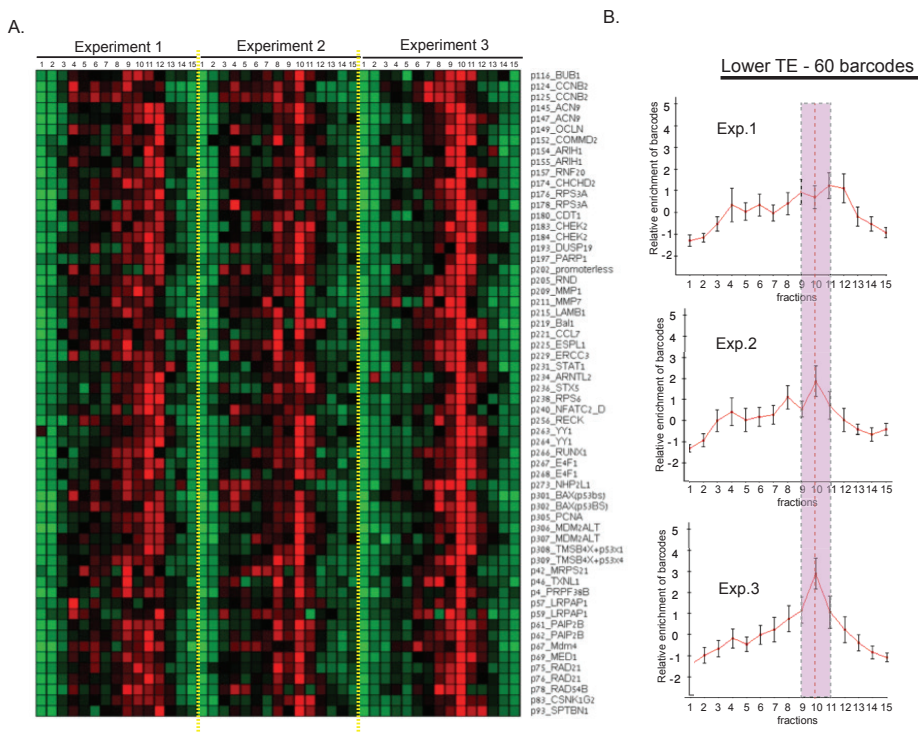
Effect of TATA box on TE, Related to Figure 2. (A) Pro-Lib vectors encoding for promoters with TA-rich regions at their 3' ends result in Rluc mRNAs shifted to the denser fractions of the gradient, compared to other transcripts (e.g., under SPTBN1 promoter). (B) Average standardized profiles of polysome fractions for the TATA+/TATA- paired constructs in Pro-Lib. (C) Promoter region of ASNSD1 gene yields better translated Rluc mRNAs when supplemented with an artificial TATA box at its 3'end. Cells expressing *Rluc* mRNAs from either native ASNSD1 promoter or with artificial TATA box were subjected to BPP and plotted for a direct comparison of TE (left panel). To test the robustness of the positive effect of TATA on TE, cells were subjected to various stress conditions, prior to BPP (right panels). Note that in all listed conditions the presence of TATA box enhances TE. (D) Point mutagenesis of the artificial TATA box reduces mRNA levels and prevents super-induction of the protein production. SZT2 and ASNSD1 promoters supplemented with an artificial TATA box were subjected to site-directed mutagenesis of the TATA box sequence (muTATA). Levels of the Rluc mRNAs and protein compared to the native promoters (lacking TATA) are shown; data are represented as mean±SEM of n=3. (E) *Rluc* mRNAs resulting from either native ASNSD1 promoter, one with the artificial TATA box (+TATA) or with a point-mutated TATA sequence (muTATA) were subjected to the BPP procedure. Note the shift towards denser fractions caused by the artificial addition of TATA box as well as the shift to the opposite direction upon mutation of the TATA sequence. (F) Screenshot of the DBTSS database (release 9.0) showing the distribution of TSS of c-Myc gene in MCF7 cells. Note that most of TSSs stem from a single TATA box-containing promoter (see zoom-in window). (G) Genotyping of the c-Myc proximal promoter region of the second clone bearing modified c-Myc TATA box. Numbers on the left represent relative abundances of the different alleles. (H) Lysates of the clone described in (G) were subjected to polysomal profiling and plotted against wild-type MCF7 cells. SEM represents three measurements of a characteristic gradient; n=2. ***p<0.005, ****p<0.001. (I) Lysates of the two clones bearing modified c-Myc TATA boxes together with wild-type MCF7 cells (WT TATA) and cells transfected with vector encoding for c-Myc (Myc OE), were resolved on SDS-PAGE and probed for c-Myc and GAPDH proteins. Replicates on the blot represent biological repeats. (J) Relative quantification of the c-Myc bands presented in (I). (K) TE of a control gene (ALG8) was assessed by probing polysomal fractions described in Figure 2J with primers detecting ALG8 mRNA. SEM represents three measurements of a characteristic gradient; n=3.



Supplemental Figure 3

Figure S3.

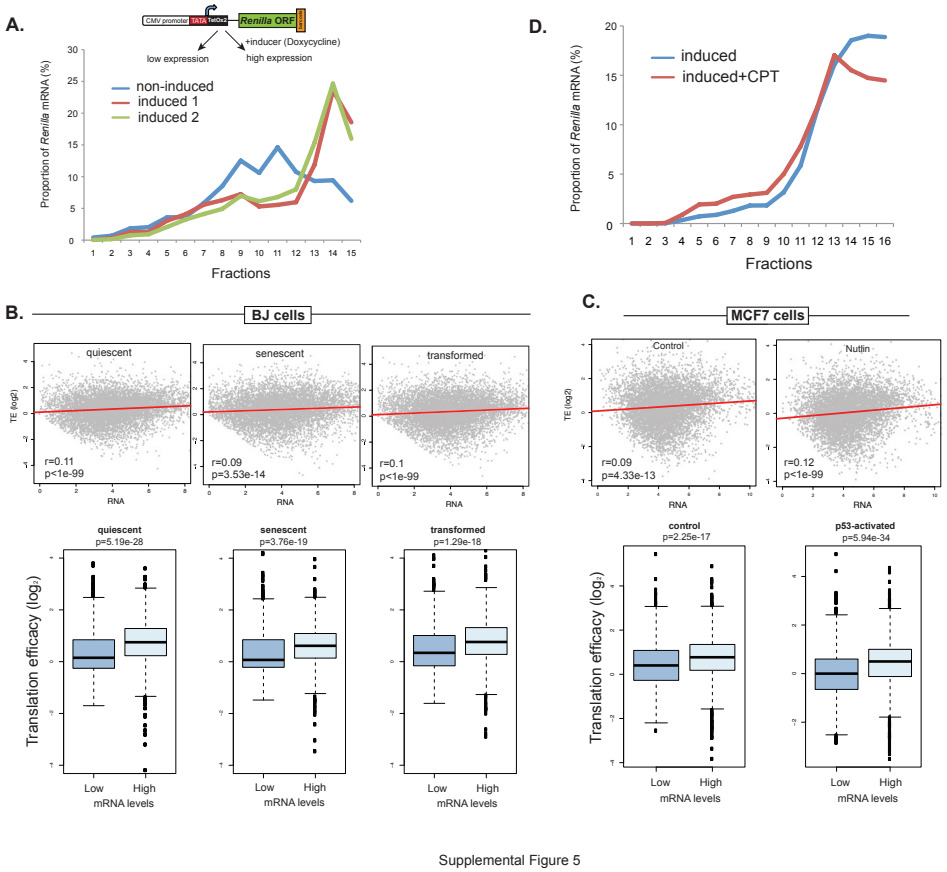
Analysis of *Rluc* mRNAs, Relates to Figures 2,3,5 and 6. *Rluc* mRNAs transcribed by ASNSD1 (A) or SZT2 (B) promoters either lacking or bearing an artificial TATA box were subjected to 5'RACE analysis in order to determine their TSSs. Reads were plotted in a quantitative manner using IGV software. Note the first ATG of the Rluc ORF and the precise focusing of the TSS by the TATA box. (C) Evaluation of the effect of the different 5'UTRs on TE. Rluc transcripts bearing the different 5'UTRs identified in A,B (see the numbered arrows) were transcribed in vitro using T7 RNA polymerase, polyadenylated and transfected into MCF7 cells in fold-wise amounts. Renilla luciferase activity was measured 24 hours later and plotted in a relative manner to draw the trend lines that represent the respective ratios of translation. The rightmost panel shows the absolute Rluc signals; n=4. (D) *Rluc* mRNAs expressed from either induced or non-induced TRex-Rluc gene were subjected to 5'RACE analysis. Reads (numbers of sequenced colonies) were plotted in a quantitative manner on a scale ranging from TATA box (-243) to ATG (+1). (E) Rluc transcripts produced *in vitro* using HeLa extract in optimal conditions (1.5mM Mg++), upon high $MgCl_2$ (3mM Mg++) or from promoter with mutated TATA element (muTATA) were subjected to 5'- and 3'-RACE analyses. The reads from both assays were analyzed and plotted in a quantitative manner using IGV software. (F) Transcripts described in (E) were subjected to analysis of the length of polyA-tails. (G) The TRex-Rluc cassette was induced for 24 hours in cells transfected with the indicated siRNAs. After isolation of total RNA, the 5'- and 3'-ends of the Rluc mRNAs were determined as detailed in (E).



Supplemental Figure 4

Figure S4.

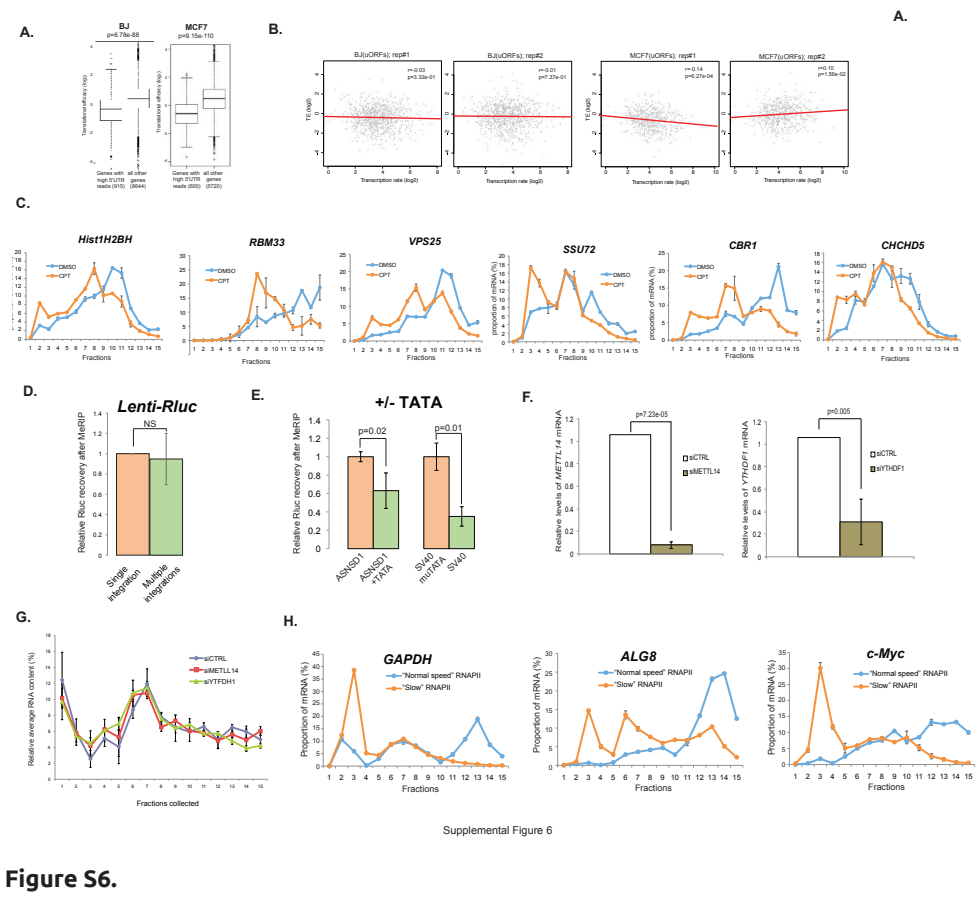
Global TE assessment of the Pro-Lib mRNAs, Relates to Figure 3. Barcoded *Rluc* mRNAs with sufficient read coverages (>10.000 reads over 15 polysomal fractions) in three independent BPP experiments were divided into two main groups by hierarchical clustering. One arm includes barcodes that showed peak in fractions 9-11 (lower TE, A-B) and the other arm includes barcodes whose polysomal profile peaked in fraction 11-13 (higher TE, C-D). Plots to the right of the heatmaps show the mean profile of the vectors assigned to each group over the three repeats; error bars=SD.



Supplemental Figure 5

Figure S5.

Transcription rates impacts translation, Relates to Figures 3 and 4. (A) Segregation of the induced versus non-induced *Rluc* mRNAs on sucrose gradients. Cells expressing barcoded Rluc gene were treated for 21 hours to induce the expression of Rluc or left untreated, and subjected to BPP procedure. (B) Upper panels: test for correlation between mRNA expression and TE, as described in Figure 3E, was performed here on data from BJ cells in quiescent, senescent, and transformed states. Lower panels: direct comparison between 10% of mRNAs with the lowest and 10% with the highest abundance. (C) In MCF7 cells, a similar correlation was tested in p53-activated (by Nutlin3 treatment) versus the control cells. Both datasets are from (Loayza-Puch et al., 2013), representation as detailed for (B). (D) Cells with induced Rluc were treated with CPT (or DMSO) for 5 hours and subjected to BPP analysis.



Supplemental Figure 6

Figure S6.

Relates to Figures 4,5 and 6. (A) Genes with high ribosomal occupancy in their 5'UTRs show overall reduced TE in both BJ and MCF7 cells; p-values were calculated Wilcoxon's test. (B) These genes do not display any clear correlation between their transcription rate and TE (showing two independent replicates), and were removed from the genome-wide analysis. (C) Treatment with CPT causes attenuation of translation. Transcripts showed in Figure 4E and additional mRNAs were detected in polysomal profiling experiments using primers against their 3'UTRs. (D) *Rluc* mRNAs produced in cells stably expressing Lenti-Rluc cassette integrated at either single- or multiple-copy numbers, were subjected to MeRIP procedure and analyzed by qRT-PCR. In each experiment, the level of recovery from cells with multi-copy integration was calculated relatively to the recovery of the single-copy population; n=4. (E) *Rluc* mRNAs transcribed from the promoter pairs with or without intact TATA box element were subjected to MeRIP procedure and analyzed by qRT-PCR; n=2. (F) Efficiency of METLL14 and YTHDF1 siRNA-mediated knock-down. MCF7 cells were transfected with siRNAs targeting METTL14 or YTHDF1 genes or control non-targeting siRNA. After 56 hours, the cells were harvested, subjected to RNA isolation and tested for the levels of the respective mRNAs using qRT-PCR. (G) MCF7 cells transfected with the indicated siRNAs were subjected to polysomal profiling procedure and an average levels of RNA in the collected fractions were measured using NanoDrop and plotted; n=2. (H) Cells expressing "normal speed" or "slow" mutants of RNAPII were subjected to polysomal profiling on sucrose gradients. Several mRNAs were detected using qRT-PCR to monitor changes in their TE; a characteristic profile is shown, n=2.

CHAPTER 3

The methylated way to translation

Ruiqi Han, Boris Slobodin and Reuven Agami

Adapted from Oncotarget (2017), 8(55): 93313-14.

Metabolism of messenger RNAs (mRNAs) consists of multiple steps, from transcription, through splicing, export to the cytoplasm, localization, translation to proteins and, finally, degradation. These steps, which are crucial to ensure correct genetic expression, have long been considered as separate events occurring at distinct time points and different locales. Recent studies suggest that they are not only interconnected, but might also be coupled to the initial process - transcription.

Initial studies in the field showed that the occurrence of exon inclusion and intron retention could be altered in HEK293 cells with “slow” and “fast” RNA Polymerase II (RNAPII) mutants, suggesting that the rate of transcriptional elongation is crucial for the appropriate splice form (Fong et al., 2014). Another important step of mRNA life is its export from the nucleus into the cytoplasm. It has been suggested that the export process is not merely a sequential event, but the regulation happens co-transcriptionally as well. In yeast, a group of factors that function in transcription elongation and mRNA export form a “Transcription/Export” (TREX) complex, which couples to RNA Pol II throughout the entire RNA molecule (Sträßer et al., 2002).

As splicing and export could occur co-transcriptionally, they might be directly affected by the dynamics of transcription. However, how post-transcriptional processes – such as degradation and translation - could be coupled to transcription? When single-molecule mRNA decay study measured stabilities of two mitotic mRNAs, SWI5 and CLB2 in yeast, they found that the precise control of their cytoplasmic decay is largely regulated by their promoters via co-transcriptional binding of a co-factor, Dbf2p (Trcek et al., 2011). Similarly, Bregman et al. showed that promoter sequences can recruit Rap1p protein and enhance the decay of its transcribed mRNA in yeast (Bregman et al., 2011). Thus promoters, elements that drive the transcription, are capable of regulating the destiny of mRNAs after its export into the cytoplasm. Recent evidences from several studies suggest that also the mRNA translation could be regulated, to a large extent, by transcription. First, yeast promoters were shown to recruit RNAPII subunits to facilitate translation of specific transcripts (Harel-Sharvit et al., 2010). Similarly, upon glucose starvation, specific yeast promoters were shown to bind Hsf1 to direct mRNA localization and the efficiency of translation in the cytoplasm (Zid and O’Shea, 2014). These two studies demonstrate the ability of yeast promoters, via co-transcriptional recruitment of effector proteins, to regulate translation of specific mRNAs.

Recently, we discovered a direct genome-wide link between transcription and translation in mammalian cells (Slobodin et al., 2017). While testing the effect of human promoters on translation, we identified a positive correlation between the levels of mRNA expression on their capabilities to bind ribosomes (i.e., translational efficiency,

or TE). However, we observed that mRNA levels present in the cytoplasm are not the factor that dictates TE directly. Instead, we found that rates of transcription (i.e. the strength of the promoter activity to recruit RNAPII and the speed of transcriptional elongation) positively regulate TE across different mammalian cell lines.

To study how this link is maintained, we examined a possible role of a specific RNA modification, N6-methyladenosine (m6A), in coordinating the two processes (Slobodin et al., 2017). Indeed, upon knock-down of several m6A-regulatory factors, both “writers” and “readers”, we observed a striking positive effect on the TE of mRNA that was characterized by slow transcription rates. Direct immunoprecipitation of m6A-modified mRNAs demonstrated significant difference between the methylation levels of the repressed and induced transcripts, suggesting that mRNAs that are “repressed” (i.e., possess low levels of transcription), have higher m6A content. This, in turn, represses TE, therefore linking inefficient transcription to inefficient translation. Finally, by artificial slow-down of RNAPII dynamics, we showed that the process of m6A modification on mRNA is likely to be co-transcriptional.

Our study provides an additional evidence that transcription machinery is intrinsically linked with the process of translation and support the theory that the fate of a mRNA molecule could be in part pre-determined during transcription. Future studies should address the hidden link between RNAPII and m6A methylation complex and reveal how the speed of RNAPII is translated into the m6A content deposited on the transcript. Additionally, the exact role of m6A in translation still remains ambiguous. While we proposed an inhibitory role of m6A in translation, several recent studies found it stimulatory (e.g., Wang et al., 2015). The paradox could be partially explained by the precise location of the modified residue within the transcript (e.g., coding regions versus untranslated regions). Future research should examine this hypothesis and uncover how cells determine the precise sites of m6A deposition.

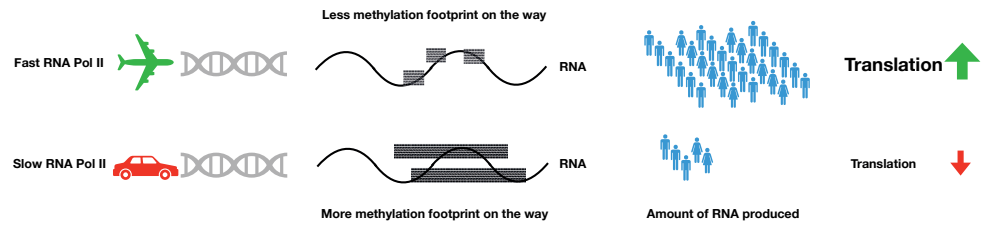


Figure 1.

Different transcription elongation rates dispose distinct m6A footprint on mRNAs, resulting in contrasting translational outcomes.

mRNAs are modified by a myriad of epigenetic alterations of which little functional knowledge is present. The functional investigation of the role of these RNA modifications is just emerging, and is expected to increase in volume in the coming few years. With the increasing knowledge of m6A regulation, its role in gene expression, involvement in diseases and the potential to serve for diagnosis (e.g. in cancer) is expected to become much clearer in the near future. The possible applications of RNA modifications, including drug development targeting m6A machineries and thus altering the epigenetic environment, could open up new avenues to a better and more effective diagnosis and treatment of human diseases stemming from imbalanced gene expression.

Affiliations:

Ruiqi Han: Division of Oncogenomics, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands.

Reuven Agami: Division of Oncogenomics, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands.
Department of Genetics, Erasmus University Medical Center, Wytemaweg 80, 3015 CN Rotterdam, The Netherlands.

Boris Slobodin: Department of Biomolecular Sciences, The Weizmann Institute of Science, Rehovot, Israel.

Correspondence to: Reuven Agami at r.agami@nki.nl

References

Bregman, A., Avraham-Kelbert, M., Barkai, O., Duek, L., Guterman, A., and Choder, M. (2011). Promoter elements regulate cytoplasmic mRNA decay. Cell *147*, 1473–1483.

Fong, N., Kim, H., Zhou, Y., Ji, X., Qiu, J., Saldi, T., Diener, K., Jones, K., Fu, X.D., and Bentley, D.L. (2014). Pre-mRNA splicing is facilitated by an optimal RNA polymerase II elongation rate. Genes Dev. *28*, 2663–2676.

Harel-Sharvit, L., Eldad, N., Haimovich, G., Barkai, O., Duek, L., and Choder, M. (2010). RNA polymerase II subunits link transcription and mRNA decay to translation. Cell *143*, 552–563.

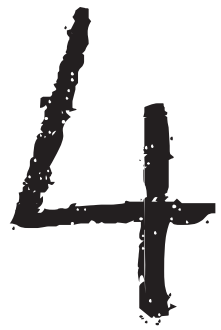
Slobodin, B., Han, R., Calderone, V., Vrielink, J.A.F.O., Loayza-Puch, F., Elkon, R., and Agami, R. (2017). Transcription Impacts the Efficiency of mRNA Translation via Co-transcriptional N6-adenosine Methylation. Cell *169*, 326–337.e12.

Sträßer, K., Masuda, S., Mason, P., Pfannstiel, J., Oppizzi, M., Rodriguez-Navarro, S., Rondón, A.G., Aguilera, A., Struhl, K., Reed, R., et al. (2002). TREX is a conserved complex coupling transcription with messenger RNA export. Nature *417*, 304–308.

Trcek, T., Larson, D.R., Moldón, A., Query, C.C., and Singer, R.H. (2011). Single-molecule mRNA decay measurements reveal promoter-regulated mRNA stability in yeast. Cell *147*, 1484–1497.

Wang, X., Zhao, B.S., Roundtree, I.A., Lu, Z., Han, D., Ma, H., Weng, X., Chen, K., Shi, H., and He, C. (2015). N^6-methyladenosine modulates messenger RNA translation efficiency. Cell *161*, 1388–1399.

Zid, B.M., and O'Shea, E.K. (2014). Promoter sequences direct cytoplasmic localization and translation of mRNAs during starvation in yeast. Nature *514*, 117–121.



CHAPTER 4

Functional CRISPR Screen Identifies AP1-associated Enhancer regulating FOXF1 to modulate Oncogene-Induced Senescence

Ruiqi Han[1,3,5], **Li Li**[1,3,5], **Alejandro Piñeiro Ugalde**[1], **Arieh Tal**[1], **Zohar Manber**[4], **Eric Pinto Barbera**[1,3], **Veronica Della Chiara**[1], **Ran Elkon**[4,6], **Reuven Agami**[1,2,3,6]

[1]Division of Oncogenomics, The Netherlands Cancer Institute, Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands.

[2]Department of Genetics, Erasmus University Medical Center, Wytemaweg 80, 3015 CN Rotterdam, The Netherlands.

[3]Oncode Institute, Amsterdam, The Netherlands.

[4]Department of Human Molecular Genetics and Biochemistry, Sackler School of Medicine, Tel Aviv University, Tel Aviv 69978, Israel.

[5]These authors contributed equally

[6]Co-corresponding authors

Adapted from Genome biology (2018), 19: 118.

Abstract

Background

Functional characterization of noncoding elements in the human genome is a major genomic challenge, and the maturation of genome-editing technologies is revolutionizing our ability to achieve this task. Oncogene-induced senescence (OIS), a cellular state of irreversible proliferation arrest that is enforced following excessive oncogenic activity, is a major barrier against cancer transformation, and therefore bypassing OIS is a critical step in tumorigenesis. Here, we aimed at further identification of enhancer elements that are required for the establishment of OIS.

Results

We first applied genome-wide profiling of enhancer-RNAs (eRNAs) to systematically identify enhancers that are activated upon oncogenic stress. DNA motif analysis of these enhancers indicated AP-1 as a major regulator of the transcriptional program induced by OIS. We thus constructed a CRISPR-Cas9 sgRNA library designed to target OIS-induced enhancers that are putatively regulated by AP-1, and used it in a functional screen. We identified a critical enhancer that we dub *Enh*$^{AP1\text{-}OIS1}$ and validated that mutating the AP-1 binding site within this element results in OIS bypass. Furthermore, we identified *FOXF1* as the gene regulated by this enhancer, and demonstrated that this target gene mediates *Enh*$^{AP1\text{-}OIS1}$ effect on the senescence phenotype.

Conclusion

Our study elucidates a novel cascade mediated by AP-1 and FOXF1 that regulates OIS and further demonstrates the power of CRISPR-based functional genomic screens in deciphering the function of noncoding regulatory elements in the genome.

Keywords

CRISPR – Functional screen – Enhancers – Oncogene-induced senescence - Gene regulation – AP1- FOS – JUN – FOXF1

Background

Over the last decade, large-scale genomic projects identified hundreds of thousands of regulatory elements in the human genome, most of them are putative enhancers(Djebali et al., 2012; ENCODE Project Consortium, 2012). Identification of candidate enhancer regions was mainly based on profiling of characteristic histone modifications (e.g., H3K27ac and H3K4me1) and binding of transcriptional activators (e.g., p300). Recently, eRNA expression, typically transcribed bi-directionally at promoter-distal cis-regulatory elements, was indicated as a

sharp feature of active enhancers, and was utilized for systematic discovery of enhancers across the genome(Andersson et al., 2014). Importantly, changes in eRNA production correlate with changes in the enhancer activity(Banerji et al., 1981; Heintzman et al., 2007). Yet, functional characterization of the plethora of candidate enhancer elements is a major genomic challenge(Rada-Iglesias et al., 2011). High-throughput reporter assays to probe the functions of regulatory regions were developed in recent years(Kim et al., 2010). However, these methods separate putative regulatory elements from their native chromosome, so that any effect of chromatin context and long range regulatory interactions is lost. Furthermore, definitive demonstration of the function of regulatory element requires their perturbation in situ. The maturation of novel genome-editing technologies is revolutionizing our ability to interrogate the function of the noncoding genome. This potential was demonstrated by pioneering CRISPR-based functional genomic screens that systematically targeted noncoding elements in the human genome(Amano et al., 2009; Bulger and Groudine, 2011; Levine, 2010; de Santa et al., 2010).

In one of these CRISPR-based functional genomic screens we focused on oncogene-induced senescence (OIS), which is a cellular state of irreversible proliferation arrest that is enforced in face of excessive oncogenic activity (oncogenic stress) [8]. OIS is a major barrier against cancer transformation (Braig et al., 2005; Lin et al., 1998) and therefore overcoming OIS is a critical step in tumorigenesis (Sage et al., 2003). Activation of this process is largely dependent on p53 (Lin et al., 1998; Serrano et al., 1997), and consequently its bypass by cancer cells is mainly achieved by emergence of somatic mutations in p53 or other components of its pathway (Korkmaz et al., 2016; Smogorzewska and de Lange, 2002). As p53 is an enhancer binding transcription factor (TF), whose function in transcriptional regulation is required for its tumor suppressive activity, we previously performed a CRISPR-based functional genomic screen that systematically targeted p53-bound enhancers(Melo et al., 2013). That screen uncovered several p53-bound regulatory elements that are required for the activation of OIS. However, aberrant expression of oncogenes can lead to activation of additional transcription factors (TFs) whose function is also critical for the establishment and/or maintenance of OIS. In the current study, we aimed at the identification of such TFs and discovery of additional enhancers that are required for the establishment of OIS. We carried out an unbiased profiling of enhancers activated upon oncogenic stress, which indicated AP-1 as a major regulator of the transcriptional program induced by OIS. We thus generated a CRISPR-Cas9 sgRNA library designed to target enhancers putatively regulated by AP-1, and used it in a functional screen to identify those required for OIS. This screen detected *Enh*$^{AP1\text{-}OIS1}$, an AP-1 bound enhancer that is

hyper-activated in OIS and whose abrogation results in OIS bypass. Furthermore, we identified *FOXF1* as the target gene of this enhancer, and demonstrated that it regulates the senescence phenotype.

Results

Genome-wide identification of OIS-induced enhancers

We previously carried out a CRISPR-based screen aimed at identification of p53-bound enhancers that are required for OIS(Korkmaz et al., 2016). That screen was confined to regions that are directly bound by p53 as detected by p53 ChIP-seq analysis, and therefore missed enhancers that are critical for OIS enforcement and are regulated by other TFs, in either p53-dependent or independent manner. To overcome this limitation and to globally screen DNA regulatory elements activated by OIS without being biased by preselection of a candidate TF, we first sought to comprehensively detect all the enhancers that are activated upon oncogenic stress. To this goal we utilized Global Run-On sequencing (GRO-seq), a nascent RNA detection method (Core et al., 2008), that allow robust determination of eRNA expression, as a quantitative measure for enhancer activity (Kim et al., 2010; Melo et al., 2013; de Santa et al., 2010). We used a cellular system in which the oncogene RAS^{G12V} was induced in hTERT-immortalized BJ cells (BJ-indRAS^{G12V}). As these cells contain wild-type p53, oncogenic stress results in a very potent activation of OIS and proliferation arrest (Drost et al., 2010). Exploiting bi-directional transcription as a hallmark of transcriptional activity at enhancers and promoters (Kim et al., 2010; Melgar et al., 2011; Wang et al., 2011), we detected 1,821 regulatory elements whose activity was induced in BJ-indRAS^{G12V} upon oncogene induction for 14 days (**Fig. 1A**; **Table S1**). Next, we bioinformatically searched for TFs that potentially mediate the activation of these regions by performing *de novo* DNA motif enrichment analysis. Remarkably, we found that the regulatory regions activated during the induction of OIS were significantly enriched for the binding motif of the AP1 (FOS:JUN) transcription factor (TF) (**Fig. 1B**). Overall, these results suggest an important role for AP1 in the regulation of the transcriptional response to oncogenic stress.

Therefore, we constructed a CRISPR library that systematically targets OIS-induced DNA elements that are putatively regulated by AP1, and performed a functional genetic screen to identify those that are required for OIS activation. The *de novo* motif analysis detected 762 AP1 motifs in 638 OIS-induced regulatory elements (over-representation p-value=$1.2*10^{-88}$). We examined which of these motif occurrences can be targeted by CRISPR-Cas9, given the requirement for the

presence of the NGG PAM motif near the AP-1 motif. WE required that the Cas9-madiated DNA cut will occur either within the motif itself or up to a margin of 5 nt with respect to it (**Fig. 1C**). 398 (62%) of the 638 OIS-induced regulatory elements (REs) containing AP1 motif met this criterion, with most motifs targeted by 2-3 distinct sgRNAs. Accordingly, we designed 840 sgRNAs that target AP1 motifs in 398 OIS-induced regulatory elements. We cloned these sgRNAs as a pool into pLentiCRISPRv2 vector, and generated a plasmid library referred herein as *CRISPR-AP1-EnhLib* (**Fig. 1D and Table S2**).

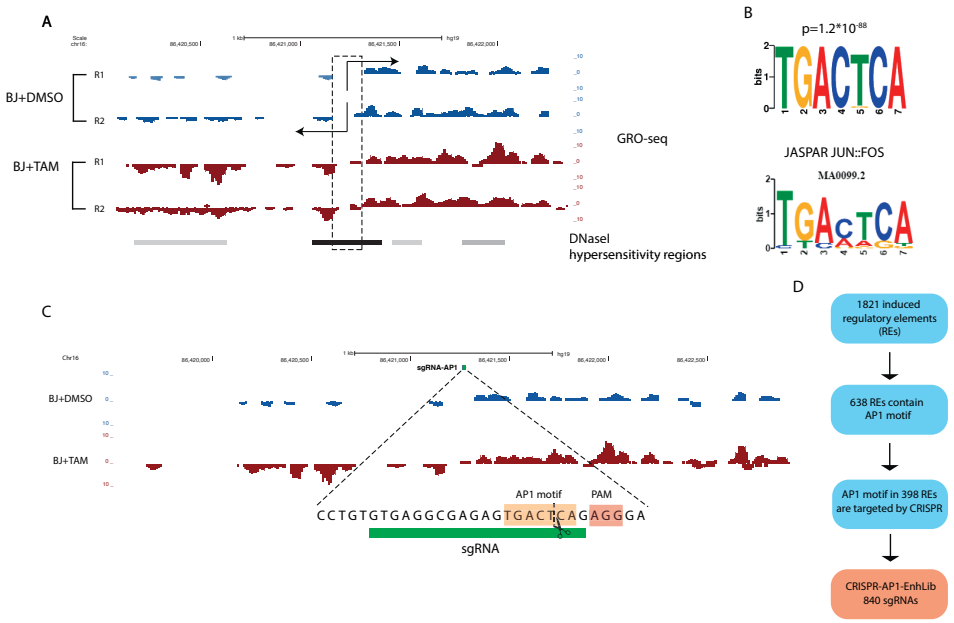


Figure 1. Design of a CRISPR screen targeting AP1 enhancers which are activated upon oncogenic stress.

A. An example of an enhancer whose activity is induced in response to oncogenic stress. Enhancer activity is inferred from the typical bi-directional transcription of eRNAs (BJ+DMSO indicates proliferating cells, and BJ+4-OHT indicates senescent cells); Genomic regions that show DNase hypersensitivity (DHS), as determined by ENCODE, are shown by the grey track). Overall, our GRO-seq analysis identified 1,821 regulatory elements (REs; enhancers or promoters) whose activity was induced in BJ cells in face of RAS activation. B. *De novo* motif analysis detected highly significant enrichment of the FOS:JUN (AP1) DNA motif in the regulatory elements that were induced upon oncogenic stress. Top – the enriched motif detected in our dataset; bottom – the AP1 motif from the JASPAR DB [49]. C. An example for occurrence of an AP1 motif within an enhancer that was induced upon oncogenic stress, that is located close enough to an NGG PAM motif, resulting in Cas9-mediated DNA cleavage that occur within the motif (Cas9 cleavage occurs ~3 nt before the PAM). Overall, we identified 398 induced REs with AP1 motif that met this requirement (Cas9 cleavage within a margin of 5 nt with respect to the motif). D. Statistical summary of the *CRISPR-AP1-EnhLib* used in our functional screen.

CRISPR screen targeting OIS-induced enhancers with AP-1 motif

We used the *CRISPR-AP1-EnhLib* library to screen for DNA elements that are putatively regulated by AP1 and are required for the activation of OIS. BJ-indRAS^{G12V} were transduced with four independent lentiviral pools of *CRISPR-AP1-EnhLib* and selected with puromycin. Then we treated the cells with 4-OHT (RAS induction) or DMSO (control) as shown in Figure 2A. Following four weeks of culturing, we harvested the cells, isolated genomic DNA, amplified integrated vectors by PCR, and used next-generation sequencing (NGS) to quantify the abundance of integrated sgRNAs present in each population. We reasoned that sgRNAs targeting regulatory elements that are required for OIS would cause bypass of senescence, sustained cell proliferation, and thus would be enriched in the cell population under oncogenic stress compared to control (**Fig. 2A**). Indeed, our screen detected several sgRNAs that were highly enriched in the OIS population (**Fig. 2B; Table S**3), among them, five showed an average enrichment fold above 1.75 over the four replicates of the screen. Notably, two of these five sgRNAs, sgRNAs-AP1^{69} and sgRNA-AP1^{71}, are independent sgRNAs that target the same enhancer region (**Fig. 2B**), hence increasing the confidence that these are true positive hits. ENCODE ChIP-seq data confirmed a strong binding of both FOS and JUN to this region (**Figure S1**). Moreover, our GRO-seq data showed ~2-fold induction of eRNA expression from this enhancer in response to oncogenic stress. Thus, we selected this regulatory region for further validation and functional characterization and named it *Enh$^{AP1\text{-}OIS1}$*.

First, using individual transductions, we validated that the introduction of sgRNAs AP1^{69} and AP1^{71} to BJ-indRAS^{G12V} cells causes a potent bypass of OIS, as judged by cell number and morphology (**Fig. 2C**). Second, we confirmed that introduction of these two sgRNAs to BJ-indRAS^{G12V} cells indeed results in an array of small deletions and insertions at the expected position within the AP1 binding motif in the targeted enhancer (**Figure S2**). Third, following induction of oncogenic stress, sgRNAs AP1^{69} and AP1^{71} transduced BJ-indRAS^{G12V} cells showed a significant reduction in senescent-associated-β-Gal (SA-β-Gal) staining (**Fig. 2D, Figure S3**) and an elevated BrdU staining (**Fig. 2E, Figure S3**), indicative of attenuated activation of cellular senescence and of sustained cellular proliferation compared to the NT control. As expected, while the effect caused by these two sgRNAs was highly significant, it was not as strong as the effect elicited by targeting p53 (**Fig. 2D-E**). Last, we examined the activity of *Enh$^{AP1\text{-}OIS1}$* following the transduction of sgRNAs AP1^{69} and AP1^{71} by measuring eRNA expression at the *Enh$^{AP1\text{-}OIS1}$* locus. As expected, targeting *Enh$^{AP1\text{-}OIS1}$* by these two sgRNAs significantly compromised its activity during OIS compared to the control sgRNA (**Fig. 2F**).

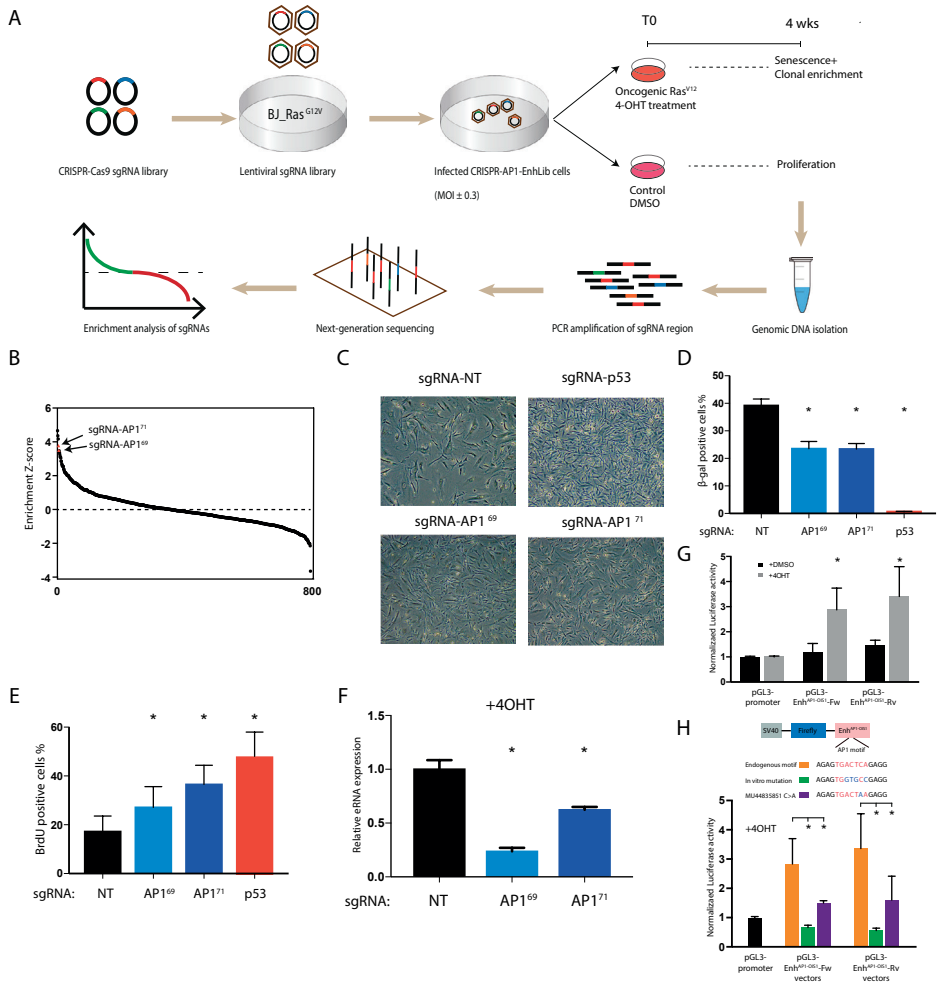


Figure 2. Functional CRISPR screen discover a novel enhancer required for OIS.

A. Schematic representation of the set-up of our functional screen. B. Results of the CRISPR screen. sgRNAs are sorted by the enrichment score based on the ratio between their prevalence in the BJ+4-OHT and BJ +DMSO control populations (measured 4 weeks after 4-OHT treatment). Y-axis shows Z scores of the mean sgRNA enrichment scores (calculated over the four replicates of the screen). Coloured in red are two sgRNAs, sgRNA-AP1^{69} and sgRNA-AP1^{71}, that target the same enhancer, called here *Enh$^{AP1\text{-}OIS1}$* C. Individual transductions of sgRNA- AP1^{69} and sgRNA-AP1^{71} validated that they cause OIS bypass. sgRNA targeting p53 was used as a positive control and a non-targeting (NT) sgRNA was used as a negative control. D. Targeting *Enh$^{AP1\text{-}OIS1}$* by either sgRNA- AP1^{69} or sgRNA-AP1^{71} caused OIS bypass as measured by β-gal staining, a canonical mark for senescence (p53ko used as a positive control). Data shown represents mean (SD), n=4. *p<0.05. E. Targeting *Enh$^{AP1\text{-}OIS1}$* by either sgRNA- AP1^{69} or sgRNA-AP1^{71} resulted in enhanced proliferation as measured by BrdU staining (p53ko used as a positive control).

Data shown represents mean (SD), n=4. *p<0.05. F. Measurement of eRNA production at *Enh*$^{AP1\text{-}OIS1}$ in cells with the indicated sgRNAs. eRNA levels are significantly decreased upon mutagenesis of the AP1 binding site caused by either sgRNA- AP1[69] or sgRNA-AP1[71]. Data shown represents mean (SD), n=3. *p<0.05. G. BJ-indRAS^{G12V} cells were transfected with the indicated plasmids, and treated with DMSO or 4-OHT for 72h. pGL3 constructs contain firefly luciferase reporter gene with the corresponding enhancer (none for pGL3-promoter, two different orientations for *Enh*$^{AP1\text{-}OIS1}$). Relative luciferase activity is calculated by dividing the firefly luciferase activity to that of Renilla luciferase. Normalized luciferase activity is calculated by dividing the relative luciferase activity to that of pGL3-promoter for each condition. Data shown represents mean (SD), n=6. *p<0.05. H. BJ-indRAS^{G12V} cells were transfected with the indicated enhancer constructs. Endogenous motif represents the original sequence of *Enh*$^{AP1\text{-}OIS1}$, in vitro mutation construct represents mutagenesis of the AP1 consensus motif, and MU44835851 represents mutant construct bearing a C>A mutation as indicated. The cells were treated with 4-OHT for 48h prior transfections. Data shown represents mean (SD), n=3. *p<0.05.

Next, we carried out *in vitro* reporter assays to verify that *Enh*$^{AP1\text{-}OIS1}$ functions as enhancer and promotes transcription of target genes. We cloned *Enh*$^{AP1\text{-}OIS1}$ downstream of the Firefly luciferase gene in two orientations in pGL3-promoter vector followed by transfection into BJ-indRAS^{G12V} cells. While we did not observe a noticeable elevation of luciferase activity in cells treated with DMSO, there was a significant increase of 3-fold in cells treated with 4-OHT **(Fig. 2G)**. To verify that mutations in the AP1 binding site disrupts the enhancer activity as we observed in sgRNA-AP1[69/71] cells, we mutated the AP1 motif of *Enh*$^{AP1\text{-}OIS1}$ and examined the effect on the enhancer activity, in condition of oncogenic stress. Indeed, the mutations completely abolished the ability of *Enh*$^{AP1\text{-}OIS1}$ to stimulate luciferase expression (**Fig. 2H**). In addition, we searched for tumor somatic mutations within the AP1 binding motif (using ICGC data), and found one case in a lung cancer patient (**Figure S4**). The mutation is a C to A substitution within the consensus motif of AP1, and located at the cut site of sgRNA-AP1[69]. To examine if this somatic mutation disrupts the activity of *Enh*$^{AP1\text{-}OIS1}$, we performed mutagenesis of *Enh*$^{AP1\text{-}OIS1}$ on the reporter construct with a single nucleotide substitution. Remarkably, we observed a 50% reduction of the enhanced luciferase activity (**Fig. 2H**), suggesting an important role of the specified nucleotide in determining binding affinity of AP1 to this enhancer. Taken together, our functional genetic screen and subsequent focused experiments have identified and validated a novel AP1-bound enhancer whose activity is required for proper induction of OIS.

FoxF1 is a target gene of *Enh*$^{AP1\text{-}OIS1}$

Next, we set up experiments to elucidate the mode of action by which *Enh*$^{AP1\text{-}OIS1}$ is required for OIS. Enhancers regulate gene expression of cis-located target genes that can reside hundreds of kbp away. Examination of our GRO-seq data indicated that *FOXF1*, the nearest gene to the *Enh*$^{AP1\text{-}OIS1}$ locus (located >100 kbp downstream

of it), was ~2 fold induced following oncogenic stress (**Fig. 3A**), suggesting a potential functional connection. No other gene in a 1 Mbp distance from *Enh*$^{AP1\text{-}OIS1}$ showed such a strong effect. Furthermore, we performed RNA-seq with cells transduced with sgRNA-AP1[69] and sgRNA-AP1[71], and observed a significant reduction in the expression of FOXF1 (**Figure S5**). We validated this result using qRT-PCR analysis, which also indicated that targeting *Enh*$^{AP1\text{-}OIS1}$ by either sgRNA-AP1[69] or sgRNA-AP1[71] results in a significant reduction in the expression level of FOXF1 under OIS conditions (**Fig. 3B**). Western blotting analysis confirmed this result at the protein level, in addition to confirming that FOXF1 expression is increased following oncogenic stress **(Fig. 3C)**. Publicly available RNA Pol II ChIA-PET data (from Hela cells) indicate physical interaction between *Enh*$^{AP1\text{-}OIS1}$ and the 3' region of *FOXF1* (**Figure S6**), which was confirmed in the chromatin conformation capture (3C) experiments of senescent BJ cells (**Figure S7**). More importantly, the physical interactions between *Enh*$^{AP1\text{-}OIS1}$ and the promoter region of *FOXF1* were significantly stronger (**Figure S7**), suggesting a robust transcriptional regulation of *Enh*$^{AP1\text{-}OIS1}$. Collectively, these results strongly point to *FOXF1* as the target gene of *Enh*$^{AP1\text{-}OIS1}$.

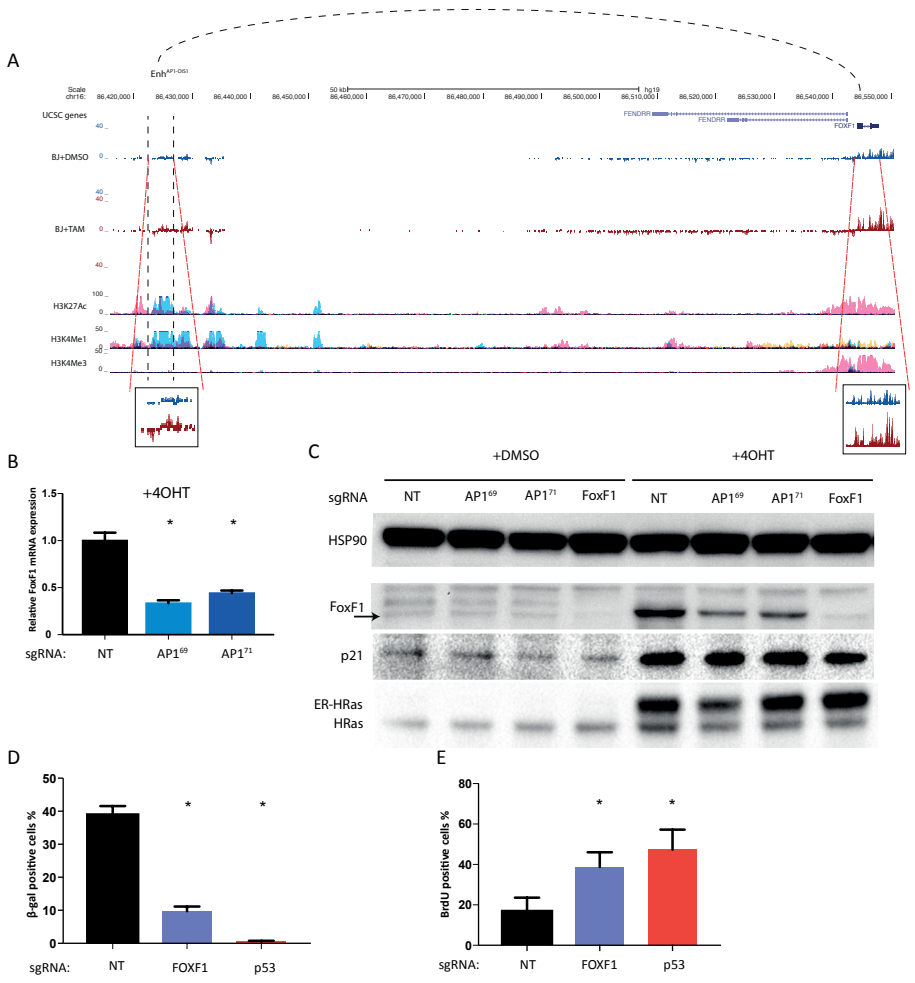
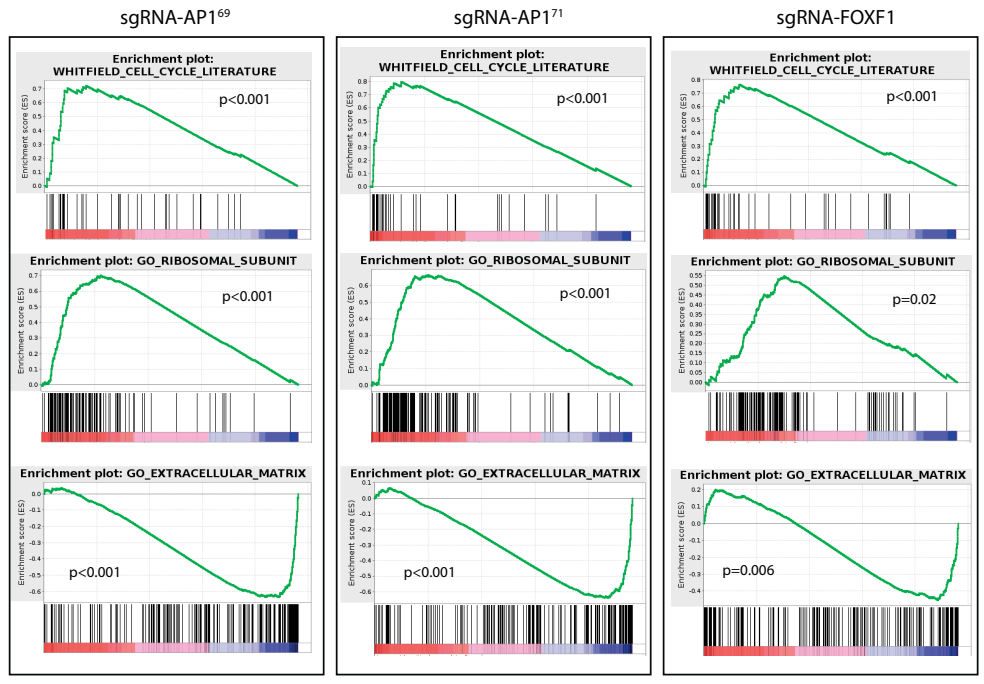


Figure 3. FOXF1 is the target gene regulated by *Enh*$^{AP1\text{-}OIS1}$.

A. UCSC screenshot of GRO-seq analysis of BJ-indRAS^{G12V} cells. BJ cells were treated with DMSO or 4-OHT for 14 days. Bidirectional transcription is represented by using positive and negative values for expression in the Crick and Watson strands, respectively. The genomic regions of *Enh*$^{AP1\text{-}OIS1}$ and *FOXF1* are enlarged. Note the enhancement in GRO-seq signal for both *Enh*$^{AP1\text{-}OIS1}$ and *FOXF1* in BJ+4-OHT (brown track) compared to BJ+DMSO (blue track). B. mRNA levels of FOXF1 are reduced in sgRNA-AP1^{69} and sgRNA-AP1^{71} targeted cells under 4-OHT treatment. Data shown represents mean (SD), n=3. *p<0.05. C. BJ-indRAS^{G12V} cells transduced with the specified sgRNAs were treated with DMSO or 4-OHT for 14 days, and FOXF1, p21, and HRas protein levels were measured by western blot. HSP90 was used as the loading control. The band of FOXF1 is marked with an arrow. ER-HRas indicates the induced version of HRas. D. Targeting the *FOXF1* and *p53* genes caused OIS bypass as measured by β-gal staining. Note the stronger effect of FOXF1ko compared to the effect elicited by targeting *Enh*$^{AP1\text{-}OIS1}$ (Fig. 2D). Data shown represents mean (SD), n=4. *p<0.05. E. Targeting *FOXF1* and *p53 gene* resulted in enhanced proliferation as measured by BrdU staining. Data shown represents mean (SD), n=4. *p<0.05.

Loss of FOXF1 causes senescence bypass and abolishes senescence expression signatures

To further establish *FOXF1* as the target gene that links *Enh*$^{AP1\text{-}OIS1}$ to senescence, we examined the phenotypic effect of knocking out *FOXF1*. Indeed, targeting *FOXF1* results in a strong senescence bypass phenotype, as evident by significant reduction in SA-β-Gal staining (**Fig. 3D**) and elevated BrdU staining (**Fig. 3E**), similar to the effect elicited by targeting *Enh*$^{AP1\text{-}OIS1}$ (**Fig. 2E-F**). Effective *FOXF1* knockout was confirmed by western blotting analysis (**Fig. 3C**). Last, we used RNA-seq to globally



Shared cell-cycle genes
CCNA2, CCNB1, CCNB2, CDC20, CDC25A, CDC25C, CDC45, CDC6, CDK1, CDKN2C, CDKN3, CENPA, CKS2, DHFR, E2F1, KIF20A, MCM2, MCM6, PLK1

Shared ribosomal-protein genes (selected)
MRPL13, MRPL15, MRPL2, MRPL21, MRPL3, MRPL33, RPL12, RPL29, RPL3, RPL30, RPL31, RPL9, RPS15, RPS16, RPS19, RPS2, RPS20, RPS21, RPS23, RPS27, RPS5, RPS7, RPS9

Shared ECM genes (selected)
AGRN, COL12A1, COL18A1, COL27A1, COL4A1, COL4A2, COL5A1, COL5A2, COL6A3, COL7A1, FBLN1, FBLN5, FBN1, FBN2, FN1, HSPG2, ITGA6, ITGB1, LAMC1

Figure 4. *Enh*$^{AP1\text{-}OIS1}$ and *FOXF1* knockouts display expression profiles of senescence bypass.

GSEA analysis of expression profiles measured in 4OHT-treated BJ-indRAS^{G12V} cells targeted by sgRNA-AP1^{69}, sgRNA-AP1^{71}, or sgRNA-FOXF1 compared to the profile of control 4OH-treated cells transduced with non-targeting sgRNA. A list of shared genes within each group is shown.

compare expression profiles in BJ-indRAS^{G12V} cells transduced with either sgRNAs targeting *Enh$^{AP1\text{-}OIS1}$* (sgRNA-AP1[69] or sgRNA-AP1[71]), sgRNA targeting *FOXF1*, or a control non-targeting sgRNA. GSEA analysis (Subramanian et al., 2005) for functional characterization of the biological processes affected by these genetic manipulations showed that cell-cycle genes and genes encoding ribosomal proteins are significantly up-regulated when targeting either *Enh$^{AP1\text{-}OIS1}$* or *FOXF1*, reflecting the bypass of OIS and the subsequent enhanced proliferation experienced by these cells following oncogene hyperactivity (**Fig. 4**). Conversely, the induction of various extra cellular matrix (ECM) components that is exhibited in OIS was largely attenuated in cells with *Enh$^{AP1\text{-}OIS1}$* or *FOXF1* knockouts (**Fig. 4**). Taken together, our results strongly indicate that *Enh$^{AP1\text{-}OIS1}$* controls OIS through the regulation of *FOXF1* expression (**Fig. 5**).

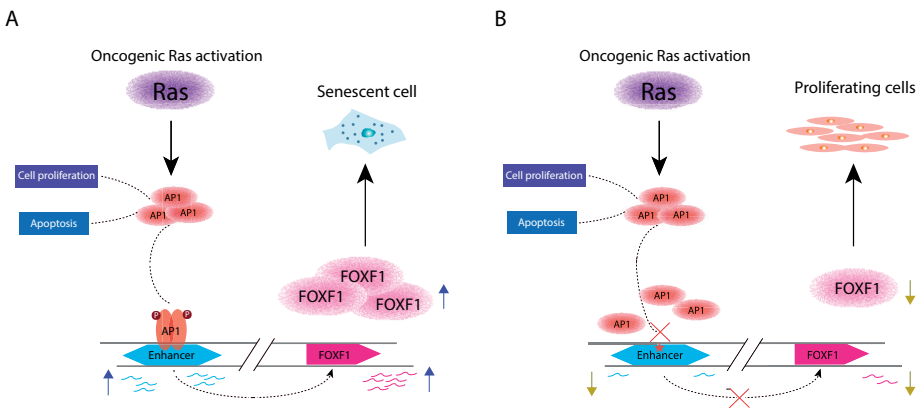


Figure 5. Model of *Enh$^{AP1\text{-}OIS1}$* regulation of OIS.

A. In normal BJ fibroblast cells, hyper-activation of RAS induces MAPK signaling cascade, including AP-1 TFs. Activated AP-1 TFs control different cellular functions, including cell proliferation and apoptosis. AP1 is recruited, among other enhancers, to *Enh$^{AP1\text{-}OIS1}$* and stimulates its activity. This in turn promotes the expression of the target gene FOXF1, diverting oncogenic signals into pre-senescent pathway. B. Mutagenesis of the AP-1 binding site in *Enh$^{AP1\text{-}OIS1}$* abrogates its enhancer activity and thus leading to decreased expression of FOXF1. This results in compromised induction of OIS and thus cells continue uncontrolled cell proliferation.

Discussion

In this study, we first found that OIS-induced enhancers are enriched for the binding motif of AP1. Based on this finding, we perform a CRISPR screen focused on AP1 motifs within enhancers that are activated upon oncogenic stress. We discovered a novel AP-1 bound enhancer, *Enh$^{AP1\text{-}OIS1}$*, that is required for establishment of OIS

and identified FOXF1 as the target that mediates this role. We propose a new role of AP1 in senescence via activation of FOXF1, providing an additional regulation of cell proliferation during senescence.

AP1 TFs are recruited to enhancer regions to drive oncogenic growth (Zanconato et al., 2015) and are broadly required for enhancer selection(Vierbuchen et al., 2017), suggesting a possible role of AP1 at enhancers. As a downstream target of RAS signalling pathway, AP1 is activated to target mitogen-responsive genes (Deng and Karin, 1994; Kampfer et al., 1998). Earlier studies have shown that mRNA level and activity of AP1 genes are attenuated upon entering replicative senescence (Irving et al., 1992; Seshadri and Campisi, 1990). Altered AP1 activity is mainly due to loss of c-FOS expression and maintained JUN proteins, thus promoting JUN-JUN homodimers instead of FOS-JUN heterodimers (Irving et al., 1992; Riabowol et al., 1992). This suggests that loss of AP1 activity is possibly responsible for the irreversible growth arrest in senescent cells. Conversely, overexpression of c-FOS with increased AP1 activity is not sufficient to initiate DNA synthesis in senescent human fibroblasts (Rose et al., 1992). Therefore, AP1 is likely not the key factor that regulates senescence, but rather a downstream factor that fine-tune the senescence program under replicative stress (e.g. H-RAS activation). In addition, previous functional genetic screens did not indicate any of the AP1 family members as critical factors in OIS(Burrows et al., 2010; Drost et al., 2010). Supporting this conclusion, CRISPR-mediated KOs of c-FOS and c-JUN did not result in any obvious bypass of OIS (**Figure S8**). However, it is possible that one or few targets of AP1 mediate OIS while others antagonize it, or are required for cell survival.

FOXF1 belongs to the Forkhead family of transcription factors. FOXA1 has been reported to promote senescence via activation of p16^{INK4a} (Li et al., 2013), and FOXO4 inhibition induces p53 nuclear exclusion, which results in apoptosis of senescent cells (Baar et al., 2017). The functions of FOXF1 remain to be determined, yet recent studies have implicated its role in lung regeneration by targeting genes of extracellular matrix and cell cycle progression (Bolte et al., 2017), as well as promoting prostate cancer growth via MAPK pathway (Fulford et al., 2016). To date, there has been no evidences of any connections between AP1, FOXF1 and OIS, possibly due to regulation via enhancers, rather than proximal promoters, as proposed in this study. It has been proposed that FOXF1 is a target gene of p53, which regulates cell migration and invasion (Tamura et al., 2014); and that FOXF1 is a potential oncogene, which promotes rhabdomyosarcoma by repressing p21^{Cip1} (Milewski et al., 2017). Here we provide the first evidence suggesting that FOXF1 is a potential tumour suppressor, regulating senescence in human cells. In addition, we generated double knockout cell lines (NT+p53 ko, sgRNA-AP1[69]+ p53 ko, sgRNA-

AP1[71]+ p53 ko, and FOXF1 ko + p53 ko), and we observed a strong senescence bypass phenotype (**Figure S9A-B**). Interestingly, we found an additive effect of proliferation in the EnhAP1-OIS1 ko and FOXF1 ko cell lines (sgRNA-AP1[69]+ p53 ko, sgRNA-AP1[71]+ p53 ko, and FOXF1 ko + p53 ko) compared with NT + p53 ko cells (**Figure S9B**). This suggests that FOXF1 regulates OIS in a p53 independent manner. In parallel with the canonical p53 pathway, FOXF1 regulates the expression of a subset of cell cycle and ribosomal genes (**Fig. 4**). Further studies should explore the exact function of FOXF1 in regulating senescence.

We propose a model in which following oncogenic induction, AP1 TFs are activated to promote cellular proliferation in response to stimuli. However, under excessive exposure to RAS, AP1 is recruited to *Enh*$^{AP1\text{-}OIS1}$ to promote the expression of FOXF1 to drive cells into senescence. Disruption of the AP1 binding site within *Enh*$^{AP1\text{-}OIS1}$ results in attenuated activation of FOXF1, hampering full execution of the OIS program. We attempted to generate a FOXF1 overexpression cell line while targeting *Enh*$^{AP1\text{-}OIS1}$, to rescue the senescence phenotype. However, this was not successful, possibly due to the intolerance of the cells under ectopic FOXF1 expression. This result show that although AP1 is activated by MAPK pathway, and stimulates the expression of many cell cycle genes, upon oncogenic stress it also mediates tumor suppressive effects.

Our current knowledge on cancer driver non-coding somatic mutations (SMs) is still very rudimentary, yet several studies suggested that the role of such SMs is underappreciated(Melton et al., 2015; Weinhold et al., 2014). Genome-wide analysis has revealed AP1 as a key factor at regulatory elements in cancers(Davie et al., 2015), and AP1 binding sites are frequently mutated in various cancer types (Kaiser et al., 2016). Analysing ICGC data, we found a somatic mutation within the AP1 motif of *Enh*$^{AP1\text{-}OIS1}$ in a lung cancer patient, and validated its functional effect by using in vitro reporter assays (**Fig. 2H**). This suggests a possible cancer driver effect for somatic mutations in AP1 binding motifs. Our study further demonstrates the power of CRISPR-based screens in exploring the function of the noncoding genome.

Conclusions

Our study provides evidence that AP1 TFs are broadly stimulated during OIS and are localized to enhancer regions to activate specific gene programs. We show that AP1 controls senescence program via *Enh*$^{AP1\text{-}OIS1}$ and its target gene *FOXF1*.

We propose that AP1 is a double-edged sword in regulating cell proliferation and senescence, providing a restrictive feedback on unlimited cell proliferation.

Methods

Cell culture

BJ/ET/Ras^{V12}and HEK293-T cells were cultured in DMEM medium (Gibco), supplemented with 1% penicillin/streptomycin (Gibco) and 10% FCS (Hyclone). To induce OIS, BJ cells were treated with 100 nM 4-OHT (Sigma) for 14days.

Analysis of GRO-seq data

GRO-seq was applied to control and RAS^{G12V}-induced hTERT immortalized BJ cells (14 days after RAS induction). These conditions were probed using biological duplicates. Sequenced reads were aligned to the human genome (hg19) using bowtie2 (Langmead and Salzberg, 2012). Transcriptional units (TUs) were inferred from the GRO-seq data using HOMER (Heinz et al., 2010). Read counts per TU were calculated using HTseq-count (Anders et al., 2015). 76,200 TUs covered by at least 20 reads in at least one sample were detected. TU expression levels were than normalized using quantile normalization to allow comparison between samples, and fold-change (FC; presented in log2 base) were calculated between the RAS-induced and control samples. To avoid inflation of high FC value for lowly expressed TUs, we set a floor value of 10 (that is, all expression levels below 10 were set to 10). Next, we defined bi-directional TUs as TUs whose start site is separated by no more than 800 bp and are transcribed on opposite strands (TU+ and TU-). As bi-directional transcription is a hallmark of transcriptional regulatory elements, we refer to these loci as regulatory elements (REs). Overall, this analysis defined 36,497 regulatory elements. Last, a regulatory element (bi-directional TU) was defined as OIS-induced if the expression level of both its mates was elevated by at least 2-folds upon RAS induction, in both duplicates. In total, 1,821 OIS-induced REs were identified in our dataset (Table S1).

Motif enrichment analysis

The sequences of the OIS-induced regulatory elements were searched for statistically over-represented TF binding motifs. We performed this *de novo* motif analysis using DREME (Bailey, 2011). For each bi-directional TUs, we scanned the region between the start site of the opposite mates (TU+ and TU-) plus a margin of 200 bp to each direction. As control sequences, we extracted adjacent sequences of the same length are immediately upstream and downstream the test sequence. The binding motif of JUN/FOS was highly enriched ($p=1.2*10^{-88}$) on the OIS-induced

REs. Specific occurrences of the enriched JUN/FOS motif were identified using FIMO with default parameters (Grant et al., 2011). Overall, 762 JUN/FOS motif occurrences were found on 638 OIS-induced REs.

CRISPR library construction and analysis

We designed a CRISPR library to target the FOS/JUN motifs in the OIS-induced REs. For 398 of the 638 OIS-induced REs with FOS/JUN motif, we found an occurrence of the NGG PAM in a location that is expected to induce a Cas9 DNA cleavage within a margin of 5 bp with respect to the motif (that is, the cut is expected to occur within the motif or up to 5 bp from its edges). Overall, we designed 840 distinct sgRNAs, collectively targeting FOS/JUN motif in 398 OIS-induced REs. We cloned these sgRNAs into pLentiCRISPRv2 vector and generated a plasmid library (which we call *CRISPR-AP1-EnhLib*). Induced and control BJ-indRAS^{G12V} were transduced with four independent lentiviral pools of *CRISPR-AP1-EnhLib*. Following four weeks of culturing, we harvested library-transduced cells, isolated genomic DNA, amplified integrated vectors by PCR, and used next-generation sequencing (NGS) to quantify the abundance of integrated sgRNAs present in each population. Read counts were normalized to 1M reads and enrichment ratio (fold-change in log2) were calculated for each sgRNA between the induced and control samples per replicate. (To avoid inflation of FC for sgRNAs covered by low number of reads, counts bellow 50 were set to 50). Next, average enrichment factor was calculated per sgRNA over the four replicates and was transformed to a Z score (**Fig. 2B, Table S3**).

Senescence-associated b-galactosidase assay

BJ cells were transduced with different sgRNA constructs and selected with puromycin. After selection, cells were seeded in triplicate in 6-well plates and treated with 100 nM 4-OHT for 14 days. β-galactosidase measurement was performed by following the protocol of the Senescence β-Galactosidase Staining Kit (Cell Signaling), and at least 1000 cells were analyzed for each condition.

BrdU proliferation assay

BJ cells were seeded in 6-well plates on day 1. Next morning, cells were incubated in fresh medium for 3 h with 30 μM bromodeoxyuridine (BrdU, Sigma) followed by two times PBS wash and then fixed with 4% formaldehyde. Cells were washed two times with PBS and treated with 5M HCl/0.5% Triton to denature DNA. Cells were neutralized with 0.1M $Na_2B_4O_7$. The cells were then treated with blocking buffer (3% BSA in 0.5% Tween PBS) for 30min and incubated with anti-BrdU antibody (Dako) with blocking buffer for 2 hours at room temperature. Cells were washed with PBS three times, and finally incubated with FITC-conjugated anti-mouse Alexa Fluor 488 secondary antibody (Dako) in blocking buffer for 1 hour, washed three

times, stained with propidium iodide for 30min. BrdU incorporation was measured by immunofluorescence (at least 1000 cells were scored for each condition). The numbers of individual nucleus and BrdU-stained nucleus were counted using imageJ software.

Luciferase reporter assay

The constructs with the enhancers were cloned based on pGL3-promoter (Promega) vector. The enhancer region was PCR amplified from BJ genomic DNA and inserted downstream of the firefly luciferase reporter gene. The transfection was performed by seeding 1x10^5 of cultured cells in 6-well plates. The next day, 500ng of each construct (pGL3-promoter, pGL3-Enh$^{AP1\text{-}OIS1}$-Fw, and pGL3-Enh$^{AP1\text{-}OIS1}$-Rv) were co-transfected with 50ng of Renilla luciferase reporter construct using Fugene-6 (Promega) following manufacturer's protocol. Luciferase reporter assay was performed 24h post transfection using Dual-Luciferase Reporter assay kit (Promega). Cells were lysed directly on the plate with passive lysis buffer for 15min at room temperature. Firefly and Renilla luciferase activity was measured with the substrates from the kit using Centro XS3 LB960 machine (Berthold technologies). For BJ-indRAS^{G12V}, cells were pre-treated with 100nM 4-OHT for 48h prior transfection. For HCT116, cells were treated with UV-C (50J/m2) or MG132 (5μM) 18h after transfection. The luciferase assay was performed 5h post treatment.

Mutagenesis of *Enh$^{AP1\text{-}OIS1}$*

Mutations of *Enh$^{AP1\text{-}OIS1}$* were performed using QuikChange Lightning site-directed mutagenesis kit (Agilent) according to the manufacture's manual. Briefly, primers for mutagenesis were designed using the online tool from Agilent. pGL3-Enh$^{AP1\text{-}OIS1}$-Fw and pGL3-Enh$^{AP1\text{-}OIS1}$-Rv were PCR amplified and transformed into DH5α bacteria. Single colonies from each mutant were sequence verified and used for transfection.

RNA isolation, reverse transcription and qRT-PCR

Total RNA was extracted by using TRIsure (Bioline) reagent and following the manufacturer's protocol. Reverse transcription was done with SuperScript III (Invitrogen) using 1 μg of total RNA per reaction. qRT-PCR reaction was performed using SensiFAST SYBR No-ROX Kit (Bioline) in LightCycler 480 (Roche). Primers used are listed in Table S4.

Western blot

1x10^6 Cells were seeded in 10cm dish and treated with DMSO or 4-OHT for 14 days. Cells were trypsinized and cell pellets were lysed with RIPA buffer supplemented

with 1x cOmplete protease inhibitor cocktail (Roche) following the manufacturer's protocol. Protein concentrations were determined using Pierce BCA protein assay kit (Thermo Scientific). Lysates were separated on SDS-PAGE gels and transferred Membranes were immunoblotted with the following antibodies: CDKN1A (Sc-397, Santa Cruz; 1: 1,000), HRAS (C-20, Santa Cruz; 1: 1,000), FoxF1 (ab168383, Abcam, 1:1000), HSP90 (610418, BD biosciences, 1:3000). Protein bands were visualized using corresponding secondary antibodies (Dako) and ECL reagent (GE Healthcare).

Lentiviruses production and infection

HEK293T cells were seeded at the density of 5 X 10^6 cells per 10cm dish one day prior transfection. Transfection was performed using PEI (Polyethylenimine, Polysciences) and medium was refreshed after 16h. Virus-containing supernatant was collected 48h post transfection by filtering through a 0.45 µm membrane (Milipore Steriflip HV/PVDF) and snap-frozen, stored at −80°C. BJ cells were infected and selected with the proper antibiotics 48 h after transduction for at least 4 days until no surviving cells remained in the no-transduction control plate.

Chromatin Conformation Capture (3C) analysis

10 x 10^6 cells were harvested in PBS for each 3C sample. Cells were centrifuged at 300xg for 5min at RT, and resuspended PBS / 10% FBS. Then cells were incubated with equal volume of 4% formaldehyde (2% end concentration) for 10min and quenched with 2M glycine solution (0.2M end concentration), followed by centrifugation at 300xg for 5min at 4°C. Cell pellet was then resuspended in PBS / 10% PBS and centrifuged at 300xg for 5min at 4°C. The supernatant was then discarded and snap-frozen, stored at -80°C. The cell pellet was lysed in 3mL lysis buffer (50mM Tris -HCl pH 7.5, 0.5% NP-40, 1% Triton X-100, 150mM NaCl, 5mM EDTA, protease inhibitor cocktail (Roche)) for 1.5h at 4°C, followed by centrifugation at 1000xg for 3min. The pellet was washed once in 1.2x restriction buffer and resuspended again in 500µL of 1.2x restriction buffer. 15µL of 10% SDS was added to the suspension and incubated at 37°C while shaking at 400rpm. 75µL of 20% Triton X-100 was added to the suspension and incubated at 37°C while shaking at 400rpm. The samples were then centrifuged at 1000 x g for 3min and resuspended in 500µL of 1x restriction buffer. The digestion was performed with addition of 200U of Csp6I (Thermo Fisher Scientific) at 37°C overnight. The digestion efficiency was assessed the next day on agarose gel. The enzyme was then inactivated at 65°C for 20min and then samples were centrifuged at 1000 x g for 3min to remove the restriction buffer. The pellet was resuspended in 7mL of 1x ligation buffer, and the ligation was performed with addition of 50U of T4 DNA ligase at 16°C overnight. Again, the ligation efficiency was examined on agarose gel. De-crosslinking was performed by addition of 30µL of protease K (Roche) at 65°C overnight. To remove residual

RNA, 15µL of RNaseA cocktail (Ambion) was added to the samples and incubated at 37°C for 45min. DNA was recovered by adding 7mL of isopropanol and 70µL of NucleoMag 96 PCR beads (Bioke) and incubated for 30min at room temperature. The samples were centrifuged for 3min at 1000 x g and washed with 80% ethanol twice. Finally, the beads were dried and eluted in 300µL of 10mM Tris-HCl pH 7.5. To assess the physical interactions between $Enh^{AP1\text{-}OIS1}$ and target regions, we designed a constant primer (C1) that amplify the $Enh^{AP1\text{-}OIS1}$ region overlapping the junction created by Csp6I enzyme. For each assessed region, we designed two primers (reverse and forward) to examine the interactions with $Enh^{AP1\text{-}OIS1}$. The first PCR was performed with primer C1 and each candidate primer for 25 cycles. Afterwards, a nested PCR was performed using a second constant primer (C2) and each candidate primer for another 18 cycles. Finally, PCR products were resolved on 2% agarose gel. To assess the primer efficiency, we PCR amplified the genomic regions of $Enh^{AP1\text{-}OIS1}$ and FOXF1, and mixed equal molar of each fragments as the template. The template DNA was then digested and ligated as mentioned. Finally, the quantifications were normalized with the primer efficiencies. To examine the sequences of the PCR products, DNA bands were cut, isolated, and sanger-sequenced.

Mutation analysis of enhancer regions

Genomic DNA of the cells transduced with sgRNAs were isolated and quantified. 500ng of the genomic DNA was used for PCR to amplify the enhancer region. We performed a two-step PCR by introducing the P5 adapter sequences in the first PCR and P7 adapters with the indexes in the second PCR. After the second PCR, the libraries were purified with CleanPCR beads (CleanNA) and quantified on 2100 Bioanalyzer using a 7500 chip (Agilent). Equimolar of each sample was taken for the final library. Libraries were sequenced using the Mi-Seq platform. Sequenced reads were aligned to the amplified enhancer region using bowtie. Bam files were analysed to count the number of mutations (mismatches, insertions or deletions) identified at each location in that region.

RNA-seq library construction

Total RNA was isolated using Trisure reagent (Bioline) following the manufacturer's protocol. Briefly, cells were lysed in Trisure, precipitated with isopropanol, and dissolved in RNase-free water. To generate strand-specific libraries, we used the TruSeq Stranded mRNA sample preparation kit (Illumina) following the manufacturer's instructions. Briefly, 1000ng of total RNA was polyA-enriched using oligo-dT beads, and the RNA was fragmented, random primed and reverse transcribed using SuperScript II Reverse Transcriptase (Invitrogen). Second strand cDNA was then synthesized, 3'-adenylated and ligated to Illumina sequencing

adapters and subsequently amplified by 12 cycles of PCR. The sequencing libraries were analyzed on a 2100 Bioanalyzer using a 7500 chip (Agilent), and pooled equimolar into a 10nM multiplex sequencing pool.

Sequencing

Sequencing of the CRISPR screen and RNA-seq was done using single reads of 65bp on the Hi-Seq2500 platform (Illumina). Mutation analysis of enhancer regions were performed with single reads of 150bp on the Mi-Seq system with Mi-Seq reagent v2 Nano kit.

RNA-seq analysis

Gene expression profiles were recorded in BJ-indRAS^{G12V} (14 days after RAS induction by 4-OHT treatment) transduced with CRISPR vectors that either targeted the *Enh*$^{AP1\text{-}OIS1}$ using sgRNA-AP1[69], sgRNA-AP1[71], targeted *FOXF1* itself or transduced with a control non-targeting sgRNA (sgRNA-NT). Sequenced reads were aligned to the human genome (hg19) using TopHat2 (Kim et al., 2013). Number of reads mapped to each annotated gene was counted using HTseq-count (Anders et al., 2015), and then converted to RPKMs (using GENCODE v25 annotations). RPKM levels were further normalized using quantile normalization and expression levels in each sample relative to the control non-targeting sample were calculated (in log2 base). Biological pathways and processes affected by targeting the *Enh*$^{AP1\text{-}OIS1}$ or *FOXF1* were sought using gene-set enrichment analysis (GSEA) (Subramanian et al., 2005).

References

Amano, T., Sagai, T., Tanabe, H., Mizushina, Y., Nakazawa, H., and Shiroishi, T. (2009). Chromosomal Dynamics at the Shh Locus: Limb Bud-Specific Differential Regulation of Competence and Active Transcription. Dev. Cell *16*, 47–57.

Anders, S., Pyl, P.T., and Huber, W. (2015). HTSeq-A Python framework to work with high-throughput sequencing data. Bioinformatics *31*, 166–169.

Andersson, R., Gebhard, C., Miguel-Escalada, I., Hoof, I., Bornholdt, J., Boyd, M., Chen, Y., Zhao, X., Schmidl, C., Suzuki, T., et al. (2014). An atlas of active enhancers across human cell types and tissues. Nature *507*, 455–461.

Baar, M.P., Brandt, R.M.C., Putavet, D.A., Klein, J.D.D., Derks, K.W.J., Bourgeois, B.R.M., Stryeck, S., Rijksen, Y., van Willigenburg, H., Feijtel, D.A., et al. (2017). Targeted Apoptosis of Senescent Cells Restores Tissue Homeostasis in Response to Chemotoxicity and Aging. Cell *169*, 132–147.e16.

Bailey, T.L. (2011). DREME: Motif discovery in transcription factor ChIP-seq data. Bioinformatics *27*, 1653–1659.

Banerji, J., Rusconi, S., and Schaffner, W. (1981). Expression of a β-globin gene is enhanced by remote SV40 DNA sequences. Cell *27*, 299–308.

Bolte, C., Flood, H.M., Ren, X., Jagannathan, S., Barski, A., Kalin, T. V., and Kalinichenko, V. V. (2017). FOXF1 transcription factor promotes lung regeneration after partial pneumonectomy. Sci. Rep. *7*.

Braig, M., Lee, S., Loddenkemper, C., Rudolph, C., Peters, A.H.F.M., Schlegelberger, B., Stein, H., Dörken, B., Jenuwein, T., and Schmitt, C.A. (2005). Oncogene-induced senescence as an initial barrier in lymphoma development. Nature *436*, 660–665.

Bulger, M., and Groudine, M. (2011). Functional and mechanistic diversity of distal transcription enhancers. Cell *144*, 327–339.

Burrows, A.E., Smogorzewska, A., and Elledge, S.J. (2010). Polybromo-associated BRG1-associated factor components BRD7 and BAF180 are critical regulators of p53 required for induction of replicative senescence. Proc. Natl. Acad. Sci. *107*, 14280–14285.

Core, L.J., Waterfall, J.J., and Lis, J.T. (2008). Nascent RNA sequencing reveals widespread pausing and divergent initiation at human promoters. Science (80-.). *322*, 1845–1848.

Davie, K., Jacobs, J., Atkins, M., Potier, D., Christiaens, V., Halder, G., and Aerts, S. (2015). Discovery of transcription factors and regulatory regions driving in vivo tumor development by ATAC-seq and FAIRE-seq open chromatin profiling. PLoS Genet. *11*, e1004994.

Deng, T., and Karin, M. (1994). c-Fos transcriptional activity stimulated by H-Ras-activated protein kinase distinct from JNK and ERK. Nature *371*, 171–175.

Djebali, S., Davis, C.A., Merkel, A., Dobin, A., Lassmann, T., Mortazavi, A., Tanzer, A., Lagarde, J., Lin, W., Schlesinger, F., et al. (2012). Landscape of transcription in human cells. Nature *489*, 101–108.

Drost, J., Mantovani, F., Tocco, F., Elkon, R., Comel, A., Holstege, H., Kerkhoven, R., Jonkers, J., Voorhoeve, P.M., Agami, R., et al. (2010). BRD7 is a candidate tumour suppressor gene required for p53 function. Nat. Cell Biol. *12*, 380–389.

ENCODE Project Consortium (2012). An integrated encyclopedia of DNA elements in the human genome. Nature *489*, 57–74.

Fulford, L., Milewski, D., Ustiyan, V., Ravishankar, N., Cai, Y., Le, T., Masineni, S., Kasper, S., Aronow, B., Kalinichenko, V. V., et al. (2016). The transcription factor FOXF1 promotes prostate cancer by stimulating the mitogen-activated protein kinase ERK5. Sci. Signal. *9*.

Grant, C.E., Bailey, T.L., and Noble, W.S. (2011). FIMO: Scanning for occurrences of a given motif. Bioinformatics *27*, 1017–1018.

Heintzman, N.D., Stuart, R.K., Hon, G., Fu, Y., Ching, C.W., Hawkins, R.D., Barrera, L.O., Van Calcar, S., Qu, C., Ching, K.A., et al. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nat. Genet. *39*, 311–318.

Heinz, S., Benner, C., Spann, N., Bertolino, E., Lin, Y.C., Laslo, P., Cheng, J.X., Murre, C., Singh, H., and Glass, C.K. (2010). Simple Combinations of Lineage-Determining Transcription Factors Prime cis-Regulatory Elements Required for Macrophage and B Cell Identities. Mol. Cell *38*, 576–589.

Irving, J., Feng, J., Wistrom, C., Pikaart, M., and Villeponteau, B. (1992). An altered repertoire of fos jun (AP-1) at the onset of replicative senescence. Exp. Cell Res. *202*, 161–166.

Kaiser, V.B., Taylor, M.S., and Semple, C.A. (2016). Mutational Biases Drive Elevated Rates of Substitution at Regulatory Sites across Cancer Types. PLoS Genet. *12*.

Kampfer, S., Hellbert, K., Villunger, A., Doppler, W., Baier, G., Grunicke, H.H., and Überall, F. (1998). Transcriptional activation of c-fos by oncogenic Ha-Ras in mouse mammary epithelial cells requires the combined activities of PKC-λ, ε and ζ. EMBO J. *17*, 4046–4055.

Khan, A., Fornes, O., Stigliani, A., Gheorghe, M., Castro-Mondragon, J.A., van der Lee, R., Bessy, A., Chèneby, J., Kulkarni, S.R., Tan, G., et al. (2017). JASPAR 2018: update of the open-access database of transcription factor binding profiles and its web framework. Nucleic Acids Res.

Kim, D., Pertea, G., Trapnell, C., Pimentel, H., Kelley, R., and Salzberg, S.L. (2013). TopHat2: Accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions. Genome Biol. *14*.

Kim, T.K., Hemberg, M., Gray, J.M., Costa, A.M., Bear, D.M., Wu, J., Harmin, D.A., Laptewicz, M., Barbara-Haley, K., Kuersten, S., et al. (2010). Widespread transcription at neuronal activity-regulated enhancers. Nature *465*, 182–187.

Korkmaz, G., Lopes, R., Ugalde, A.P., Nevedomskaya, E., Han, R., Myacheva, K., Zwart, W., Elkon, R., and Agami, R. (2016). Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9. Nat. Biotechnol. *34*, 1–10.

Langmead, B., and Salzberg, S.L. (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods *9*, 357–359.

Levine, M. (2010). Transcriptional Enhancers in Animal Development and Evolution. Curr. Biol. *20*, R754–R763.

Li, Q., Zhang, Y., Fu, J., Han, L., Xue, L., Lv, C., Wang, P., Li, G., and Tong, T. (2013). FOXA1 mediates p16INK4a activation during cellular senescence. EMBO J. *32*, 858–873.

Lin, A.W., Barradas, M., Stone, J.C., Van Aelst, L., Serrano, M., and Lowe, S.W. (1998). Premature senescence involving p53 and p16 is activated in response to constitutive MEK/MAPK mitogenic signaling. Genes Dev. *12*, 3008–3019.

Melgar, M.F., Collins, F.S., Sethupathy, P., Maniatis, T., Reed, R., Komili, S., Silver, P., Wyrick, J., Young, R., Kim, H., et al. (2011). Discovery of active enhancers through bidirectional expression of short transcripts. Genome Biol. *12*, R113.

Melo, C.A., Drost, J., Wijchers, P.J., van de Werken, H., de Wit, E., Vrielink, J.A.F.O., Elkon, R., Melo, S.A., Léveillé, N., Kalluri, R., et al. (2013). ERNAs Are Required for p53-Dependent Enhancer Activity and Gene Transcription. Mol. Cell *49*, 524–535.

Melton, C., Reuter, J.A., Spacek, D. V., and Snyder, M. (2015). Recurrent somatic mutations in regulatory regions of human cancer genomes. Nat. Genet. *47*, 710–716.

Milewski, D., Pradhan, A., Wang, X., Cai, Y., Le, T., Turpin, B., Kalinichenko, V. V., and Kalin, T. V. (2017). FoxF1 and FoxF2 transcription factors synergistically promote rhabdomyosarcoma carcinogenesis by repressing transcription of p21 Cip1 CDK inhibitor. Oncogene *36*, 850–862.

Rada-Iglesias, A., Bajpai, R., Swigut, T., Brugmann, S.A., Flynn, R.A., and Wysocka, J. (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature *470*, 279–285.

Riabowol, K., Schiff, J., and Gilman, M.Z. (1992). Transcription Factor Ap-1 Activity Is Required for Initiation of DNA-Synthesis and Is Lost during Cellular Aging. Proc Natl Acad Sci U S A *89*, 157–161.

Rose, D.W., McCabe, G., Feramisco, J.R., and Adler, M. (1992). Expression of c-fos and AP-1 activity in senescent human fibroblasts is not sufficient for DNA synthesis. J. Cell Biol. *119*, 1405–1412.

Sage, J., Miller, A.L., Pérez-Mancera, P.A., Wysocki, J.M., and Jacks, T. (2003). Acute mutation of retinoblastoma gene function is sufficient for cell cycle re-entry. Nature *424*, 223–228.

de Santa, F., Barozzi, I., Mietton, F., Ghisletti, S., Polletti, S., Tusi, B.K., Muller, H., Ragoussis, J., Wei, C.L., and Natoli, G. (2010). A large fraction of extragenic RNA Pol II transcription sites overlap enhancers. PLoS Biol. *8*.

Serrano, M., Lin, A.W., McCurrach, M.E., Beach, D., and Lowe, S.W. (1997). Oncogenic ras provokes premature cell senescence associated with accumulation of p53 and p16(INK4a). Cell *88*, 593–602.

Seshadri, T., and Campisi, J. (1990). Repression of c-fos transcription and an altered genetic program in senescent human fibroblasts. Science *247*, 205–209.

Smogorzewska, A., and de Lange, T. (2002). Different telomere damage signaling pathways in human and mouse cells. EMBO J. *21*, 4338–4348.

Subramanian, A., Tamayo, P., Mootha, V.K., Mukherjee, S., Ebert, B.L., Gillette, M.A., Paulovich, A., Pomeroy, S.L., Golub, T.R., Lander, E.S., et al. (2005). Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. Proc. Natl. Acad. Sci. *102*, 15545–15550.

Tamura, M., Sasaki, Y., Koyama, R., Takeda, K., Idogawa, M., and Tokino, T. (2014). Forkhead transcription factor FOXF1 is a novel target gene of the p53 family and regulates cancer cell migration and invasiveness. Oncogene *33*, 4837–4846.

Vierbuchen, T., Ling, E., Cowley, C.J., Couch, C.H., Wang, X., Harmin, D.A., Roberts, C.W.M.M., and Greenberg, M.E. (2017). AP-1 Transcription Factors and the BAF Complex Mediate Signal-Dependent Enhancer Selection. Mol. Cell *68*, 1067–1082.e12.

Wang, D., Garcia-Bassets, I., Benner, C., Li, W., Su, X., Zhou, Y., Qiu, J., Liu, W., Kaikkonen, M.U., Ohgi, K.A., et al. (2011). Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. Nature *474*, 390–397.

Weinhold, N., Jacobsen, A., Schultz, N., Sander, C., and Lee, W. (2014). Genome-wide analysis of noncoding regulatory mutations in cancer. Nat. Genet. *46*, 1160–1165.

Zanconato, F., Forcato, M., Battilana, G., Azzolin, L., Quaranta, E., Bodega, B., Rosato, A., Bicciato, S., Cordenonsi, M., and Piccolo, S. (2015). Genome-wide association between YAP/TAZ/TEAD and AP-1 at enhancers drives oncogenic growth. Nat. Cell Biol. *17*, 1218–1227.

Zhou, X., Maricque, B., Xie, M., Li, D., Sundaram, V., Martin, E.A., Koebbe, B.C., Nielsen, C., Hirst, M., Farnham, P., et al. (2011). The human epigenome browser at Washington University. Nat. Methods *8*, 989–990.

Supplemental data

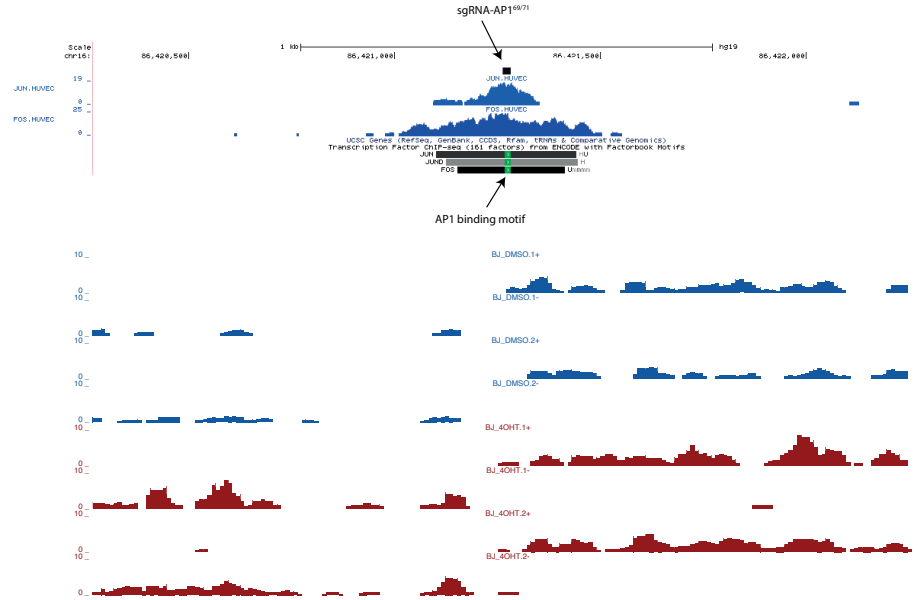


Figure S1. UCSC screenshot of Enh$^{AP1\text{-}OIS1}$ with ENCODE CHIP-seq data.

ENCODE CHIP-seq data shows a significant binding of c-Fos and c-Jun to *Enh*$^{AP1\text{-}OIS1}$. The binding of these factors coincides with the AP1 family consensus motif (Jun, JunD, and Fos in this case).

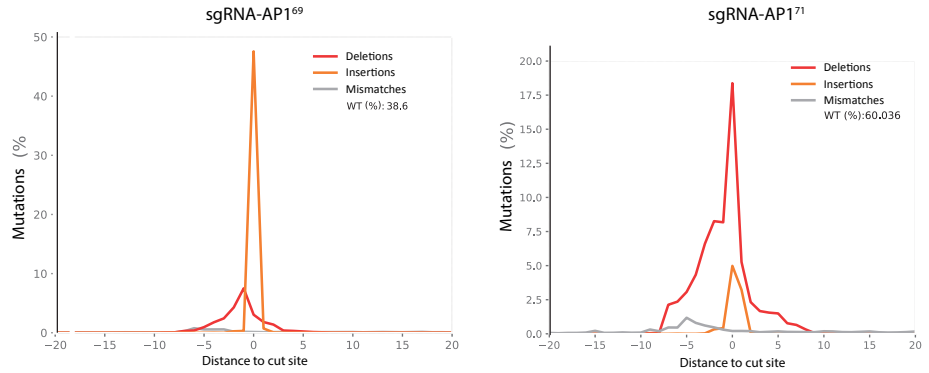


Figure S2. Mutation profiles of BJ-indRAS^{G12V} cells transduced with sgRNA-AP1^{69} and -AP1^{71}.

Genomic DNA of BJ-indRAS^{G12V} cells with the indicated sgRNAs was isolated, and *Enh*$^{AP1\text{-}OIS1}$ region was PCR amplified, subjected to deep sequencing and analysed for mutations. Proportion of wild-type or mutated (by mismatch, deletion or insertion) base calls is indicated as a function of its distance from the Cas9 cleavage site. For sgRNA-AP1^{69}, the majority of the mutations are single nucleotide insertions, while for sgRNA-AP1^{71} deletions around the cleavage site were more prevalent.

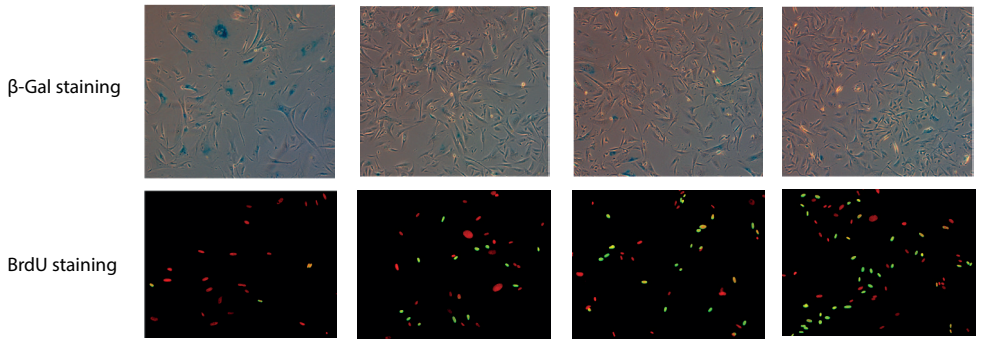


Figure S3. Representative pictures of *β*-Gal assay and BrdU staining.

A. Representative pictures of *β*-Gal assay and BrdU staining experiments used for the analysis of the results shown in Fig. 2-3. Each column represents the cells infected with the indicated sgRNA. The cells were treated with 4-OHT for 14 days prior analysis. Upper panel shows the *β*-Gal staining experiments with senescent cells stained as blue. Lower panel shows the BrdU staining experiments with cells incorporated with BrdU stained as green and nucleus stained as red.

Summary

ID	MU44835851
DNA change	chr16:g.86421280C>A
Type	single base substitution
Reference genome assembly	GRCh37
Allele in the reference assembly	C
Functional Impact	Unknown

Project	Site	Tumour Type	Tumour Subtype	# Donors affected
LUSC-KR	Lung	Lung cancer	Adenocarcinoma, Squamous cell carcinoma	1 / 66 (1.52%)

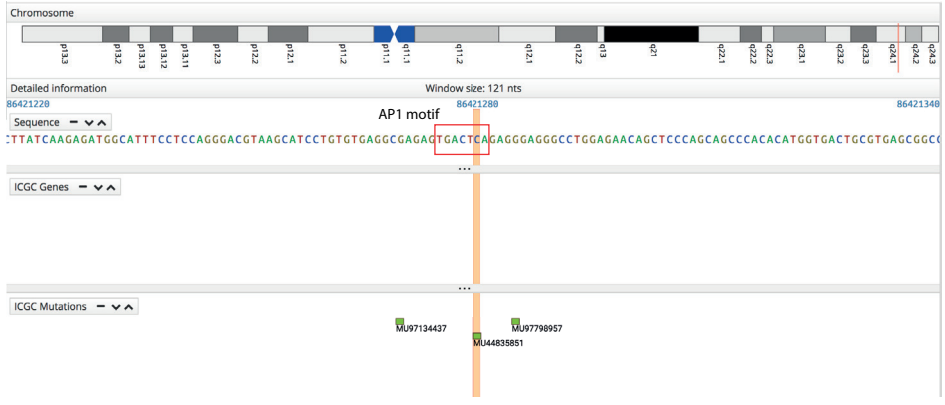


Figure S4. Somatic mutation in the AP1 motif within *Enh*$^{AP1\text{-}OIS1}$.

Screenshots from ICGC portal show a single nucleotide substitution within the AP1 motif in *Enh*$^{AP1\text{-}OIS1}$ detected in a patient with lung cancer.

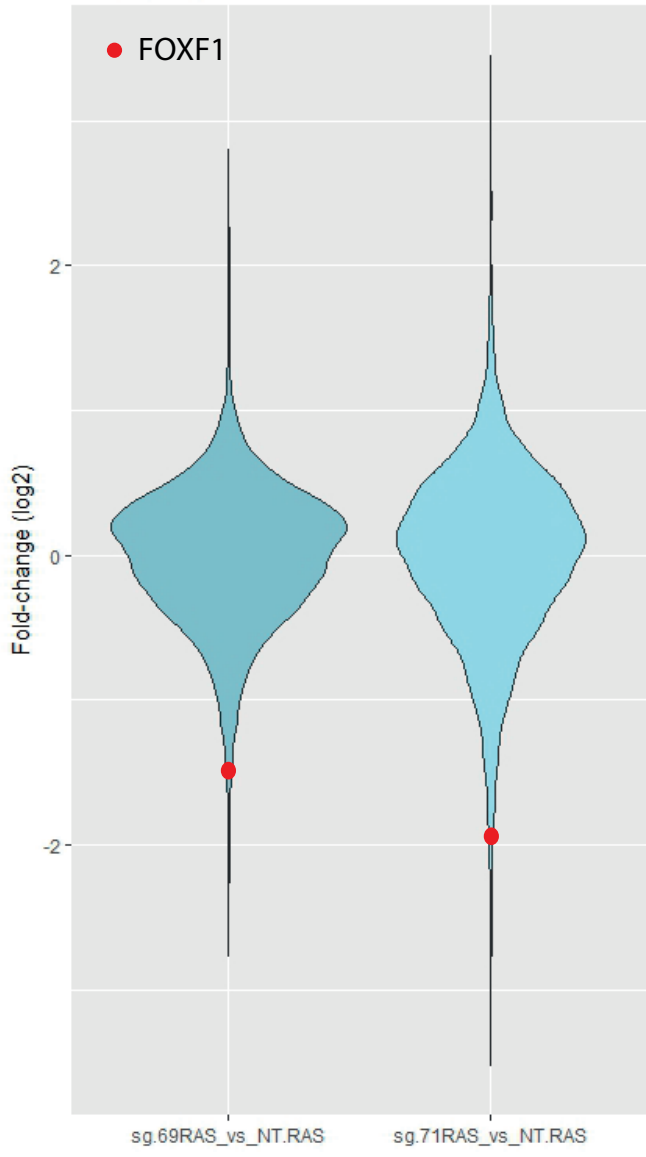


Figure S5. RNA-seq analysis indicates FOXF1 as a target gene of *Enh*$^{AP1\text{-}OIS1}$.

Gene expression levels were measured in BJ cells targeted by either sgRNA-AP1[69], sgRNA-AP1[71] or sgRNA-NT negative control. Violin plots show the distribution of fold change of gene expression (in log2 base) calculated for the comparison between the sgRNA-AP1[69] and sgRNA-AP1[71] samples and the sgRNA-NT control. *FOXF1* is marked by a rod dot. Its expression was markedly decreased by targeting *Enh*$^{AP1\text{-}OIS1}$ by either sgRNA-AP1[69] or sgRNA-AP1[71].

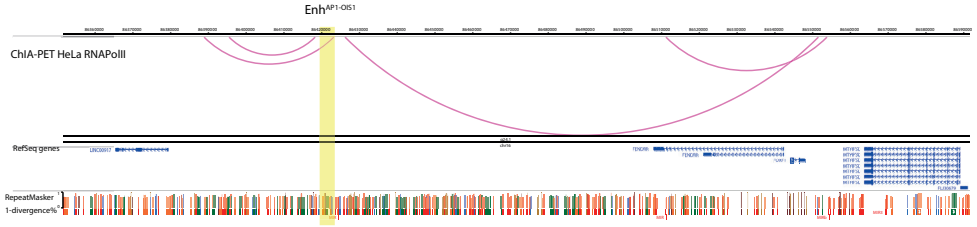


Figure S6. RNA PolII ChIA-PET data support direct interaction between Enh$^{AP1\text{-}OIS1}$ and FOXF1.

Each CHIA-PET experiment is indicated with the name of cell lines and the antibody used for the pull down on the left. *Enh$^{AP1\text{-}OIS1}$* is highlighted in yellow and inferred physical interactions between two regions are presented by loops. Data is taken from the WashU epigenome browser [50].

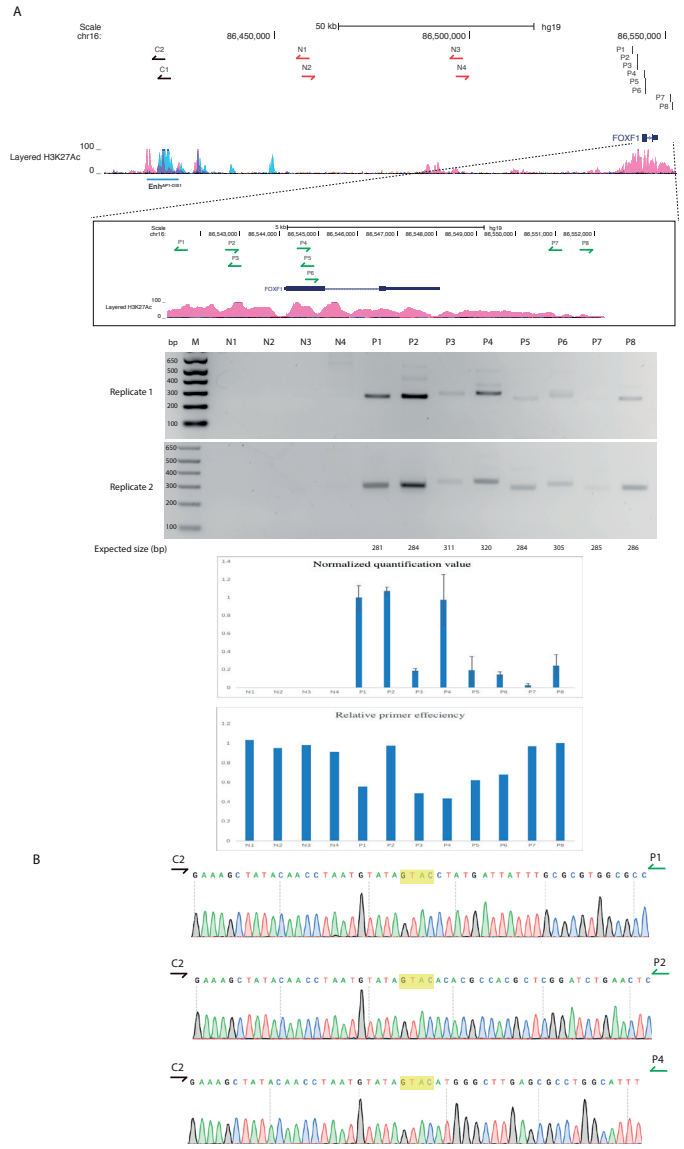


Figure S7. 3C experiment reveal direct interaction between *Enh$^{AP1\text{-}OIS1}$* and FOXF1.

A. Genome browser presentation of the location of each primer used in 3C analysis. Constant primers (C1, C2) used to amplify *Enh$^{AP1\text{-}OIS1}$* are indicated in black arrows. Negative control regions with no interaction with *Enh$^{AP1\text{-}OIS1}$* are amplified with primers indicated in red arrows (N1-N4). *FOXF1* regions with potential interactions with *Enh$^{AP1\text{-}OIS1}$* are amplified with primers indicated in green arrows (P1-P8). Agarose gel images from two independent biological replicates are shown with the expected sizes of the PCR products. The quantification of the gels was performed by normalizing to the primer efficiencies. Values shown are further normalized to the quantification value of P1. B. Sanger sequencing results from the indicated PCR products. Csp6I restriction sites are highlighted in yellow. The PCR were performed using the indicated primer pairs.

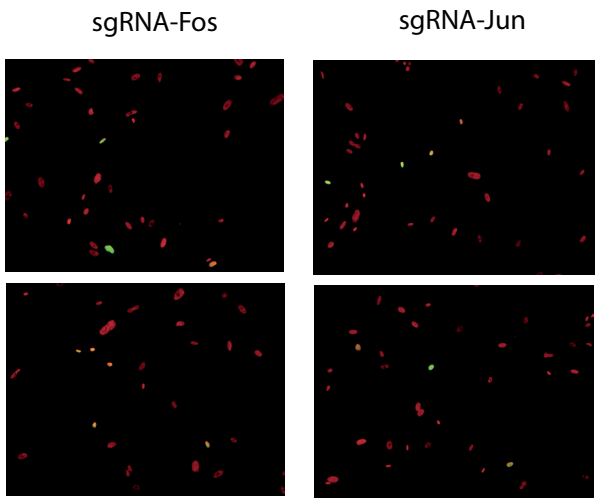


Figure S8. Disruption of Fos and Jun could not bypass senescence.

Representative pictures of BrdU staining experiments with sgRNAs against c-Fos and c-Jun genes. BJ cells were treated with 4-OHT for 14 days. Each column shows two pictures from the same experiment.

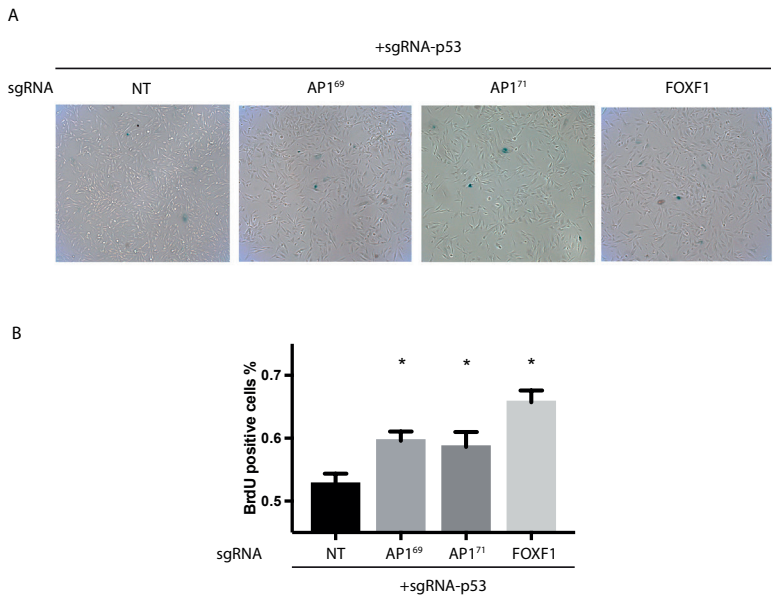


Figure S9. FOXF1 function is independent of p53 during OIS.

A. Representative images of the SA-β-gal staining of cells under the indicated conditions and after 14 days 4-OHT treatment. B. The proliferation levels of the various sgRNA-transduced BJ-indRAS^{G12V} cells (indicated in the figures) was quantified using BrdU assay, (* p <0.05, two-tailed Student's t-test). For every condition, the percentage of BrdU-positive cells was normalized to control cells (NT+p53 ko).

CHAPTER 5

General discussion

5

Transcription as a regulator of translation

As a starting point of the study, **Chapter 2** hypothesizes that promoters as the driving force in regulating translation. Firstly, it is recognized that transcription starts at a stretch of DNA sequences, which are known as promoters. Moreover, efficient transcription requires the presence of a promoter. Secondly, several studies have provided some insights into the regulatory functions of promoter sequences on mRNA metabolism (Bregman et al., 2011; Harel-Sharvit et al., 2010; Zid and O'Shea, 2014). Hence, we generated a promoter library to evaluate their effects on translation. To minimize any bias toward certain groups of promoters, we selected candidate promoters with distinct cellular functions (e.g., ribosomal proteins, cell cycle, transmembrane receptors, etc.).

Moreover, since the defined promoter regions of different genes vary in length, we took a region of 0.5-2.5kb upstream of the main TSS for cloning. This was supplemented with the information of different promoter-associated histone modification marks (e.g., H3K9me3, H3K27Ac). This consideration covers most of the functional region for promoters (e.g., transcription factor binding sites, DNA modifications, etc.). In the promoter library, we introduced pairs of promoters with or without TATA box, which is originated from the β-actin promoter, to directly compare the effect of TATA box on translation between these promoter pairs. Since it is difficult to specify any endogenous TATA box within each promoter, we cannot exclude the presence of any intrinsic TATA boxes within the promoter sequence. Still, the additional TATA box should govern any changes in translation within the TATA box pairs.

We noticed increased translational efficiency with an additional TATA box, despite further data analysis suggesting a weaker, yet significant effect upon the removal of the TATA box containing promoters from the promoter library. This finding implies that the TATA box might not be the sole factor in regulating translation in our experiments. Indeed, downstream experiments supported our assumption that TATA box is not the causal element, but that a boosted transcription rate could induce a more efficient translation.

There are three things that matter in m6A: location, location, location.

The occurrence of RNA modifications studies emerged over 60 years ago, yet their biological relevance and functions have not been well characterized. With the

rise of next-generation sequencing, genome-wide examinations of DNA and RNA have both become feasible and affordable. In 2012, Dominissini et al. and Meyer et al. developed in parallel, a method for detecting the footprints of m6A using anti-m6A antibodies (Dominissini et al., 2012; Meyer et al., 2012). By combining immunoprecipitation of m6A-bound antibody and next-generation sequencing, researchers have broadened our understanding of m6A biogenesis. Before our study, m6A was described as promoting translation by recruiting translational regulatory factors (eIF4F, HSP70) (Coots et al., 2017; Zhou et al., 2016). However, these findings were attributed to the m6A modifications in the 5'UTR or 3'UTR of genes, suggesting an indirect control of translation via co-factors. While we discovered a similar distribution of m6A in the UTR, we also identified a differential distribution of m6A along the coding region of genes. More importantly, we proposed a controversial effect of m6A on translation (repressive), possibly by disrupting translation elongation dynamics. This inhibitory effect on translation was also backed up by two studies, which showed that m6A at coding regions interferes with optimal codon usages (Choi et al., 2016; Hoernes et al., 2016). Together, we suggest that the exact functions of m6A are at least partially determined by its location. At UTR, m6A is recognized by m6A readers to promote translation actively; while at coding regions, m6A attenuates translation via suboptimal codon usage and impaired transcription elongation. However, further studies should address the significance of the locations for m6A, along with codon preferences during the translation process in more detail.

Co-transcriptional deposition of m6A – two birds one stone

The mammalian cell is a very efficient factory for optimal protein production, of which many steps occur simultaneously. While RNA pol II is transcribing along the DNA, assembly of the splicing complex can already take place. Thus, it is not surprising to realize that the deposition of m6A exists co-transcriptionally as described in our study. Interestingly, this finding is further supported by the fact that METLL3, an m6A writer, binds to genes with high levels of m6A in multiple m6A profiling datasets (Knuckles et al., 2017). In our study, we performed co-immunoprecipitation experiments to show that RNA Pol II and METLL3 interact during the impaired transcription elongation.

Further studies are required to reveal the dynamics of such interaction. For example, protein-protein interaction experiments (e.g., FRET) could uncover whether the interaction is conducted continuously or in an "on and off" manner. Moreover, it is not well-understood if such interactions could potentially affect

the rate of transcription elongation and other activities of RNA Pol II. Addressing these aspects could provide more insight into how the interaction is established, and why it is found more frequently in certain conditions.

AP1 feedback loop, a naughty boy with self-control

Characterization of the AP1 complex has been researched for decades, resulting in the discovery of its roles in cellular proliferation, differentiation, and cell death. As a well-interpreted transcription factor, AP1 is involved in many aspects of carcinogenesis, where it is primarily activated by various growth factors, cytokines, and oncogenic stimuli (Hess, 2004). In our study, we describe a more complex model explaining how mammalian cells respond to provide a checkpoint of tumorigenesis. During oncogenic RAS activation, AP1 is induced to promote cellular proliferation. There are various consequences for this process. First, p53 signaling pathway is activated and p21 locks the cells in the senescence state; second, AP1 is recruited to responsive elements where it induces a likely p53-independent senescence pathway. Thus, maintaining double insurance on cellular fitness.

There are a few drawbacks of our study. Firstly, we did not observe any effect in FOS or JUN knockout cells. This could be explained by the broad functions of AP1, including the involvement of different family members, and association with co-factors. As AP1 collaborates with cell-type specific transcription factors to select enhancers (Vierbuchen et al., 2017), it is reasonable to speculate that AP1 plays novel roles in the non-coding genome as well. Secondly, the senescence phenotype of enhancer or FOXF1 ko cells is not as potent as p53 ko cells, suggesting p53-regulated senescence program is still dominant during OIS. The supplemental checkpoint by AP1-regulated senescence is possibly a failsafe mechanism to help arrest cells in case of a non-functional p53 pathway. Third, double knockout of p53 and FOXF1 boosts the level of cell proliferation versus the single knockout cells, indicating (at least partially) a p53-independent nature of AP1-regulated senescence program. Notably, the knockout of FOXF1 does not affect the p21 protein level. Giveb the indispensable function of p53 during senescence, it will be difficult to dissect the effect of other regulatory elements or genes from that of p53 ko cells, which will likely cover any minor phenotypes. To resolve this issue, an appropriate model is demanded to identify the hidden regulatory networks of OIS.

Forkhead box protein, a new class of senescence modifier

Forkhead box proteins are large families of transcription factors that act on cell proliferation, differentiation, and senescence. To date, few studies have focused on the functions of Forkhead box proteins in cellular senescence. Phosphorylation of FOXO1 accelerates the senescence of liver in mouse (Tomobe et al., 2013). Protein levels of FOXA1, FOXO3, and FOXO4 are induced in senescent human fibroblast cells (Baar et al., 2017; Li et al., 2013). Conversely, down-regulation of FOXO3a stimulates the process of senescence in human fibroblast (Kyoung Kim et al., 2005). In our study, we discovered that FOXF1 is a positive regulator of senescence; however, the downstream mechanism is not well-understood.

Therefore, the exact mechanism of the regulatory role of FOX proteins during senescence is largely unknown, making it a matter of debate. Recently, the knowledge of FOX proteins has been applied to potentially reversing the process of aging in mammals (Baar et al., 2017). In the study, researchers developed a polypeptide to compete with the binding of FOXO4 to p53 in senescent cells, causing the exclusion of p53 from the nucleus – eventually leading to cell death. It is intriguing to discover that the use of the polypeptide has successfully reverted the fitness level and senescent phenotypes of mice. The relevance of such treatment for longevity in humans has yet to be carefully examined, but the functions of FOX proteins in senescence cannot be overlooked.

References

Baar, M.P., Brandt, R.M.C., Putavet, D.A., Klein, J.D.D., Derks, K.W.J., Bourgeois, B.R.M., Stryeck, S., Rijksen, Y., van Willigenburg, H., Feijtel, D.A., et al. (2017). Targeted Apoptosis of Senescent Cells Restores Tissue Homeostasis in Response to Chemotoxicity and Aging. Cell *169*, 132–147.e16.

Bregman, A., Avraham-Kelbert, M., Barkai, O., Duek, L., Guterman, A., and Choder, M. (2011). Promoter elements regulate cytoplasmic mRNA decay. Cell *147*, 1473–1483.

Choi, J., Ieong, K.W., Demirci, H., Chen, J., Petrov, A., Prabhakar, A., O'Leary, S.E., Dominissini, D., Rechavi, G., Soltis, S.M., et al. (2016). N6-methyladenosine in mRNA disrupts tRNA selection and translation-elongation dynamics. Nat. Struct. Mol. Biol. *23*, 110–115.

Coots, R.A., Liu, X.M., Mao, Y., Dong, L., Zhou, J., Wan, J., Zhang, X., and Qian, S.B. (2017). m6A Facilitates eIF4F-Independent mRNA Translation. Mol. Cell *68*, 504–514.e7.

Dominissini, D., Moshitch-Moshkovitz, S., Schwartz, S., Salmon-Divon, M., Ungar, L., Osenberg, S., Cesarkas, K., Jacob-Hirsch, J., Amariglio, N., Kupiec, M., et al. (2012). Topology of the human and mouse m6A RNA methylomes revealed by m6A-seq. Nature *485*, 201–206.

Harel-Sharvit, L., Eldad, N., Haimovich, G., Barkai, O., Duek, L., and Choder, M. (2010). RNA polymerase II subunits link transcription and mRNA decay to translation. Cell *143*, 552–563.

Hess, J. (2004). AP-1 subunits: quarrel and harmony among siblings. J. Cell Sci. *117*, 5965–5973.

Hoernes, T.P., Clementi, N., Faserl, K., Glasner, H., Breuker, K., Lindner, H., Hüttenhofer, A., and Erlacher, M.D. (2016). Nucleotide modifications within bacterial messenger RNAs regulate their translation and are able to rewire the genetic code. Nucleic Acids Res. *44*, 852–862.

Knuckles, P., Carl, S.H., Musheev, M., Niehrs, C., Wenger, A., and Bühler, M. (2017). RNA fate determination through cotranscriptional adenosine methylation and microprocessor binding. Nat. Struct. Mol. Biol. *24*, 561–569.

Kyoung Kim, H., Kyoung Kim, Y., Song, I.-H., Baek, S.-H., Lee, S.-R., Hye Kim, J., and Kim, J.-R. (2005). Down-regulation of a forkhead transcription factor, FOXO3a, accelerates cellular senescence in human dermal fibroblasts. J. Gerontol. A. Biol. Sci. Med. Sci. *60*, 4–9.

Li, Q., Zhang, Y., Fu, J., Han, L., Xue, L., Lv, C., Wang, P., Li, G., and Tong, T. (2013). FOXA1 mediates p16INK4a activation during cellular senescence. EMBO J. *32*, 858–873.

Meyer, K.D., Saletore, Y., Zumbo, P., Elemento, O., Mason, C.E., and Jaffrey, S.R. (2012). Comprehensive analysis of mRNA methylation reveals enrichment in 3??? UTRs and near stop codons. Cell *149*, 1635–1646.

Tomobe, K., Shinozuka, T., Kawashima, T., Kawashima-Ohya, Y., and Nomura, Y. (2013). Age-related changes of forkhead transcription factor FOXO1 in the liver of senescence-accelerated mouse SAMP8. Arch. Gerontol. Geriatr. *57*, 417–422.

Vierbuchen, T., Ling, E., Cowley, C.J., Couch, C.H., Wang, X., Harmin, D.A., Roberts, C.W.M.M., and Greenberg, M.E. (2017). AP-1 Transcription Factors and the BAF Complex Mediate Signal-Dependent Enhancer Selection. Mol. Cell *68*, 1067–1082.e12.

Zhou, J., Wan, J., Gao, X., Zhang, X., Qian, S., and Biology, C. (2016). Dynamic m6 A mRNA methylation directs translational control of heat shock response. Nature *526*, 591–594.

Zid, B.M., and O'Shea, E.K. (2014). Promoter sequences direct cytoplasmic localization and translation of mRNAs during starvation in yeast. Nature *514*, 117–121.

SUMMARY

The nature of life is intrinsically simple and complex, with merely four basic nucleotides that ultimately make up countless lives with distinct appearances. Darwin's theory of evolution reveals how time shapes the inhabitants of the planet, whereby it is the fundamental components that determine who we are. It is not until several decades ago, where we finally cracked the mystery of the DNA molecule with the swirling structure. Since then, waves of biology studies have entered an exponential phase. With the development of technologies, scientists have increased their understanding of the mechanisms behind biological questions in various organisms. As the initial step of all biological processes, transcription from DNA to RNA has been extensively characterized. We have acknowledged that this series of action is not at all effortless, but largely regulated. To name a few, the state of chromatin structure, level of transcription factor binding, initiation, and elongation of transcription, the post-transcriptional regulation of messenger RNA, and not to mention the influence of non-coding RNA. Moreover, the messenger RNA is stepwisely exported to the cytoplasm of the cells and translated into protein in mammalian cells. Despite being located in two spatially separate compartments, the crosstalk is never ignored.

Hence, it is of great interest to investigate whether this crosstalk between transcription and translation has any consequences to cellular functions. Our study looks at the starting point of transcription – promoter, by performing an unbiased screen for candidate promoters that could potentially affect the translation process Interestingly, we discovered a set of promoters with TATA box, which boosts the translation of messenger RNA. Using CRISPR-Cas9 genome editing, genome-wide analysis, and *in vitro* experiments, we found that the effect is not limited to promoters with TATA box. Instead it is a rather general mechanism dictated by the rate of transcription, more specifically transcription elongation. More importantly, the performance requires cap-dependent translation. To examine the responsible factor, we assess the level of RNA modification - m6A on the messenger RNA and observe a reverse correlation between the status of translation and the level of m6A modification. Notably, although m6A are mostly found around stop codons, in our study m6A modification within the coding region of genes are crucial for the suppressive effect on translation. Finally, we propose a model that co-transcriptional deposition of m6A via retarded RNA Pol II elongation could lead to translational control. The study opens new doors to grasping the hidden regulators of gene expression.

The advancement of CRISPR-Cas9 genome editing has led to a myriad of genome-wide studies in various organisms. While most studies focus on the coding regions of the genome, we have previously established a systemic approach to study the functions of non-coding genome by targeting transcription factor binding sites at enhancer elements. To broaden our understanding about how enhancer elements coordinate gene expression, we performed a genetic screen of enhancers with AP1 binding motif in oncogene-induced senescence. We successfully validated two sgRNAs targeting the same AP1 motif, showing it can affect the process of senescence. Mutations of the core AP1 binding site, in turn, abolished the enhancer function. Later, it was discovered that the FOXF1 gene was the interacting target, where disruption of FOXF1 lead to similar effects on senescence. Together, we suggest that AP1, a downstream target of many oncogenes, maintains a negative feedback loop to counteract the unlimited cell proliferation driven by oncogene activation, thus adding an extra tier of cellular surveillance to tumorigenesis.

SAMENVATTING

De aard van het leven is intrinsiek eenvoudig en complex, met slechts vier fundamentele nucleotiden die ontelbare levens met verschillende verschijningsvormen codeert. Volgens Darwins evolutietheorie is onthuld hoe tijd de bewoners van de planeet vormgeeft, terwijl het fundamentele element dat codeert voor wie we zijn blijft bestaan. Pas enkele decennia geleden is eindelijk het mysterie van het DNA-molecuul met de helix structuur opgehelderd. Sindsdien is de golf van biologiestudies een exponentiële fase ingegaan. Met de ontwikkeling van technologieën zetten wetenschappers een grote stap richting het begrijpen van de mechanismen achter biologische vragen in verschillende organismen. Als de eerste stap van alle biologische processen is transcriptie van DNA naar RNA uitgebreid gekarakteriseerd. We hebben erkend dat deze reeks acties helemaal niet moeiteloos, maar grotendeels gereguleerd is, van de staat van de chromatine structuur, niveau van transcriptiefactor binding, initiatie en verlenging van transcriptie, tot de post-transcriptionele regulatie van boodschapper-RNA, om nog maar te zwijgen van de invloed van niet-coderend RNA. Bovendien wordt in zoogdier cellen het boodschapper-RNA stapsgewijs geëxporteerd naar het cytoplasma van de cellen en daar vertaald in eiwit. Ondanks dat deze zich in twee ruimtelijk gescheiden compartimenten bevindt, wordt de crosstalk nooit genegeerd.

Het is daarom van groot belang om te onderzoeken of deze crosstalk tussen transcriptie en translatie consequenties heeft voor cellulaire functies. We zetten onze voet op het startpunt van de transcriptie – de promotor - door een onbevooroordeelde screen uit te voeren voor kandidaat-promotors die mogelijk het vertaalproces zouden kunnen beïnvloeden. Interessant is dat we ontdekken dat een aantal promotors met TATA-boxen de vertaling van messenger-RNA stimuleren. Met behulp van CRISPR-Cas9 gemedieerde bewerking van het genoom, genoom-brede analyse en *in vitro* experimenten onthullen we dat het effect niet beperkt is tot promotors met een TATA-box, maar eerder een algemeen mechanisme dat wordt gedicteerd door de snelheid van transcriptie, meer specifiek transcriptie-verlenging. Wat nog belangrijker is, is dat de uitvoering een cap-afhankelijke translatie vereist. Om de verantwoordelijke factor te onderzoeken, beoordelen we het niveau van een RNA-modificatie - m6A op het boodschapper-RNA en observeren we een omgekeerde correlatie tussen de snelheid van translatie en het niveau van m6A-modificatie. Opmerkelijk is dat, hoewel meestal aangetroffen rond stopcodons, de M6A-modificaties in onze studie binnen het coderende gedeelte van genen causaal is voor het onderdrukkende effect op translatie. Ten slotte stellen we een model voor dat co-transcriptionele depositie van m6A via vertraagde RNA PolII-verlenging kan leiden tot translationele controle. De studie opent nieuwe deuren voor het ontdekken van de verborgen regulatoren van genexpressie.

De vooruitgang van CRISPR-Cas9 genoom editing heeft geleid tot een groot aantal genoom-brede studies in verschillende organismen. Hoewel de meeste studies zich richten op de coderende gebieden van het genoom, hebben we eerder een systemische benadering gebruikt om de functies van het niet-coderende genoom te bestuderen via targeting van transcriptiefactor bindingsplaatsen op enhancer-elementen. Om onze kennis over hoe enhancer-elementen genexpressie coördineren te verbreden, voeren we een genetische screening uit van enhancers met een AP1-bindingsmotief in door oncogen geïnduceerde senescence. We valideren met succes twee sgRNA's die gericht zijn op hetzelfde AP1-motief en die het proces van senescence kunnen beïnvloeden. Mutaties van de kern van de AP1-bindingsplaats heffen de enhancer-functie op. Later ontdekken we dat het FOXF1-gen het interacterende doelwit is en dat verstoring van FOXF1 leidt tot vergelijkbare effecten op senescence. Alles samen, stellen we voor dat AP1, een stroomafwaarts doelwit van veel oncogenen, een negatieve feedback in stand houdt om de onbeperkte celproliferatie, aangedreven door activering van het oncogen, tegen te gaan door een extra laag van cellulair toezicht voor tumorgenese toe te voegen.

CURRICULUM VITAE

Ruiqi Han was born in Changchun, China on February 5th, 1988. After completing his high school study in Changchun, he was admitted to Qingdao University of Science and Technology in 2006. Before starting his university life, he made the decision to depart from the university for pursuing his study in the Netherlands. In 2007, he was admitted to the international program of life sciences of the Hogeschool van Nijmegen en Arhnem (HAN) in Nijmegen. During his bachelor study, he worked on the functions of microRNAs in bladder and prostate cancer in the group of prof. Jack Schalken under the supervision of Dr. Gerald Verhaegh. He completed his bachelor thesis on the microRNA targets on ABC transporters in liver cancers in the RNAi section of Uniqure B.V. (formerly known as Amsterdam Molecular Therapeutics B.V.), and published his first peer-reviewed publication under the supervision of Dr. Pavlina Konstantinova. Based on this study, he was offered a PhD position in the group, of which he gratefully declined. In 2011, he joined the research master program of oncology at the Vrij University of Amsterdam. During his master study, he joined Dr. Anthony Kong's group at the Weatherall Institute of Molecular Medicine (WIMM) of Oxford University to study how Hsp90 inhibitors could overcome Herceptin-resistant breast cancer. Upon return to Amsterdam, he joined prof. Reuven Agami's group at the Netherlands Cancer Institute (NKI) to study how transcription could affect translation. After receiving his master degree, he stayed in the group to continue his work as described in this book.

LIST OF PUBLICATIONS

Boris Slobodin#, **Ruiqi Han#**, Vittorio Calderone, Joachim AF Oude Vrielink, Fabricio Loayza-Puch, Ran Elkon, Reuven Agami*. Transcription impacts the efficiency of mRNA translation via co-transcriptional N6-adenosine methylation. **Cell** 169(2): 326-337. (2017)

Ruiqi Han, Boris Slobodin, Reuven Agami. The Methylated way to translation. **Oncotarget** 8(55): 93313-14. (2017)

Ruiqi Han#, Li Li#, Alejandro Piñeiro Ugalde, Arieh Tal, Zohar Manber, Eric Pinto Barbera, Veronica Della Chiara, Ran Elkon, Reuven Agami. Functional CRISPR screen identifies AP1-associated enhancer regulating FOXF1 to modulate oncogene-induced senescence. **Genome biology** 19: 118. (2018)

- **Highlighted in the Genome Biology special issue of Genome editing**

Li Li, Pieter C van Breugel, Fabricio Loayza-Puch, Alejandro P Ugalde, Gozde Korkmaz, Naama Messika-Gold, **Ruiqi Han**, Rui Lopes, Eric P Barbera, Hans Teunissen, Elzo de Wit, Ricardo J Soares, Boye S Nielsen, Kim Holmstrøm, Dannys J Martínez-Herrera, Maite Huarte, Annita Louloupi, Jarno Drost, Ran Elkon, Reuven Agami; LncRNA-OIS1 regulates DPP4 activation to modulate senescence induced by RAS. **Nucleic Acids Research** gky087, https://doi.org/10.1093/nar/gky087. (2018)

Gozde Korkmaz#, Rui Lopes#, Alejandro P Ugalde, Ekaterina Nevedomskaya, **Ruiqi Han**, Ksenia Myacheva, Wilbert Zwart, Ran Elkon, Reuven Agami*. Functional genetic screens for enhancer elements in the human genome using CRISPR-Cas9. **Nature Biotechnology** 34(2): 192-198. (2016)

Florie Borel, **Ruiqi Han**, Allerdien Visser, Harald Petry, Sander J.H. van Deventer, Peter L.M. Jansen, Pavlina Konstantinova, with collaboration of the Réseau Centre de Ressources Biologiques Foie (French Liver Biobanks Network), France. Adenosine triphosphate-binding cassette transporter genes up-regulation in untreated hepatocellular carcinoma is mediated by cellular microRNAs. **Hepatology** 55: 821–832. doi:10.1002/hep.24682. (2012)

Co-first authors

ACKNOWLEDGE-
MENTS

After an exciting and meaningful 5-year journey, finally it is time to look back to the past for a better start of the future. It goes without saying that I have learned so much from all the colleagues and friends that I encountered, from top scientific discussions to deep thoughts of life. As a proper ending of every novel, film, drama, and last but not least - a Phd study, I would like to express my appreciation to all of you:

First of all, I would like to thank my supervisor Reuven, who brought me into this, in my opinion, “all about exciting science” lab. I am grateful that you offered me, a final-year master student, as an internship to initiate a journey on a world-class challenging project. Later on, you kept me as a Phd student to continue this journey; and together we witnessed a happy ending to it. Your scientific ideas were inspiring, which helped me finalize my study smoothly.

Boris, we knew each other from the very first day of the interview with Reuven. From then, we together worked on a dream “Nature” paper for the next four years. At the bench side, I always regarded you as my second supervisor who was there for answering my questions and generously sharing your valuable experiences and knowledge. Remembering the times when we were about to drop the project, I am glad that we kept our faith in our research, getting both of us a rocket boost to our career. As to personal life, I enjoyed, as a friend, talking with you about life, cultures, and families. I will always have time reserved for you to grab a beer somewhere in this world.

Alejandro, it was a pleasure to taste your home-brewed beer; it was pure but sophisticated, just like how I felt about you. You have so much knowledge and countless tricks, and you were never too shy to share them with others. Your determination and confidence have surely influenced me, and have shed light on many dark hours. Fabricio and Pieter, the first day when I started in the group, I was placed in an office with three postdocs, you guys and Boris. I have to say it was not easy for me to catch up with you guys scientifically at the beginning, but that also helped me to grow fast and to question like a postdoc – working independently and think creatively. Without the generous help and guidance from you guys, my Phd study would have become much more difficult. And Rani, thank you for all the beautiful bioinformatic analysis you performed, and the ideas you initiated. During the time when you were in Amsterdam, you could always give me valuable suggestions. It was a privilege to have you as a strong support for all the publications of my Phd study.

Remco, it was you who brought order to our chaotic lab. Your talent in organizing and managing the lab not only helped us to work efficiently, but also made us realize the fact that, we could have done it better. Thank you for your help with the Dutch translation of this thesis, and more importantly, for all the last-minute ordering you placed for me. And to probably the most “senior” member of the group ever, Arieh, many appreciations for your never-ending attitude of pursuing what you like. Thanks for all the analysis you performed, and all the “lessons” you gave inside offices, by the coffee machine, and by the lunch table. You are my number one role model for a true free spirit.

It is fascinating to watch the transition of the lab, from RNA biology to translation, metabolism, genomics, as a junior member of the group to one-of-the-left seniors. I have to mention all the previous and current members of the group, Carlos, Nick, Gozde, Koos, Rui, Yuval, Abhi, Julien, Behzad, and Itamar, thank you for being so patient to my questions, helping me out with experiments, analysis, urgent reagent requests and making my Phd life a very memorable moment.

And to the master students that I worked with, Eric, it was a very comfortable experience to work with you. You have surprised me and Li with your skills and knowledge, and I am glad that you are now on the track to a great future. Veronica, it was a, hard-to-describe, full-of-gossip, dramatic, experience to work with you. I still remember you came as a clueless volunteer student, trying to fetch a good opportunity for a Phd position. During those periods, we initiated and established an exciting project together; although I am still working on it, I promise that I will do my best to finish it well. It was a very pleasant time to gather together with you, Alejandro, Fabricio, and Li, which would always result in endless conversations. Lyubomir, thank you for helping out with the cloning and the cell culture of Li and me. We have had a good time during the “outdoor hours”, it was a pleasure to work with you.

Many thanks to my Phd evaluation committee, Fred van Leeuwen, Heinz Jacobs, and Jacqueline Jacobs. Every year I received very useful suggestions, which always kept me on the right track of my Phd study. And my thesis committee - Wilbert Zwart, Peter Verrijzer, and Raymond Poot, thank you all for taking the time to read and assess my thesis.

I have to express my gratitude to the ladies that made our stay hassle-free. Suzanne, Thea, Anne, and Elise, thank you all for arranging the FedEx packages, dealing with HR issues, organizing social activities, and many other things. Without your help, it would be impossible to put all my focus on my work.

Since the group has physically moved twice to different departments, I was very fortunate to get to know many colleagues from different backgrounds. First of all, B4 was where it all started, Bas, Marcel, Joris, Laura, Eva, Ludo, Sandra, Fred, Deepani, Tibor, Hanneke, Tessy, Thom, and everyone else from the old-time B4, thank you all for bearing me during my stay in the department and all the interesting discussions during the department meeting. Then the story moved to H3, Hans, Heinz, Muhammad, Farshid, Alessandra, Paul, Bas, Aldo, Hein, Bente, Thomas, Wiestke, Marleen, Ben, and all the colleagues in the department, it was probably difficult for you guys to accept the sudden incorporation of a new group in the department, especially when we were being messy and unorganized; but thank you all for tolerating us. Finally, it comes to the current H5, Wilbert, Suzan, Karianne, Simon, Stacey, Stefan, Anniek, Isabel, Jeroen, Tesa, Abishek, Jacqueline, Inge, Aurora, Judit, Zeliha, Santiago, William, Eric, Joana, Rob, Piet, and Henry, thank you all for the unforgettable moments and discussions in the department. This is a lively and friendly department, which I believe will get better and better in the coming years.

Next, to all the colleagues that I have had the honor to work with, Elzo and Hans, thank you for the collaborations and answering all the questions we had during the experiments. Ron, Wim, and all the colleagues from the sequencing facility, I think I was blessed to have met you guys. Your work had saved a considerable amount of valuable time during revision periods of my manuscripts, and the communications were always straightforward, which made everything simple. Desiree, Frank, Annita, and Martijn, thank you for your help with using the different machines. Your generous help had made my days at NKI warm memories of my life.

接下来，我需要感谢在NKI的中国同事和盆友们。**李礼**，我们同时开始了博士之旅，从一开始我们就一直站在一条战线上，无论是对于科学上的合作，实验室里的烦心事还是生活中的互相帮助，我们都有着高度的一致。我很幸运能够和你共同合作完成了很多项目，并几乎同时完成博士的学业。虽然我届时已经离开了NKI，但是你也要在这虎狼之地坚强地支撑到最后，完成你最后的收尾之作。**孙建慧**，你虽然是后来才加入到组里的，但是从一开始我们就希望能够真正帮助到你，因为我知道读博士不易，且发且珍惜。这些年你帮助了我很多，一点一滴的我都记在了心里，希望你能顺利完成学业，成为一个内心强大的人。

黄馨瑶同学，我与你的渊源是在太久远，一直追溯到本科第一年的学习生涯。没想到过了11年后，我们还能一起站在这博士毕业的舞台上，不得不说是一种不平凡的缘分。这5年的NKI学习生涯我们都成长了不少，从当初一起喝酒打麻将到现在能一起合作解决科学问题。谢谢你一直以来的帮助和鼓励，希望以后还能有机会我们一起共商大事。**乔晓航**，虽然在刚来NKI的时候我们并不是很熟，但是后来我们渐渐地成为了好朋友。也许是同为东北人的原因，我们沟通起来从来都不拖泥带水非常顺畅。这些年我从你身上学习了不少东西，同时在生活中也麻烦了你不少事，这些点点滴滴让我和丁丁都非常感动。衷心希望你能找到属于自己的那份安宁与幸福。

陈涛，从我来NKI实习那时开始，我就发现了你的与众不同。接下来的几年，我就更加确定你是一个不走寻常路之人，在会议上鬼斧神工的问题和生活中放荡不羁的态度都让人印象深刻。非常感谢你对我在科学上的指导和提供那些帮助我提升自己的机会，这些都让我成长为了一个自己不曾想过的更好的人。虽然你现在与科学本身渐行渐远，但我希望以后不管在地球的哪一边，我们都能有合作的机会。**孙崇**，你让我见识到了NKI真是个卧虎藏龙之地，未曾想到这里会有如此出色的科学家，同时不乏出色的娱乐精神。看到你现在的成就，也会激励我不能倦怠，毕竟天外有天人外有人。说到这里就不的不提一些已经离开了荷兰的小伙伴们，**张平**，你说好回来答辩的那天好像过了3年了也没有到来，不知道你现在当医生是否真的事事如愿，期待你能回荷兰来叙旧。**徐国泰**，那些在你家一起打够级，每周五在篮球场上驰骋的画面还历历在目，感谢这些年我们曾经一起制造的欢乐。**陈湜** （话说你的名字输入法里为什么找不到），**曲璇**，**刘真**，三位都已经是或即将成为人母的女士。非常开心能与你们共同度过了那么多愉快的时光，特别是那次单身派对的惊喜，着实让我又惊又感动。希望你们和孩子们都健康成长，同时还记得有我这个穿着连衣裙飞舞的男人。

王存，**孔祥俊**，**王耀庚**，**朱燕韵**，**薛征**，这些年多亏了你们在科学上的帮助并且无私提供你们的试剂，这些都是我博士学习期间无价的财富。同时还有在NKI的所有中国人群体，Chun-pu，**甘常培**，**郭小虎**，**金浩杰**，**李文龙**，**王力勤**，**王静**，**王伟**，**肖艳玲**，非常幸运能够结识同在一个屋檐下的你们，你们共同铸就了亚洲人在研究所良好的口碑。

最后，我要感谢养育我的父母，是你们对我所做决定的支持成就了今天站在这里的我。感谢你们从小给了我良好的成长环境和正确的价值观和人生观，这些都对我之后的人生产生了非常积极的影响。虽然我19岁就离开了祖国来到了这个未知的国度，但是你们一直也永远都是我最强大的后盾和支持。之所以这些年我能安心地完成学业，你们给予我物质与精神上的支持是不可或缺的。年轻时，你们奉献了自己成就了我；当岁月无情，抚去你们年轻时的光彩，我会努力成为一个有能力的人，有能力照顾你们和家庭，有能力为朋友遮风挡雨，有能力为人类的命运带来微薄的贡献。同时我要感谢我的岳父岳母，虽然我们缘起数年前，但是你们带给我的温暖和支持确是终生难忘的。从结识这段姻缘开始，我就暗立誓言，要像对待我的父母一样照顾你们。这些年多亏有你们时不时的关怀和教育，才能帮助我们从婚姻里的婴儿成长为婚姻中的少年。感谢你们对于我工作的理解，我相信你们的无私和慷慨会继续鼓舞我朝着一个正确的方向前进。我希望我们的小家和整个大家在一起的时光都会成为我们彼此人生不想抹去的快乐回忆。

当然还有Kenzo，呼噜，鱼丸，你们的存在让我们这个小家不再安静，有时甚至过分的热闹，但是感谢你们这几个小生命选择并容忍我们成为你们的主人，你们带给我们的快乐我们会用一辈子的时间来照顾你们。

最后的最后，我要感谢带给我这辈子最多快乐的老婆，周之丁。还记得多年前，那个还一无所有的稚气男孩遇到了花样年华的你，从此他的人生不再灰暗无色。你像是打开了我人生的魔盒，那些奇妙的事情接二连三地都出现在了我的眼前，而你就像个挥舞魔棒脚踏彩虹的少女，默默地小心翼翼地导演着这一幕幕的剧情。因为爱情，你选择同我一同离开了熟悉的故土，走上了一条对你来说陌生而又孤独的道路。感谢你这一路无私的陪伴，让我感受到了夜归遍寻处，盏灯暖身屋。你的包容和照顾让我能够更专心地投身我的事业，而你的支持和对我的信心更是驱动我不停向前的动力。也许目前我还不能给你一个我们都梦想的生活，但是我相信只要我们都怀抱着那个最初的信念，一切都不会遥远。

朋友们，同事们，家人们，在这一刻，当一切尘埃落定，庆幸我们还年轻，还有能力好好准备自己，来无畏地直面那外面的世界。

Special thanks to: Cover design by Caiqiu Yang from Sydney, Australia.
特别鸣谢：封面设计由远在悉尼的杨采秋倾力设计。