Regulation of Coding and Non-coding Genes;
New insights obtained through analysis of high-throughput sequencing data

Koos Rooijers

PART A

Cover (Part A): A digital view of translation, Koos Rooijers, 2015
PDB entry used: 4UG0

ISBN: 978-94-6233-187-7

Regulation of Coding and Non-coding Genes;
New insights obtained through analysis of high-throughput sequencing data

Regulatie van coderende en niet-coderende genen;
Nieuwe inzichten verkregen door analyse van high-throughput sequencing data

Thesis

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

Prof.dr. H.A.P. Pols

and in accordance with the decision of the Doctorate Board

The public defence shall be held on

Tuesday 2nd of February, 2016 at 15:30 hrs

by

Koos Rooijers,

born Haarlem, the Netherlands, 7th of October, 1986

Erasmus University Rotterdam

Doctoral Committee

Promotor: prof.dr. R. Agami

Other members: prof.dr.ir. G. Jenster
dr. M. von Lindern
prof.dr. W. Vermeulen

Contents

Part A – Protein translation studied by ribosome profiling

Chapter 1 Introduction to Part A **A 9**

Chapter 2 Ribosome profiling reveals features of normal and disease-associated mitochondrial translation **A 21**

Chapter 3 Sensing amino acid deficiencies through differential ribosome codon occupancies **A 43**

Chapter 4 Specific changes of ribosome distribution over messenger-RNAs during mitotic arrest **A 77**

Chapter 5 Discussion of Part A **A 103**

Part B – Transcription of the non-coding genome

Chapter 6 Introduction to Part B **B 5**

Chapter 7 Genome-wide profiling of p53-regulated enhancer RNAs uncovers a subset of enhancers controlled by a lncRNA **B 9**

Chapter 8 Discussion of Part B **B 47**

Summary **B 49**

Samenvatting **B 51**

Curriculum Vitae **B 53**

List of publications **B 54**

Abbreviations

3-AT	3-Amino-1,2,4-triazole
ALL	Acute lymphoblastic leukemia
bp	Basepair
CDS	Coding DNA sequence (often also used when concerning an RNA molecule)
ChIP	Chromatin immunoprecipitation
ChIRP	Chromatin isolation by RNA purification
diricore	Differential ribosomal codon reading
EMT	Epithelial to mesenchymal transition
FACS	Fluorescence-activated cell sorting
GO	Gene ontology
ID	Identifier
IRES	Internal ribosome entry site
kb / kbp	Kilobasepair
nt	Nucleotide(s)
ORF	Open reading frame
PCR	Polymerase chain reaction
PEG	Poly-ethylene glycol
ROS	Reactive oxygen species
RP	Ribosomal profiling
RPF	Ribosome protected fragment
SNP	Single-nucleotide polymorphism
STLC	S-trityl-L-cysteine
TE	Translational efficiency
TF	Transcription factor
uORF	Upstream open reading frame
wt	Wild-type
YFP	Yellow fluorescent protein

There are times when I find myself transfixed by a shadow on the wall, or the splashing of water against a stone. I stare at it, the hours pass, the world around me drops away, replaced by worlds being created and destroyed by my imagination. A way to focus the mind.

ST:VOY s03e26, sample featured in Dom&Roland’s ‘Imagination’

CHAPTER 1

Introduction to Part A – Protein translation studied by ribosome profiling

In the study of a large organism such as the human, the cell provides a convenient unit. An adult human consists of many cells (in the order of approximately 10^{14}, that is: a hundred million millions) and in many ways all these cells share many commonalities. In multicellular organisms such as the human, these cells are ordered into larger macro-structures which form tissues, organs and body parts. It should be obvious that the cells across the body have different characteristics, properties and presentations, which allow the different tissues, organs and body parts to develop. Indeed, among the cells in a large organism different specialized populations of cells can be identified, which gives rise to a classification of cell types. Cells of the same cell type share certain phenotypic and/or morphological traits, which set them apart from other cell types[†]. What gives cells their phenotype, morphology, behavior? An important constituent of the cell, and major determinant of its properties, are proteins.

Proteins as determinant of cellular phenotype

Protein is a class of biomolecule that is built up from simpler building blocks, the amino acids. The amino acids are joined (polymerized) into long strings. To get a sense of scale: the median length of the human protein is about 300 amino acids. There are 20 standard amino acids, plus a few uncommon amino acids. The different physiochemical properties of the amino acids ultimately determine the physiochemical properties of the protein, and thereby its functions and capabilities. Of great importance is the *order* of the amino acids in the protein. Thus, for a protein of median length there are 20^{300} theoretical different proteins possible ($=\sim 10^{390}$, a number much greater than the number of proteins which will ever be generated in the lifetime of the universe across all living cells). However, the vast majority of these theoretical combinations would yield non-functional proteins, and if they were to be generated in a cell, the cell would readily get rid of them (by the proteasome, a structure composed of other proteins!). Nevertheless it should be clear that even with “only” 20 primary building blocks there is substantial room for diversity in the physiochemical properties of a protein. And this diversity is readily observed in the different functions that different proteins can have: through enzymatic activity proteins can convert catalyze a great spectrum of metabolic reactions, or act on other proteins, polymerize other biomolecules (such as DNA and RNA, thereby allowing cells to be self-replicating), and have structural functions important for the formation of extra-cellular matrix, or larger components of the human body such as hair and bone[‡].

[†] In my view this is a dazzling feat considering the fact that all cells in the body, with its many cell types, originate from a single cell

The genetic code as blueprint for proteins

The exact sequence of amino acids in a protein is ultimately determined by the genome of the cell, recorded in its DNA (see also Figure 1). The blueprint of a protein is stored in the genome in a unit referred to as a gene. Human cells which are diploid have two copies of most genes, a consequence of the merging of two gametes (sperm/egg cell) at the conception of a new human. The two copies are in stark contrast with the many copies that can exist of a protein in a cell at a given time (>1000 protein copies per cell for some abundant proteins). This *amplification of expression* is achieved by a two-step process. The first step is transcription, where the DNA is expressed in the form of RNA, a macromolecule sharing many physiochemical properties with DNA.

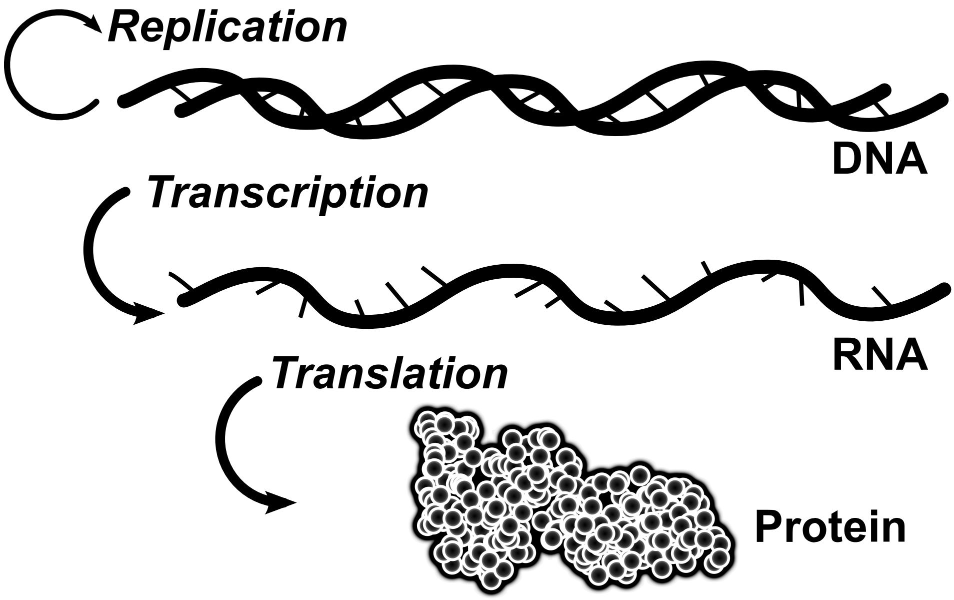


Figure 1: Central dogma of biology
A cell's genetic information is stored in its DNA. By replication of its DNA it can copy its hereditary content and spread this over two daughter cells, allowing cells to multiply. When genes come to "expression" the flow of genetic information follows two steps: (1) transcription, in which RNA is transcribed from the DNA and (2) translation, where the RNA is translated into protein.

However, unlike DNA, it can be moved out of the nucleus, into the cells' cytoplasm. It is there that the second process takes place: translation. At this step the *messenger RNA* (mRNA) serves as template for the protein, and the order of the bases in the mRNA will dictate the order in which the amino acids will be placed in the protein. The translational machinery responsible for translation of mRNA into protein consists of many parts. The most voluminous parts are the ribosomes, themselves composed of two major subunits (small and large ribosomal subunits), which are again built up from smaller parts (ribosomal RNA and ribosomal proteins). The ribosomes assemble on the mRNA which they will translate, and thread over it, reading its sequence while simultaneously building an amino acid chain, which folds up into a protein co-translationally. Of importance here is that while transcription of DNA into mRNA occurs into a similar alphabet (4 symbols in both DNA and RNA), translation of mRNA into protein has to be more complex, indicated by the existence of 20 standard amino acids, the symbols in the alphabet in the protein sequence. Indeed, to specify one of the 20 amino acids, 3 mRNA nucleotides (the symbols in RNA) are used. A unit of such 3 nucleotides is referred to as a *codon*. The mapping of codons to amino acids is determined by transfer-RNAs (tRNAs), specialized molecules that carry an *anti-codon* and an amino acid moiety. They are used by ribosomes to supply the amino acids for the growing chain. The ribosome binds the tRNAs and checks whether the anti-codon of the tRNA matches with the codon in the mRNA. A proper match will cause the ribosome to catalyze the transfer of the amino acid from the charged tRNA to the growing amino acid chain. After the incorporation, elongation factors (eEF-2 in human cells) move the ribosome one codon downstream on the mRNA, so that it will face the next codon, and the translation of mRNA can continue (for a comprehensive introduction to and overview of translation, see (Lehninger et al., 2005), chapter 27).

[‡] Outside the scope of this thesis are post-translational modifications, which further diversify the range of functional proteins which a cell has at its disposal

There are some consequences of a two-step process for the expression of genes into proteins. First, amplification is possible, both at the step of transcription (many mRNA molecules can be generated from a single gene) and the translation step (many protein molecules can be generated from a single mRNA molecule). Secondly, it allows for more opportunities of regulation of expression, as each step can be controlled. Thirdly, since the mRNA molecules are free to diffuse in the cytoplasm, in contrast to the DNA which is constrained to the nucleus, translation of the mRNAs can be localized to specific subcellular compartments (organelles) in the cell, which happens for example with many mitochondrially-localized proteins.

Studying gene expression; from DNA to RNA to protein

The study of gene expression, from DNA to RNA and RNA to protein, has been, and still is, of great importance for the understanding of the functioning of cells. Developments in the mid-90s paved the way to profile abundances of RNA molecules using *microarrays*, a technology that became immensely popular and transformed how molecular biology research was done. More recently, a technique with very different technical foundations known as *high-throughput sequencing* or *deep-sequencing* is enabling researchers to get insight in the abundances of RNA (or DNA) in their samples (see also Box 1 and Figure 2). Exhaustive, genome-wide (or transcriptome-wide) techniques profiling the DNA and/or RNA of a sample are intrinsically easier than techniques that attempt to exhaustively profile the proteins in a sample. In my view, two reasons are at the core of this difference: first, the alphabet of proteins has 20 common symbols, while RNA and DNA have only 4 common symbols. Secondly, molecular tools are available to *reverse transcribe* RNA into DNA, and to *replicate* DNA (i.e. make two identical DNA molecules from a single template). These molecular tools, together with clever chemistry, open the possibility for the profiling techniques: to amplify RNA or DNA, and to attach a signal to its replication (such as the release of a fluorophore). Naturally, this pushed exhaustive, genome-wide techniques towards the profiling of the transcriptome (the mRNA), while exhaustive profiling of a cells' proteins (the proteome) remained difficult or impossible. The current state of the art techniques to measure protein levels by mass spectrometry are orders of magnitudes less sensitive than techniques that profile transcript levels.

Box 1 – A few words on sequencing

Deep sequencing and next-generation sequencing are two terms that are often interchangeably used to refer to current state-of-the-art DNA sequencing techniques with high throughput. While next-generation sequencing is a catch-all term to refer to modern sequencing techniques, deep sequencing in fact refers to the sequencing of a DNA sequence at high coverage, i.e. having repeatedly sequenced the same sequence at least a couple of times. Deep sequencing is valuable and necessary in order to sequence novel genomes (e.g. bacterial genomes) or to confidently determine single nucleotide polymorphisms (SNPs) or mutations in samples. The state-of-the-art methods are in contrast with the conventional chain termination method of sequencing (“Sanger sequencing”) which is low-throughput and labor intensive. Next-generation sequencing methods are both qualitative (they determine the order and identity of nucleotides of DNA fragments in the sample with high fidelity) and quantitative (they allow the concentration of a DNA fragment in a sample to be measured). Several next-generation sequencing technologies have been developed and/or refined over the past decade, however here I discuss only Illumina (sometimes referred to as “Solexa”) sequencing, which is the technique used throughout this thesis. Advantages of Illumina sequencing are high throughput generation of data, at relative low costs. Furthermore its popularity and “network effects” amongst users have made reagents, kits and other experimental materials adapted to Illumina sequencing ubiquitous and cost-effective.

The sequencing reaction requires a sample containing a so-called “library”: fragmented DNA molecules which contain special sequences (adapter sequences) at their termini. Upon loading of the sample onto the flowcell (a disposable but integral part of the Illumina platform) the molecules in the library attach themselves to the bottom of the flowcell by binding of the adapter sequences to molecules that form a “lawn” at the surface of the flowcell. After attachment of the DNA molecules to the flowcell, the DNA molecules are multiplied by “bridge amplification”, a special twist on PCR that keeps the amplified DNA molecules attached to the flowcell and physically close to the original DNA molecule (see Figure 2). The bridge amplification is possible due to the presence of adapters on the DNA molecules in the library and the surface of the flowcell. After bridge amplification, many

(~1000) identical DNA molecules are clustered together on the flowcell. Finally this allows the sequencing reaction, which takes place by a final round of synthesis of a new DNA strand. The reaction takes place in rounds, in which distinct dye-labeled versions of the four nucleotides are added. Incorporation of the nucleotides can be registered by a high-resolution microscope. At this point all molecules in the cluster (~1000 molecules) emit the same color, which shows that the bridge amplification serves as a signal amplification step allowing the microscope to detect nucleotide incorporation. Image processing is used to identify the clusters and determine the identity of the nucleotide that was incorporated at each cluster by the color that was emitted. The unincorporated nucleotides are washed away and new dye-labeled nucleotides are added. The process repeats in many more cycles to identify the order of the next nucleotides, on nucleotide per cycle. The high-throughput nature of Illumina sequencing stems from the fact that DNA molecules are small, and a large number of clusters can be generated on a flowcell (in the order of a million per square millimeter). A comprehensive review of Illumina sequencing and other next-generation sequencing techniques can be found in (Metzker, 2010). Since Illumina (and other next-generation sequencing techniques) sequence the bases in DNA, transcriptome profiling (where RNA instead of DNA sequences are determined and quantified) requires some extra experimental procedures. The key step is “reverse transcription” where RNA is generated back into DNA by a *reverse transcriptase* enzyme (first isolated from retroviruses).

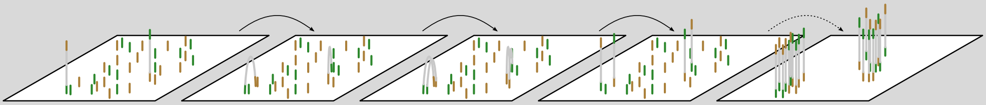


Figure 2: Illumina sequencing

Diagram of bridge amplification and cluster formation. A library with DNA fragments (grey) harboring special adapter sequences (green and brown) is dispersed over the flowcell's surface (white). The surface of the flowcell contains molecules to which the DNA fragments can attach. Important is that both ends of the DNA fragments can attach to the surface, allowing “bridging” of the fragments. Performing several rounds of polymerase chain reaction (PCR) amplification causes dense clusters of DNA fragments with identical sequence to be formed.

Other essential experimental steps include adapter ligation (for attachment of the sequences to the flowcell and to allow bridge amplification) and usually PCR amplification (to multiply the material in cases where too few molecules are present in the sample to saturate the flowcell), but they are common in all sequencing library preparation protocols, and not unique to transcriptome profiling. Finally, ribosome profiling, the central technique in this thesis, is an adaptation of transcriptome profiling, where not the full transcriptome is profiled but only those RNA fragments which are occupied by ribosomes. These ribosome protected fragments (RPFs) are extracted by digestion of unprotected RNA, sedimentation of the *monosomes* (single ribosomes plus RPF) and separation and isolation of the RPF, after which a sequencing library of the RPFs can be made analogously to transcriptome profiling.

Ribosome profiling has emerged as a technique to investigate the level of translational regulation in a fashion similar to transcriptome profiling. The experimental protocol (Figure 3) generates ribosome protected fragments (RPFs) by digestion of unoccupied mRNA and isolation of the remaining fragments, which correspond to the mRNA positions occupied by ribosomes. As a quantitative technique, it allows for the measurement of translation rate of

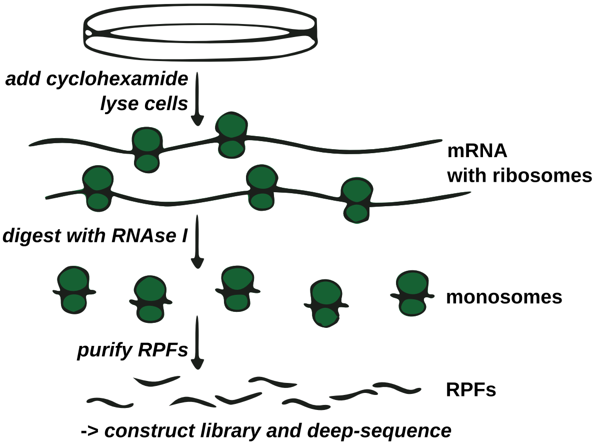


Figure 3: Summary of the ribosomal profiling protocol
Cyclohexamide is added to cells to block movement of ribosomes on the mRNA and prevent drop-off. The ribosomes and mRNA molecules are isolated together. mRNA regions that are unoccupied by ribosomes are vulnerable to digestion by RNAse, but ribosome-protected mRNA regions are not. Thus, when treated with RNAse I, single ribosomes occupying short fragments of mRNA are obtained. These so-called monosomes are isolated together with the short mRNA fragments (named ribosome protected fragments or RPFs). Finally, the RPFs are isolated from which a library for deep-sequencing is constructed.
Figure adapted from (Guo et al., 2010).

a gene, an important factor in determining translational regulation. In contrast, proteome profiling by e.g. mass spectrometry measure the actual levels of protein, which are determined by both the translation (production) rate and the rate of degradation of the protein. As such, ribosome profiling does not provide an alternative to proteome profiling, but rather the two augment each other, as ribosome profiling provides insight in the translational process. Ribosome profiling led to the identification of several translational programs (programs where translation of mRNAs is differentially regulated) (Gonzalez et al., 2014; Stumpf et al., 2013). Furthermore, open reading frames upstream of the canonical protein-coding open reading frame in an mRNA (uORFs) have been identified as regulatory components of mRNA translation by attracting ribosomes (Ingolia et al., 2011). Ribosome profiling has fully opened the door to investigate the level of translational regulation in cancer and tumorigenesis. Research on the level of transcriptional regulation has been fruitful in this area, as cells exhibit distinct patterns of gene expression when they are or become malignant. These transcriptional changes have been a cornerstone in cancer research and a starting point for understanding of cellular processes that underlie tumorigenesis, development of biomarkers and providing handles for therapeutics. The level of translational regulation has remained largely unexplored. Ribosome profiling allows us to close this gap in understanding, and investigate whether cancer cells exploit levels of translational regulation to facilitate their growth and replication, escape of cell death and other hallmarks of cancer (Hanahan and Weinberg, 2011). Would there be, for instance, key proteins that act as proto-oncogenes or tumor suppressors whose translation is altered to give cancer cells a proliferative advantage?

miRNAs as potent regulators of gene expression at the mRNA level

miRNAs are small RNA molecules acting on other RNAs by sequence complementarity as post-transcriptional regulators of gene expression. Their discovery (Lee et al., 1993) was an overture to the discovery of a completely novel level of gene regulation, as most mRNAs were found to be regulated by miRNAs (Friedman et al., 2009). Their mode of gene repression was initially thought to involve repression of translation of the mRNA without decreasing the abundances of the mRNAs (Wightman et al., 1993). However, transcriptome-wide analyses enabled by microarrays later indicated that the mechanism of action of miRNAs is through decreasing mRNA abundances (Giraldez et al., 2006; Lim et al., 2005; Rehwinkel et al., 2006). Ribosome profiling has also been employed to distinguish the mode of gene repression elicited by miRNAs (Guo et al., 2010) and again indicated that mRNA destabilization is the major consequence of miRNA-mediated repression. Work of Bazzini et al. showed that temporal effects may be important in the interpretation of the ribosome profiling data, as the authors claimed that translational repression precedes transcriptional destabilization for miRNA *miR-430* in zebrafish (Bazzini et al., 2012). While delineation of the order of events following miRNA-mRNA hybridization is important, a phase of translational repression is to be expected and the impact of the short duration of this phase on final protein output (and thus gene expression) is unsure[†].

Mechanisms for regulation of translation

Upstream open reading frames (uORFs) have already been identified as potent regulators of translation of the downstream canonical protein-coding ORF (Calvo et al., 2009). The observation that they attract ribosomes and yield RPFs by ribosome profiling is an interesting observation that opens the possibility to further investigate the regulation of (and by) uORFs. The uORFs reduce translational rate of downstream ORF by detracting ribosomes, which when occupying uORFs require shunting or reinitation to translate the downstream ORF (Ryabova and Hohn, 2000). Another mechanism of translational regulation is the association of the translation initiation complex (eIF4) with the mRNA, that is required to load the pre-initiation complex, of which the small ribosomal subunit is a major part, onto the mRNA. The initiation complexes consist of many subunits of which many are known to be regulated by post-translational modifications such as phosphorylations. Furthermore the initiation steps depend on several factors that are present in the mRNA, such as the capped 5' guanine, sequence elements in the 5'-region of the mRNA (the 5'TOP sequence (Avni et al., 1997)) and structural elements in the mRNA (e.g. IRESs (Pelletier and Sonenberg, 1988)). Also RNA-binding proteins (RBPs) are known to regulate the translational activities of mRNAs (Ascano et al., 2012). It is clear that cells have developed regulatory mechanisms to control translation of mRNAs. Ribosome profiling provides the most direct method to investigate these mechanisms, and is to some extent capable of distinguishing between the mechanisms of regulation that act on mRNAs.

[†] postscriptum: We have investigated whether patterns of miRNA-mediated translational repression could be identified in the ribosome profiling datasets we have generated, but found no indications consistent with translational repression elicited by miRNAs. The topic of miRNAs is therefore not recurring in this thesis.

Regulation of amino-acid and translational machinery components

Besides regulation of mRNA translation by differential initiation rates, we expect that translational efficiencies will also depend on amino acid availabilities, as amino acids form essential components of proteins. Ribosomes depend on tRNAs charged with amino acids to elongate the peptide chain and progress translation of the mRNAs. When amino acid availabilities fluctuate, activated-tRNA availabilities may also fluctuate. Note that other components such as the tRNA levels themselves and activity levels of the aminoacyl-transferase proteins (that catalyze the activation of the tRNA) will elicit indistinguishable responses. Nevertheless, if ribosome profiling is able to provide a view of regulation at the amino acid and/or tRNA level, it would be another valuable tool in the study of cancer cells and tumorigenesis. Metabolic rewiring and differential requirements of amino acids of cancers has been long known (Broome, 1963) and this observation is becoming increasingly of interest as therapeutic window (Vander Heiden, 2011). We hypothesize that ribosome profiling will accelerate the discovery of therapeutic opportunities by pinpointing metabolic rewiring and resulting differential nutrient requirements.

Inferring ribosome dimensions from ribosome profiling data

Important in the interpretation of ribosome profiling data to detect differential amino acid (or tRNA) availabilities are the dimensions of the ribosome. Initial data of ribosome profiling (Ingolia et al., 2009) shows that RPFs are around 30nt. in size, and the RPFs are predominantly abundant 12 nt. upstream of initiation START codons and show periodic abundance downstream of this position. From these observations the relative positions of tRNA binding pockets within the RPFs can be determined (Figure 4). Thus, owing to the nucleotide resolution of the data, ribosome profiling yields a snapshot of which ribosomal pockets are occupying which codons, in the population of ribosomes in a sample. The chapters in this part of the thesis build on these observations in order to analyze differential activated tRNA availabilities in samples, possibly caused by metabolic rewiring. Several key issues will be addressed: can ribosome profiling be used to investigate the metabolic state of cells in a sample? If it is technically possible to infer differences in metabolic state, can ribosome profiling be used to uncover novel molecular targets or biomarkers for cancer therapies?

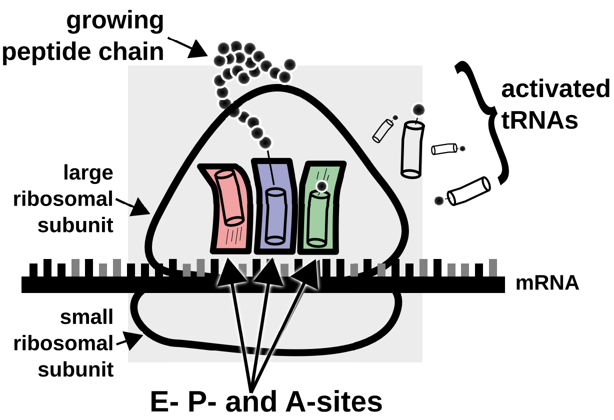


Figure 4: Schematic overview of the translational machinery Ribosomes comprise a small and large subunit, each consisting of several RNA and protein subunits. Together they protect a chunk of mRNA of about 30nt. (not faithfully shown in figure), known as the ribosome protected fragment (RPF). In order to translate mRNA into protein, the ribosomes use activated tRNAs, that deliver amino acid moieties. The structure of the large ribosomal subunit has three distinct pockets, or tRNA binding sites. The A-site (green) allows tRNAs to enter the ribosome. When the anti-codon of the tRNA matches the codon in the mRNA, the ribosome moves 3nt (one codon) over the mRNA, and by doing so moves the tRNA into the P-site (blue). This is also accompanied with the conjugation of the amino acid onto the growing peptide chain. The tRNA in the P-site carries the peptide chain. As the cycle repeats, the tRNA occupying the P-site is moved to the E-site (red), which allows exit of the tRNA from the ribosome.

rRNA contamination in ribosome profiling data hampers its sensitivity

The ribosome profiling protocol involves RNA digestion in order to obtain RPFs and map ribosomal positions with nucleotide resolution. Because the ribosomes themselves are complexes of protein and ribosomal RNA (rRNA), they are also prone to RNA digestion. This causes certain rRNA fragments to become a common and highly abundant contaminant in current ribosome profiling datasets (see Figure 5). The high abundance of this contamination can reduce the sensitivity of the obtained data because it limits the amount of useful reads that the dataset contains (the RPFs, derived from mRNA fragments). However, mapping of the contaminating rRNA fragments on a structure of the ribosome reveals that the most abundant rRNA fragments are derived from the exterior of the ribosome (Figures 5c and 5d). Indeed, within the contaminating reads the sequence diversity is small, which may prove useful to counteract the contaminating reads.

Outline of part A

This part of the thesis explores control of translation using ribosome profiling, a technique which maps the positions of ribosomes on mRNAs with nucleotide resolution. Chapter 2 highlights the consequences of a mitochondrial tRNA mutation on translation and ribosome progression. Using a cellular system where a mitochondrial mutation is introduced independent of the nuclear genetic background we show that mitochondria harboring a tRNA-Trp$^{5556G>A}$ mutation yield more ribosomes stalling with Trp codons in their A-site, indicating the reduced availability of the tRNA. Chapter 3 extends the concept of differential availabilities of tRNAs further upstream by investigation of the consequences of differential amino acid availabilities. Furthermore, it takes ribosome profiling to a clinical

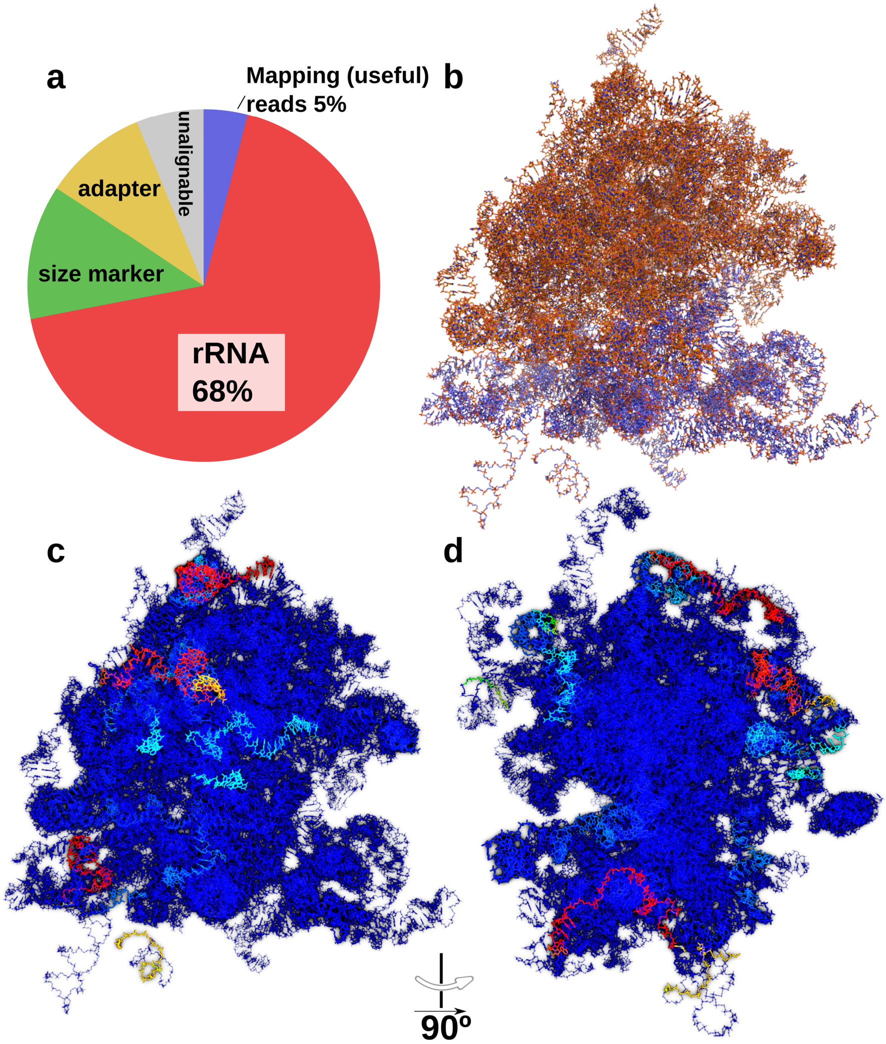


Figure 5: Contamination issues common in ribosome profiling data
Because ribosomes are large macromolecues themselves comprised of RNA, they are vulnerable to RNAse I digestion. Some of the RNA content of ribosomal RNA (rRNA) ends up as contaminant in the deep-sequencing libraries. Shown here is a meta-analysis of a replicate of the published dataset with GEO ID GSM1047584. **(a)** From the total number of reads, over two-thirds were found to be rRNA contamination. Other contaminants (size marker, adapter sequence) are also abundant but not as pervasive as rRNA. The contamination limits the fraction of data which is informative (actual RPFs of mRNA), in this case to approx. 5% of the raw data. **(b)** The structure of the nuclear-encoded ribosome of Saccharomyces cerevisiae. Only ribosomal RNA is shown. The large subunit has carbon atoms shown in orange, the small subunit has carbon atoms shown in blue. PDB entries used: 3IZE and 3IZF. **(c, d)** Occurrence of rRNA fragment as contaminant in the ribosomal profiling sample. Dark blue indicates that the rRNA has not been found in the raw data. Bright red indicates the most abundant contaminating rRNA fragments. Orientation of the structure in (c) is as in (b).

setting in order to pinpoint differential nutrient availabilities, by performing ribosome profiling on kidney tumor samples. In chapter 4 the level of translational regulation during mitosis is investigated. Again ribosome profiling is used, and coupled to RNA sequencing data, to determine translational efficiencies of mRNAs. Furthermore, the tools we developed thus far are used to identify differential amino acid and/or tRNA availabilities.

References

Ascano, M., Mukherjee, N., Bandaru, P., Miller, J.B., Nusbaum, J.D., Corcoran, D.L., Langlois, C., Munschauer, M., Dewell, S., Hafner, M., et al. (2012). FMRP targets distinct mRNA sequence elements to regulate protein expression. Nature *492*, 382–386.

Avni, D., Biberman, Y., and Meyuhas, O. (1997). The 5′ Terminal Oligopyrimidine Tract Confers Translational Control on Top Mrnas in a Cell Type-and Sequence Context-Dependent Manner. Nucl. Acids Res. *25*, 995–1001.

Bazzini, A.A., Lee, M.T., and Giraldez, A.J. (2012). Ribosome profiling shows that miR-430 reduces translation before causing mRNA decay in zebrafish. Science *336*, 233–237.

Broome, J.D. (1963). Evidence that the L-asparaginase of guinea pig serum is responsible for its antilymphoma effects. J Exp Med *118*, 99–120.

Calvo, S.E., Pagliarini, D.J., and Mootha, V.K. (2009). Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. PNAS *106*, 7507–7512.

Friedman, R.C., Farh, K.K.-H., Burge, C.B., and Bartel, D.P. (2009). Most mammalian mRNAs are conserved targets of microRNAs. Genome Res. *19*, 92–105.

Giraldez, A.J., Mishima, Y., Rihel, J., Grocock, R.J., Van Dongen, S., Inoue, K., Enright, A.J., and Schier, A.F. (2006). Zebrafish MiR-430 promotes deadenylation and clearance of maternal mRNAs. Science *312*, 75–79.

Gonzalez, C., Sims, J.S., Hornstein, N., Mela, A., Garcia, F., Lei, L., Gass, D.A., Amendolara, B., Bruce, J.N., Canoll, P., et al. (2014). Ribosome profiling reveals a cell-type-specific translational landscape in brain tumors. J. Neurosci. *34*, 10924–10936.

Guo, H., Ingolia, N.T., Weissman, J.S., and Bartel, D.P. (2010). Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature *466*, 835–840.

Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: the next generation. Cell *144*, 646–674.

Vander Heiden, M.G. (2011). Targeting cancer metabolism: a therapeutic window opens. Nat Rev Drug Discov *10*, 671–684.

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

Ingolia, N.T., Lareau, L.F., and Weissman, J.S. (2011). Ribosome Profiling of Mouse Embryonic Stem Cells Reveals the Complexity and Dynamics of Mammalian Proteomes. Cell *147*, 789–802.

Lee, R.C., Feinbaum, R.L., and Ambros, V. (1993). The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell *75*, 843–854.

Lehninger, A.L., Nelson, D.L., and Cox, M.M. (2005). Lehninger principles of biochemistry (New York: W.H. Freeman).

Lim, L.P., Lau, N.C., Garrett-Engele, P., Grimson, A., Schelter, J.M., Castle, J., Bartel, D.P., Linsley, P.S., and Johnson, J.M. (2005). Microarray analysis shows that some microRNAs downregulate large numbers of target mRNAs. Nature *433*, 769–773.

Metzker, M.L. (2010). Sequencing technologies — the next generation. Nat Rev Genet *11*, 31–46.

Pelletier, J., and Sonenberg, N. (1988). Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. Nature *334*, 320–325.

Rehwinkel, J., Natalin, P., Stark, A., Brennecke, J., Cohen, S.M., and Izaurralde, E. (2006). Genome-wide analysis of mRNAs regulated by Drosha and Argonaute proteins in Drosophila melanogaster. Mol. Cell. Biol. *26*, 2965–2975.

Ryabova, L.A., and Hohn, T. (2000). Ribosome shunting in the cauliflower mosaic virus 35S RNA leader is a special case of reinitiation of translation functioning in plant and animal systems. Genes & Development *14*, 817–829.

Stumpf, C.R., Moreno, M.V., Olshen, A.B., Taylor, B.S., and Ruggero, D. (2013). The Translational Landscape of the Mammalian Cell Cycle. Molecular Cell *52*, 574–582.

Wightman, B., Ha, I., and Ruvkun, G. (1993). Posttranscriptional regulation of the heterochronic gene lin-14 by lin-4 mediates temporal pattern formation in C. elegans. Cell *75*, 855–862.

CHAPTER 2

Ribosome profiling reveals features of normal and disease-associated mitochondrial translation

Koos Rooijers[1,*], Fabricio Loayza-Puch[1,*], Leo G. Nijtmans[2] and Reuven Agami[1,3]

1 Division of Gene Regulation, The Netherlands Cancer Institute
Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands
2 Nijmegen Center for Mitochondrial Disorders
Geert Grooteplein Zuid 10, 6500 HB Nijmegen, The Netherlands
3 Erasmus MC, Rotterdam University,
Dr Molewaterplein 50, 3015 GE Rotterdam, The Netherlands
* These authors contributed equally to this work

Adapted from Rooijers et al., Nature Communications 4, article number: 2886, doi:10.1038/ncomms3886

Abstract

Mitochondria are essential cellular organelles for generation of energy and their dysfunction may cause diabetes, Parkinson's disease and multi-systemic failure marked by failure to thrive, gastrointestinal problems, lactic acidosis and early lethality. Disease-associated mitochondrial mutations often affect components of the mitochondrial translation machinery. Here we perform ribosome profiling to measure mitochondrial translation at nucleotide resolution. Using a protocol optimized for the retrieval of mitochondrial ribosome protected fragments (RPFs) we show that the size distribution of wild-type mitochondrial RPFs follows a bimodal distribution peaking at 27 and 33 nucleotides, which is distinct from the 30-nucleotide peak of nuclear RPFs. Their cross-correlation suggests generation of mitochondrial RPFs during ribosome progression. In contrast, RPFs from patient-derived mitochondria mutated in tRNA-Tryptophan are centered on tryptophan codons and reduced downstream, indicating ribosome stalling. Intriguingly, long RPFs are enriched in mutated mitochondria, suggesting they characterize stalled ribosomes. Our findings provide the first model for translation in wild-type and disease-triggering mitochondria.

Introduction

The primary function of the mitochondrion is ATP production via the oxidative phosphorylation pathway (Smits et al., 2010). The human mitochondrial genome contains 37 genes of which 13 are protein-coding. These genes are essential for normal energy production function of the mitochondria but also other processes such as cell signaling and cell death (Koc and Koc; Suen et al., 2008). The mitochondria harbor their own ribosomes, which are structurally more similar to bacterial ribosomes than the cytosolic ribosomes (Sharma, 2003). However, the protein-synthesizing system of mitochondria and the template mRNAs contain unique features not observed in prokaryotes or the eukaryotic cell cytosol, including arginine codons AGG and AGA, which are unassigned while UGA stop codon codes for tryptophan; both AUG and AUA serve as START codons; messenger RNAs (mRNAs) essentially lack untranslated regions, are uncapped and contain a poly(A) tail, and mitochondria use a simplified decoding mechanism that allows translation of all codons with only 22 tRNAs (Smits et al., 2010). Because of their function in the oxidative phosphorylation pathway, mitochondria are recognized for their role in the production of reactive oxygen species (ROS) and nitric oxide (Navarro and Boveris, 2007). The production of ROS may also be associated with aging since mitochondria increase production of ROS during human lifespan (Lin and Beal, 2006). Interestingly, mutations in mitochondrial translation machinery, such as their tRNAs, are the source of various human diseases, typically causing multi-systemic disorders and early fatality (Scharfe, 2009).

To characterize mitochondrial translation we used ribosome profiling, a deep sequencing-based technology that allows quantitative analysis of translation at nucleotide resolution (Ingolia et al., 2009, 2012). Ribosome profiling maps the positions of ribosomes on transcripts by nuclease footprinting, generating nuclease-protected mRNA fragments bound to ribosomes that are converted into a DNA library suitable for deep sequencing. We observed a poor capacity of the established ribosome profiling protocol to capture mitochondrial ribosome protected fragments (RPFs) and developed a suitable method for this purpose. We then used the adapted protocol to show that the mitochondrial tRNA(Trp)$^{5556G>A}$ mutation induces strong ribosome stalling.

Results

Mitochondrial RPFs are poorly captured by current protocols

The abundance of RPFs was demonstrated to correlate with the level of translation of a gene (Ingolia et al., 2009). Figure 1a shows the translational efficiency (TE; normalized abundance measured by ribosome profiling divided by normalized abundance measured by RNA sequencing) of nuclear- and mitochondrial-encoded genes. Curiously, all mitochondrial transcripts (marked green) showed a high level of signal in the RNA-seq while having a very low amount of reads in the ribosome profiling data compared with the bulk of the nuclear transcripts, which had TE values centered on the norm and up to fourfold difference in the vast majority of the cases, resulting in relatively low TE values for mitochondrial genes. A similar observation was seen when we used several recently published ribosome profiling data sets (Liu et al., 2013; Shalgi, 2013) (see Supplemental Fig. S1 and Supplemental Table S1). In contrast, histone genes, which are not polyadenylated and therefore are poorly retrieved during mRNA sequencing, showed as expected relatively extremely high TE values (marked red in Fig. 1a). Two potential explanations can clarify the low detection level of mitochondrial genes in the ribosome profiling data.

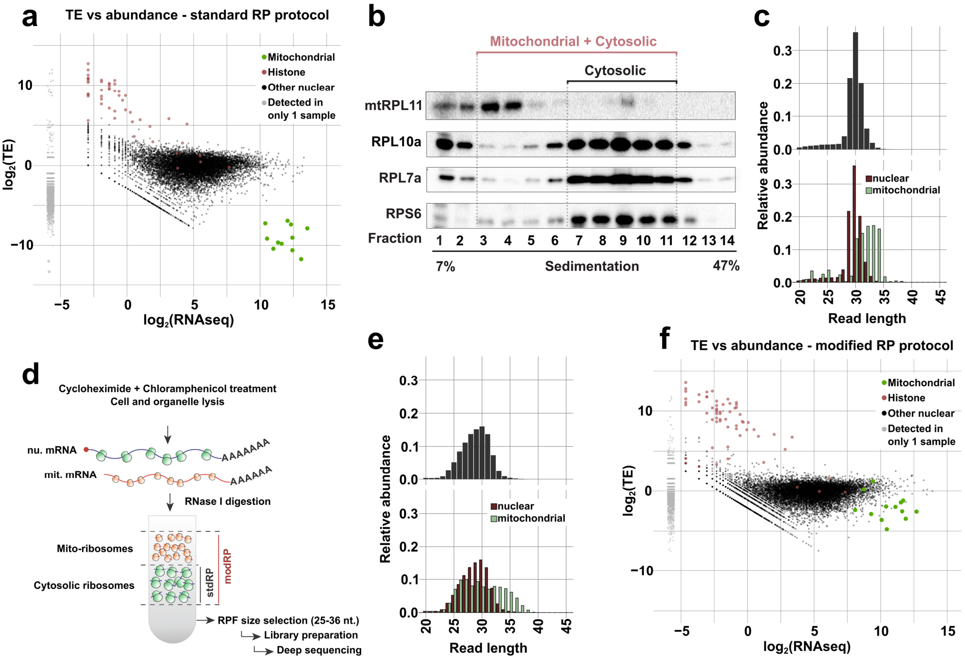


Figure 1: A modified ribosomal profiling protocol improves detection of mitochondrial translation.
(a) A plot showing the TE as a function of mRNA abundance in a standard ribosomal profiling protocol. Mitochondrial and histone genes (in green and red, respectively) appear as outside groups. **(b)** Western blots showing the abundance of mitochondrial (mtRPL11) and cytosolic (RPL10a, RPL7 and RPS6) ribosomal proteins in different sucrose gradient fractions after RNAse I treatment. **(c)** Size distribution of RPFs in the standard ribosome profiling. The upper panel shows all reads, the lower panel shows nuclear-encoded genes (red) and mitochondrially encoded genes (green) separately. **(d)** Overview of our modified ribosome profiling protocol. **(e)** Size distribution of RPFs obtained by our modified ribosome profiling. **(f)** Same as in (a) for modified protocol. Marked improvement in the detection of TE in mitochondrial genes is observed.

First, mitochondrial genes might be translated at a remarkably different rate compared with the cytosolic mRNAs. Alternatively, the ribosome profiling protocol does not efficiently capture mitochondrial RPFs as a result of either a distinct size of the mitochondrial ribosomes or a distinct size of RPFs protected by the mitochondrial ribosomes, compared to the nuclear-ribosomes. As mitochondrial ribosomes are very different from the nuclear ones, and mitochondrial proteins are produced in high rate, inefficient capturing of mitochondrial RPFs the RP protocol seems more likely. Therefore, we first determined the fractions where mitochondrial monosomes localize after RNAse I treatment and sucrose gradient sedimentation, and indeed this analysis indicated altered sedimentation (Fig. 1b). Second, we also examined the size of RPFs in mitochondrial- and nuclear-encoded genes. Figure 1c shows that while nuclear genes feature the expected enrichment of 30-nt long RPFs, the size distribution of mitochondrial RPFs is clearly larger and seems truncated. As the ribosome profiling protocol we used included a selection step for fragments sized 27–33 nt, this result suggests that most of the mitochondrial RPFs may have a different size and thus are vastly excluded during this selection step.

Improving retrieval of mitochondrial RPFs

We modified the ribosome profiling protocol to collect more mitochondrial ribosomes in the sucrose gradient and relaxed the size of RPFs cutoff to 25–36 during size excision of adapter ligated reads (Fig. 1d). Indeed, Fig. 1e demonstrates that this protocol results in RPFs whose global size distribution coincides with the expected expansion in fragments cutoff. Moreover, it also shows that while nuclear RPFs showed a peak at 30 nt as expected, a much broader distribution of the mitochondrial RPFs with a bimodal size peaking at 27 and 33 nt was observed. Last, Fig. 1f shows that our modified ribosome profiling protocol indeed improved the apparent TE of mitochondrial mRNAs to levels comparable to the nuclear-encoded mRNAs without compromising the quality of libraries. As shown in Fig. 2a, the reproducibility between RP samples is high (R^2 0.91), comparable to RNA-seq data (Fig. 2b). The data shows periodicity (Fig. 2c) and the modifications in the protocol do not lead to increase of ribosomal RNA contamination (Fig. 2d). Lastly, as a characteristic of ribosome profiling data, the vast majority of reads originate from CDSs and a steep increase in RPF density is observed at the START codons, and similarly, a steep decrease is observed at the STOP codons (Fig. 2e).

Mitochondrial RPF length shows bimodality

The apparent bimodal distribution of mitochondrial RPFs length, peaking at 27 and 33 nt, may indicate alternative mitochondrial ribosome configurations or different positioning on mRNA. To examine this point we divided the mitochondrial RPFs into short (24–29 nt) and long (31–36 nt) fragments and measured their distribution over the mitochondrial transcripts. Figure 3a shows an overall similar pattern of distribution between the two types of mitochondrial-RPFs, both were distributed similarly along the transcripts, though some

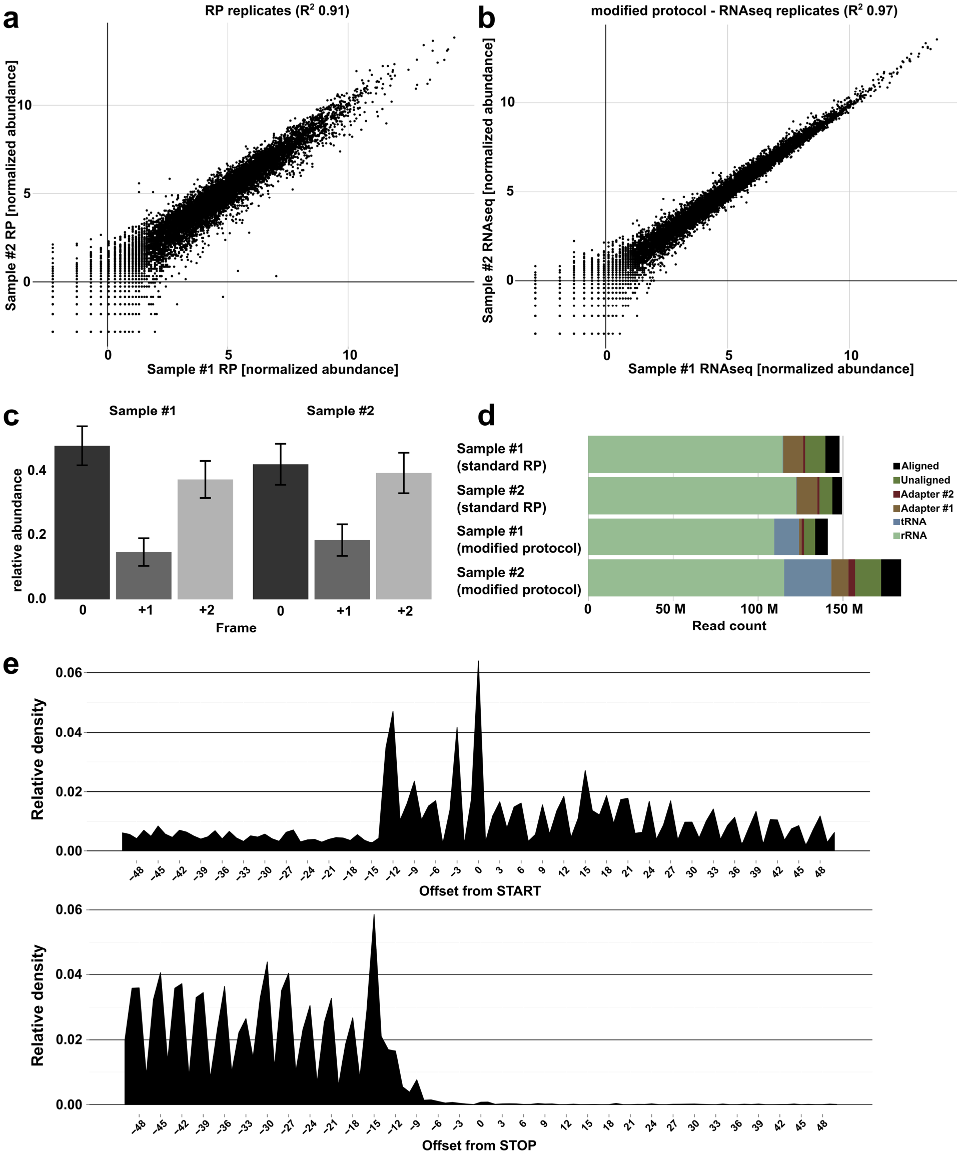


Figure 2: Reproducibility and quality control of modified ribosome profiling protocol. **(a)** Correlation between ribosome profiling data set replicates in the modified ribosome profiling protocol. **(b)** Correlation between RNAseq data sets in the modified ribosome profiling samples. **(c)** In-frame RPF abundance for nuclear-encoded genes in two modified ribosome profiling replicate samples. Error bars indicate standard deviations across the genes with at least 100 RPFs (n=5579 for sample #1 and n=7108 for sample #2). **(d)** Abundances of mRNA and contamination in the conventional and modified ribosome profiling protocol samples. **(e)** Coverage around START and STOP codons in modified ribosome profiling sample #1.

positions were reproducibly slightly more enriched with one type of fragment.

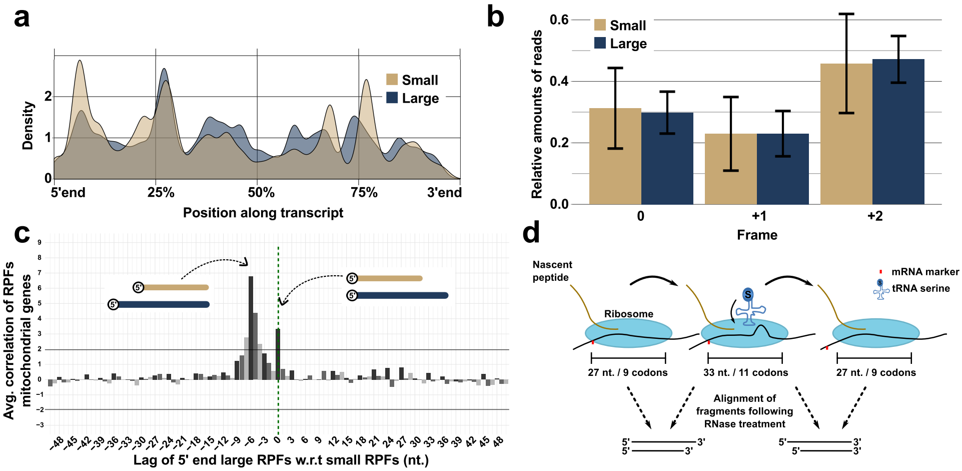


Figure 3: Size distribution of mitochondrial RPFs suggests mechanism of action. **(a)** The normalized abundance of short and long RPFs along the 13 protein-coding mitochondrial genes. **(b)** Normalized frame abundance of short and long RPFs on all mitochondrial protein coding genes. Error bars indicate s.d. across the mitochondrial protein-coding genes. **(c)** Correlation between the 27 nt length (short) and 33 nt length (long) mitochondrial RPFs averaged over all mitochondrial protein-coding genes (n=13) in the control sample #1. Error bars indicate s.d.'s. **(d)** A schematic model showing how the long and short fragments might be produced during mitochondrial ribosome procession.

Additionally, analyzing the orientation of mitochondrial-RPFs with respect to the START codon, we found no apparent periodicity difference between the short and long fragments (Fig. 3b). These results may indicate that the two mitochondrial-RPF types are either generated from the same ribosome during the translation process, or from two types of related ribosomes both engaged in active translation of the mitochondrial genes. Interestingly, if the first is the case, a global correlation between the positions of the two types of RPFs should exist. We therefore compared the correlation between the distribution of reads with exactly 27 and 33 nt in length and found two preferential relationships (Fig. 3c). The peak at -6 indicates an overhang of 6 nt at the 5′ end, while the peak at 0 indicates alignment of the 5′ ends and thus an overhang of 6 nt at the 3′ end. Though not entirely excluding the existence of two separate mitochondrial ribosome types, this result strongly supports the hypothesis that the two RPF types are generated from one mitochondrial ribosome during its progression in a manner depicted in Fig. 3d. We envision that during progression the mitochondrial ribosome encompasses additional 6 nt.

tRNA(Trp)$^{5556G>A}$ mutation causes ribosome stalling

Next, to learn more about the protein translation process in mitochondria and relate it to mitochondria malfunction, we utilized trans-mitochondrial cybrids, rho(0) cells that lack mitochondrial DNA and fused with patient-derived mitochondria (Smits, 2010). We performed our ribosomal profiling on cybrids containing either wild-type (wt) or the naturally occurring pathogenic mutant mitochondria at tRNA(Trp)$^{5556G>A}$, as this was shown to cause impairment of mitochondria function and reduce mitochondrial protein translation (Smits, 2010). Although mitochondria-targeted TagYFP expression showed no major abnormalities in terms of number of mitochondria between wt and mutant mitochondria (Supplemental Fig. S2a), mitochondrial activity assays (Table 1) and acidification of medium (Supplemental Fig. S2b) indicated reduced functionality, consistent with reported OXPHOS defects (Smits, 2010). The mutation was verified using tRNA sequences retrieved in the ribosome profiling and purity was >99% (Supplemental Fig. S5).

Table 1: Enzymatic activity of respiratory complexes I–V in the tRNA(Trp)$^{5556G>A}$ mutant cybrid and control cells.

	tRNA(Trp)5556G>A	**Control**
CI-Q	11.9	316.7 μU mU−1 CS
CI-D	17.9	501.7 μU mU−1 CS
CII	334.4	746.3 μU mU−1 CS
CIII	161.7	895.0 μU mU−1 CS
COX	43.4	333.4 μU mU−1 CS
SCC	122.8	406.1 μU mU−1 CS
CS	1000	1000.0 μU mU−1 CS

Intriguingly and most importantly, our ribosomal profiling method revealed a distinct distribution pattern of mitochondrial RPFs between wt and tRNA(Trp)$^{5556G>A}$ mitochondria (Fig. 4a). In particular, peaks of the tRNA(Trp)$^{5556G>A}$ mitochondria typically tended to be centered on tryptophan codons. Data analysis confirmed increased abundance of RPFs with the mitochondrial-tryptophan codons TGA and TGG (Fig. 4b). Moreover, RPF alignment analysis indicated that the tryptophan codons are centered in the middle of the RPFs, positions 15–18, implying stalling of the ribosome when encountering tryptophan codons. This effect was specific to mitochondrial tryptophan codons, as it was not observed in other codons (for example, the mitochondrial phenylalanine codon) or in nuclear-encoded tryptophan codons (Fig. 4c). Data from biological replicates showed that this was reproducible (Supplemental Fig. S3a). qRT–PCR measurements indicated comparable expression levels of the mutated tRNA(Trp)$^{5556G>A}$ and tRNA(Trp)wt in wt and tRNA(Trp)$^{5556G>A}$ mutant mitochondria (Fig. 4d, upper panel). Moreover, immune-precipitation analysis of mitochondrial ribosomes showed comparable enrichment levels (Fig. 4d, lower panel). This indicates that the 5556G>A mutation in mitochondrial tRNA(Trp) does not lead to rapid degradation and does not affect loading into the ribosomes. However, despite the dramatic changes observed in the distribution pattern of mitochondrial RPFs, the TE was not affected in the tRNA(Trp)$^{5556G>A}$ sample (Fig. 4e). Altogether, our observations indicate that the expression of tRNA(Trp)$^{5556G>A}$ enforces relocation of mitochondrial ribosomes to tryptophan codons.

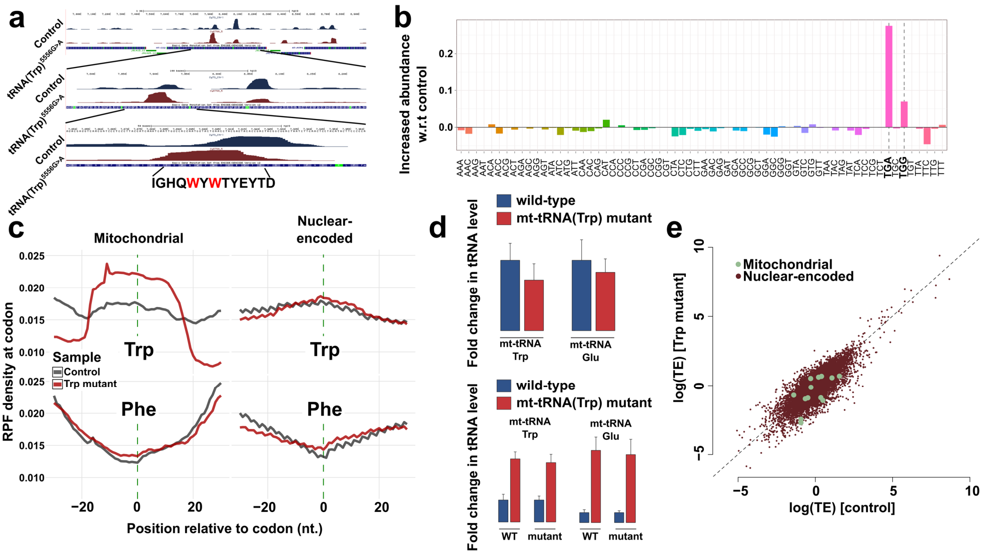


Figure 4: tRNA(Trp)$^{5556G>A}$ causes mitochondria malfunctioning and an increase of RPFs at tryptophan codons.

(a) Screenshots from the UCSC genome browser showing RPFs of MT-CO2 gene in wt and tRNA(Trp)$^{5556G>A}$ mutant cybrids. **(b)** Differential codon abundance in the tRNA(Trp)$^{5556G>A}$ mutant cybrids compared with control wt. **(c)**. The tRNA(Trp)$^{5556G>A}$ mutant cybrid (red line) shows increased RPF density with respect to wt controls (gray line) centered at tryptophan codon but not at a phenylalanine codons, nor at nuclear encoded codons (right panels). **(d)** Quantitative reverse transcriptase PCR (qRT–PCR) of tRNA(TRP) and control tRNA(GLU) from total RNA extracts from cybrids with either (Trp)$^{5556G>A}$ mutant or wt mitochondria (upper panel). Extracts from cybrids were immunoprecipitated by anti-mt-RPL11 or control IgG antibodies. Immunoprecipitated RNAs were detected by qRT–PCR using primers for mt-tRNA(TRP) and mt-tRNA(GLU) (lower panel). Mean values and s.d.'s were calculated from three independent experiments. **(e)** Log-transformed translational efficiencies in the (Trp)$^{5556G>A}$ mutant versus the wt control. Mitochondrially encoded genes are highlighted in green.

Next, we asked whether the relocation of ribosomes to tryptophan codons in the tRNA(Trp)$^{5556G>A}$ mitochondria is a result of stalling, and therefore impacts on the downstream translation process. We aligned the first tryptophan of each mitochondrial transcript and plotted RPF density. Compared with wt mitochondria, tRNA(Trp)$^{5556G>A}$ mitochondria manifested increased abundance of RPFs centered at the first tryptophan codons of mRNAs, and reduced abundance of RPFs downstream of it (Fig. 5a). This trait indicates stalling of the ribosome at tryptophan. To support global ribosome stalling, we aligned all mitochondrial transcripts and plotted RPF density along the transcripts. Figure 5b shows a cumulative density of RPFs and clearly indicates a significant reduction in RPF density at the 3′ half of the transcripts in the tRNA(Trp)$^{5556G>A}$ mitochondria compared with tRNA(Trp)wt. Furthermore, this global effect was observed at the level of the individual transcripts. Figure 5c presents two examples (MT-ND3 and MT-ND5) of the change in

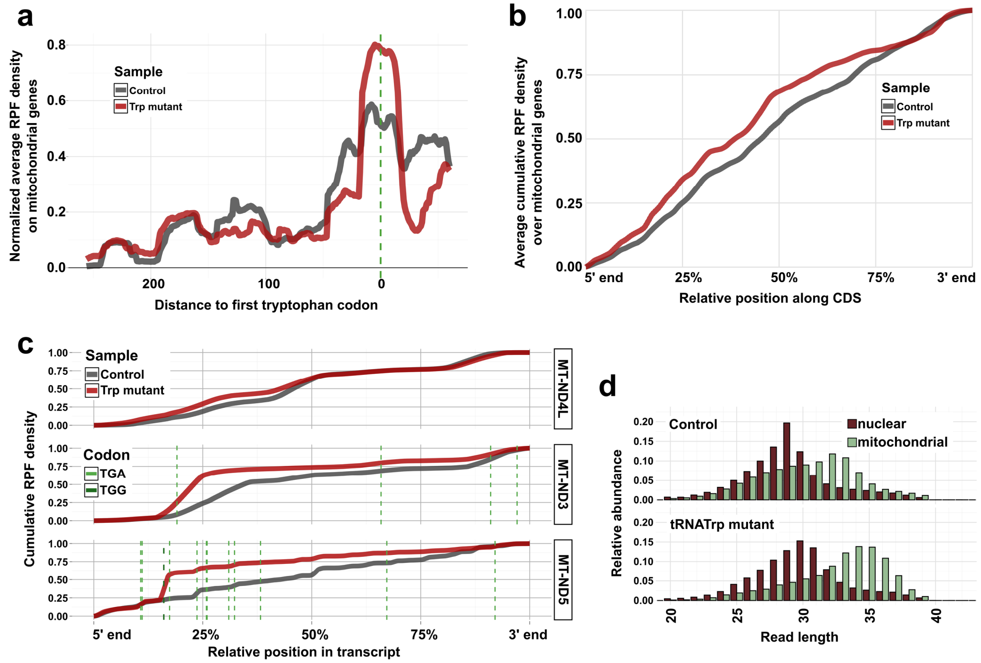


Figure 5: tRNA(Trp)$^{5556G>A}$ stalls ribosomes at tryptophan codons and causes a change in RPF size distribution.
(a) RPF density from the 5′ end of transcripts up to the first tryptophan codon occurrence averaged over the 13 protein-coding mitochondrial transcripts. **(b)** Cumulative RPF density normalized over the 13 protein-coding mitochondrial transcripts. **(c)** Example of the cumulative density of RPF density along the mitochondrial genes MT-ND4L (having no Trp codons), MT-ND3 and MT-ND5, in wt and tRNA(Trp)$^{5556G>A}$ mutant cybrids. **(d)** The relative abundance of RPF length in control and tRNA(Trp)$^{5556G>A}$ mutant cybrids.

RPF density along mitochondrial genes in tRNA(Trp)$^{5556G>A}$ mutant and wild-type mitochondria. The change in RPF density generally showed good correlation with the appearance of tryptophan codons along the mitochondrial transcripts (green dashed lines in Fig. 5c and Supplemental Fig. S4), while the tryptophan-less MT-ND4L transcript showed a similar RPF pattern in control and mutant mitochondria. Thus, our results demonstrate the general effect of tRNA(Trp)$^{5556G>A}$ mutation on mitochondrial translation, pausing ribosomes at tryptophan codons and altering ribosomal density on mitochondrial mRNAs.

Last, the stalling of ribosomes in the tRNA(Trp)$^{5556G>A}$ mutant mitochondria presented us with the possibility to examine the RPFs generated during ribosome progression. In particular, whether stalled ribosomes preferentially produce a certain type of RPFs. As tryptophan codons are enriched at the center of the RPFs in the tRNA(Trp)$^{5556G>A}$ mutant mitochondria, our model (Fig. 3d) predicts accumulation of long RPFs. Figure 5d shows that ribosomes from cybrids with tRNA(Trp)$^{5556G>A}$ mitochondria produce longer RPFs than cells with wt mitochondria. This phenomenon was reproduced in biological replicates

(Supplemental Fig. S3b). A very moderate increase in cytosolic RPFs was observed too, perhaps due to metabolic changes the cells experience in the presence of dysfunctional mitochondria. Altogether, our results show, for the first time at a nucleotide resolution, how a mutation in tRNA disturbs translation, and how this information can be used to decipher ribosome progression.

Discussion

Impaired mitochondrial protein translation affects organismal development, function and ageing, and is a cause for early mortality in humans. Precise measurement of protein translation in the mitochondria is therefore crucial in understanding why malfunctioning occurs. Distinct characteristics of mitochondrial and cytosolic translation machineries however hampered faithful translation measurements using ribosomal profiling. By including several modifications to the ribosomal profiling protocol we show that efficient capturing of cytosolic and mitochondrial ribosomes RPFs in a single experiment is possible.

Our adapted protocol provides a useful new tool for mitochondrial research. It allowed us to show that a genetic disorder caused by a mutated tRNA(Trp)$^{5556G>A}$ leads to stalling of ribosomes on mitochondrial genes. The stalled positions are strongly enriched for tryptophan codons, and inversely tryptophan codons show a large increase in RPFs with respect to wt mitochondria. Furthermore, mitochondria expressing tRNA(Trp)$^{5556G>A}$ showed an increase of ribosomes at the 5′ half of transcripts, and a decrease at the 3′ half. Thus, the point mutation in the mitochondrial-tRNA encoding for tryptophan has a vast effect on ribosome progression and rate of translation along the transcripts. We did not detect any prominent change in the translation in the cytosol, indicating limited effect to the mitochondria. Intriguingly, not all mitochondrial tryptophan codons induced stalling (Fig. 5c and Supplemental Fig. S4), suggesting that other factors, such as mRNA structure, modification and composition, might influence the usage of mutated tRNA by mitochondrial ribosomes.

Our findings also infer, for the first time, mitochondrial ribosome progression from sequencing-based data analyses. Mitochondrial ribosomes display a bi-modal size distribution of RPFs, peaking at 27 and 33 nt, in contrast to the single peak at 30 nt observed from cytosolic ribosomes. This observation suggests a parallel between mitochondrial and bacterial ribosomes, since a similar observation was recently made for the latter (O’Connor et al., 2013). Altogether, the distribution of mitochondrial RPFs, their relative position with respect to mRNA, and changes due to a disease-causing tRNA mutation, suggest a model for ribosome movement (Fig. 3d) and stalling (Fig. 6).

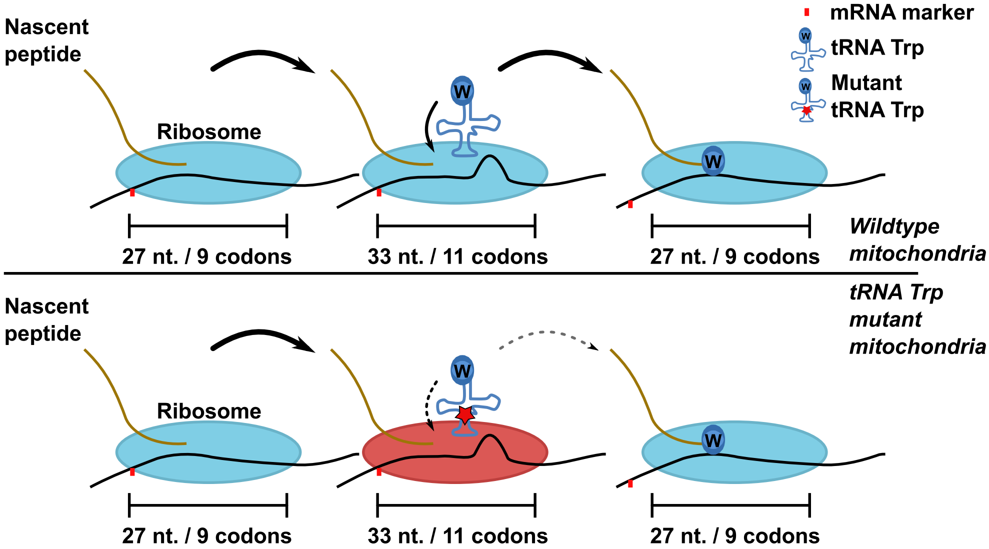

Nascent peptide
Ribosome
W
27 nt. / 9 codons
33 nt. / 11 codons
27 nt. / 9 codons
mRNA marker
tRNA Trp
Mutant tRNA Trp
Wildtype mitochondria
tRNA Trp mutant mitochondria
Nascent peptide
Ribosome
27 nt. / 9 codons
33 nt. / 11 codons
27 nt. / 9 codons

Figure 6: Model for ribosome stalling.
Schematic model for ribosome stalling at codons translated by affected tRNAs, explaining the unbalanced generation of long RPFs in mutant mitochondria.

Materials and methods

RNA sequencing

RNA was isolated using Trizol reagent (Invitrogen). RNASeq libraries were prepared using the TruSeq RNA Sample Prep Kit v2 (Illumina) following the manufacturer's instructions. In brief, RNA was fragmented, then first-strand cDNA was prepared using the kit-supplied 1st Strand Master Mix and Superscript III (Life Technologies) followed by second strand cDNA synthesis. Libraries were amplified by PCR for 12 cycles and sequenced on a HiSeq 2000 System (Illumina).

Alignment of raw data

5′ and 3′ adapters were clipped from ribosome profiling reads prior to alignment. rRNA and tRNA fragments were cleaned by aligning to databases of rRNA and tRNA (compiled from Ensembl (http://dec2011.archive.ensembl.org/index.html) categories 'rRNA', 'rRNA_pseudogene' and 'Mt_rRNA' and the genomic tRNA database (Chan and Lowe, 2009) for hg19 respectively) in two alignment steps and removing reads with one or more hits. Cleaned data was aligned to hg19 using Tophat (Trapnell et al., 2009) and splice junctions of the transcriptome that was assembled by merging Ensembl, Refseq and Broad novel transcripts (Cabili, 2011) using the gffread utility supplied with the cufflinks package (Trapnell, 2012). Only primary alignments (that is, not having SAM flag 0x256) with mapping quality ≥10 were kept.

Size distribution analysis

Valid alignments were assigned to genes using htseq-count[†] and only reads unambiguously assigned to a gene were counted. The gene-wise read density over the read sizes was averaged across genes with at least 100 reads.

Periodicity analysis

Coding sequence (CDS) regions were extracted from Ensembl and the 5′ ends of reads were offset 12 nucleotides in the 5′ direction to match the P-site location of ribosomes (that is, the currently 'read' codon). If reads aligned unambiguously to a single gene and unambiguously to a single frame of one or more transcripts of that gene they were assigned to that frame.

TE analysis

Genewise readcounts (in CDS for ribosome profiling data; over all exons for RNAseq data) were obtained using htseq-count (in 'union' mode, the default) and Ensembl annotations. Translational efficiencies were calculated as the normalized ratios (by total read counts) of RNAseq and ribosome profiling data over the mitochondrial genes. TMM normalization was applied to RNAseq and ribosome profiling data using the R package 'edgeR' (Robinson et al., 2010).

Correlation of short and long RPFs

Abundances of RPFs of 27 and 33 nt in length per mitochondrial protein coding transcript were used to calculate cross-correlation. The cross-correlation values were adjusted by multiplying with the length of the transcript. The normalized cross-correlation was averaged over the 13 transcripts.

Codon enrichment

RPFs assigned uniquely to the CDS of a mitochondrial gene were used to determine codon enrichment. For each RPF the current P-site was estimated by adjusting 17 nt from the 5′ end of the RPF. Only P-sites which were in-frame with the CDS were counted. The total number of each of the 64 codon occurrences was counted in the control and mutant sample, and the relative frequencies were calculated. From this, the differences between mutant and control were calculated.

Cell culture and western blotting

BJ primary fibroblast and cybrid cells were cultured in DMEM supplemented with 10% heat-inactivated fetal calf serum (FCS) in 5% CO_2 at 37 °C. Cell extracts were separated on

---

[†] http://www-huber.embl.de/users/anders/HTSeq/doc/overview.html

10% SDS–PAGE gels and transferred to Immobilon-P membranes (Milipore). Antibodies used were mt-RPL11 (Cell Signaling, 1:1,000), RPL10a (Santa Cruz, 1:1,000), RPL7 (Abcam, 1:1,000) and RPS6 (Cell Signaling, 1:1,000). Validation and homoplasmy of the cybrids harboring the tRNA(Trp)$^{5556G>A}$ mutation were verified by tRNA sequences obtained in the ribosome profiling protocol (see Supplemental Fig. S6, which also shows full-length images of immunoblots).

Ribosome profiling

Approximately 30 × 10^6 cells were treated with chloramphenicol (100 μg ml^{-1}) for 15 min and cycloheximide (100 μg ml^{-1}) for 5 min. Cells were lysed in buffer B (20 mM Tris–HCl, pH 7.8, 100 mM KCl, 10 mM $MgCl_2$, 1% Triton X-100, 2 mM DTT, 100 μg ml^{-1} chloramphenicol, 100 μg ml^{-1} cycloheximide, 1 × Complete protease inhibitor). Lysates were centrifuged at 1300g and the supernatant was treated with 2 U $μl^{-1}$ of RNase I (Ambion) for 40 min at room temperature. Lysates were fractionated on a linear sucrose gradient (7–47%) using the SW-41Ti rotor at 36,000 r.p.m. (221632.5g) for 2 h. Fractions enriched in mito-monosomes and cytosolic monosomes were identified by western blotting, pooled and treated with proteinase K (Roche) in 1% SDS. Released RPFs were purified using Trizol reagent (Invitrogen) following the manufacturer's instructions. Libraries for deep sequencing were prepared as described previously (Loayza-Puch, 2013). In brief, digested fragments between 25–36 nucleotides were gel-purified and dephosphorylated using T4 polynucleotide kinase (New England Biolabs) for 5 h at 37 °C in buffer containing 100 mM MES-NaOH, pH 5.5, 10 mM $MgCl_2$, 10 mM β-mercaptoethanol and 300 mM NaCl. 3′ adaptor was added with T4 RNA ligase 1 (New England Biolabs) for 2.5 h at 37 °C. Ligation products were 5′-phosphorylated with T4 polynucleotide kinase for 30 min at 37 °C. 5′ adaptor was added with T4 RNA ligase 1 for 2 h at 37 °C. All the sequencing experiments were performed on a HiSeq 2000 System (Illumina).

Mitoribosome–tRNA complex inmunoprecipitation

Cybrid cells were fixed with 1% formaldehyde for 10 min at 37 °C, quenched with 2.5 M glycine for 5 min and washed with 1 × PBS. Cell pellets were resuspended in lysis buffer (150 mM NaCl, 50 mM Tris, 0.5% Sodium deoxycholate, 0.2% SDS, 1% NP-40) supplemented with RNAseOUT (Invitrogen) and Complete protease inhibitors (Roche). Protein-G beads were washed and pre-incubated with mt-RPL11 antibody (Cell Signaling) (1:100) for 1 h at room temperature. Extracts and beads were incubated at 4 °C for 1 h. Beads were washed thrice with 1 × PBS, 0.1% SDS, 0.5% NP-40 and twice with 5 × PBS, 0.1% SDS, 0.5% NP-40. RNA was reverse crosslinked in the presence of proteinase K (Roche) at 65° for 3 h. RNA was purified using Trizol reagent (Invitrogen).

Primer sequences

mt-tRNA(TRP) forward:

5′-GAAATTTAGGTTAAATACAGACCAAGA-3′;
mt-tRNA(TRP) reverse:
5′-GAAATTAAGTATTGCAACTTACTGAGG-3′;
mt-tRNA(GLU) forward:
5′-ACAACGATGGTTTTTCATATCATT-3′;
mt-tRNA(GLU) forward:
5′-TTCTCGCACGGACTACAACC-3′;
Gapdh forward:
5′-ACCCAGAAGACTGTGGATGG-3′;
Gapdh reverse:
5′-TCTAGACGGCAGGTCAGGTC-3′

References

Cabili, M.N. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev. *25*, 1915–1927.

Chan, P.P., and Lowe, T.M. (2009). GtRNAdb: a database of transfer RNA genes detected in genomic sequence. Nucleic Acids Res. *37*, D93–D97.

Ingolia, N.T., Ghaemmaghami, S., Newman, J.R.S., and Weissman, J.S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science *324*, 218–223.

Ingolia, N.T., Brar, G.A., Rouskin, S., McGeachy, A.M., and Weissman, J.S. (2012). The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. Nature Protocols *7*, 1534–1550.

Koc, E.C., and Koc, H. Regulation of mammalian mitochondrial translation by post-translational modifications. Biochim. Biophys. Acta *1819*, 1055–1066.

Lin, M.T., and Beal, M.F. (2006). Mitochondrial dysfunction and oxidative stress in neurodegenerative diseases. Nature *443*, 787–795.

Liu, B., Han, Y., and Qian, S.-B. (2013). Cotranslational response to proteotoxic stress by elongation pausing of ribosomes. Mol. Cell *49*, 453–463.

Loayza-Puch, F. (2013). p53 induces transcriptional and translational programs to suppress cell proliferation and growth. Genome Biol. *14*, R32.

Navarro, A., and Boveris, A. (2007). The mitochondrial energy transduction system and the aging process. Am. J. Physiol. Cell Physiol. *292*, C670–C686.

O’Connor, P.B.F., Li, G.-W., Weissman, J.S., Atkins, J.F., and Baranov, P.V. (2013). rRNA:mRNA pairing alters the length and the symmetry of mRNA-protected fragments in ribosome profiling experiments. Bioinformatics *29*, 1488–1491.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Scharfe, C. (2009). Mapping gene associations in human mitochondria using clinical disease phenotypes. PLoS Comput. Biol. *5*, e1000374.

Shalgi, R. (2013). Widespread regulation of translation by elongation pausing in heat shock. Mol. Cell *49*, 439–452.

Sharma, M.R. (2003). Structure of the mammalian mitochondrial ribosome reveals an expanded functional role for its component proteins. Cell *115*, 97–108.

Smits, P. (2010). Functional consequences of mitochondrial tRNATrp and tRNAArg mutations causing combined OXPHOS defects. Eur. J. Hum. Genet. *18*, 324–329.

Smits, P., Smeitink, J., and van den Heuvel, L. (2010). Mitochondrial translation and beyond: processes implicated in combined oxidative phosphorylation deficiencies. J. Biomed. Biotechnol. *2010*, 1–24.

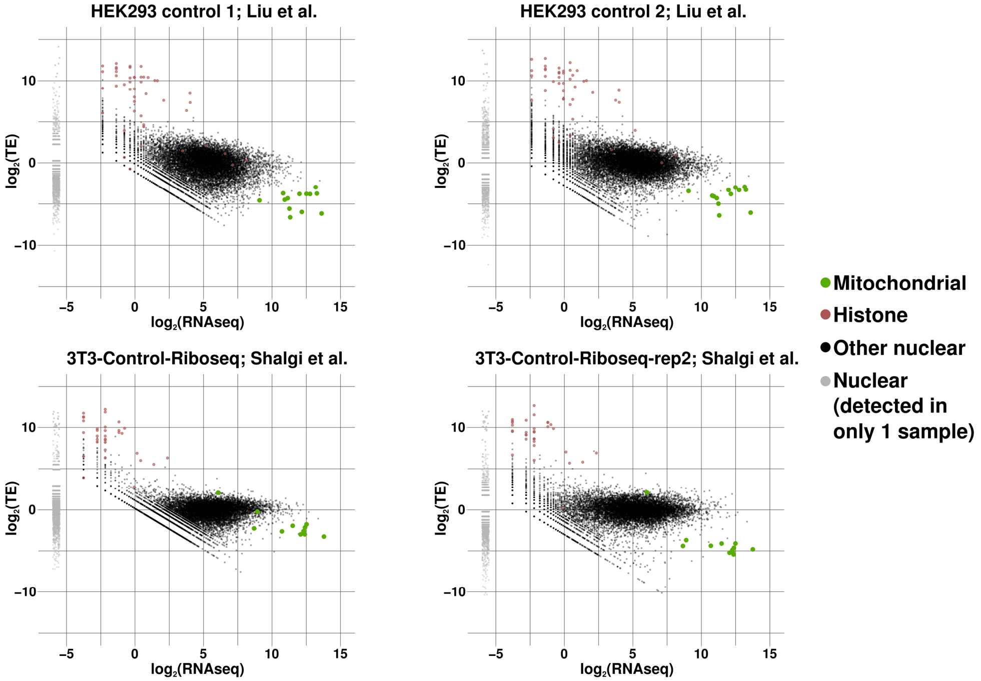
Suen, D.-F., Norris, K.L., and Youle, R.J. (2008). Mitochondrial dynamics and apoptosis. Genes Dev. *22*, 1577–1590.

Trapnell, C. (2012). Differential analysis of gene regulation at transcript resolution with RNA-seq. Nat. Biotechnol. *31*, 46–53.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105–1111.

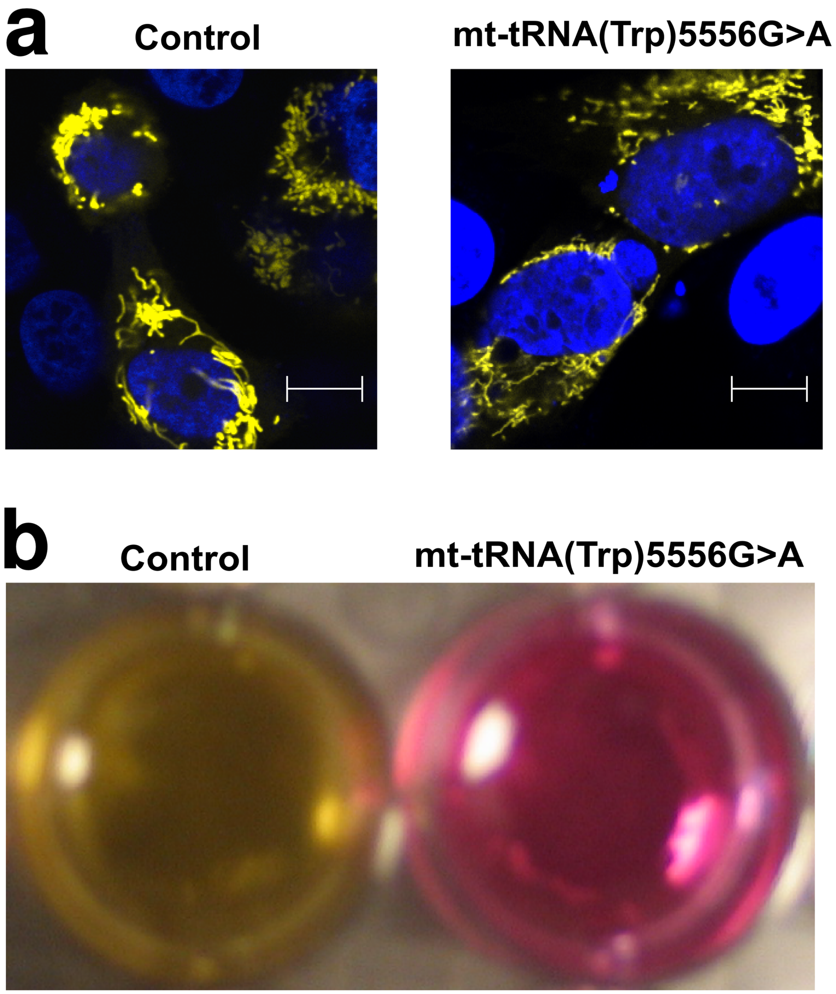
Xiao, X., Wang, Z., Jang, M., Nutiu, R., Wang, E.T., and Burge, C.B. (2009). Splice site strength-dependent activity and genetic buffering by poly-G runs. Nat. Struct. Mol. Biol. *16*, 1094–1100.

Supplemental figures and tables

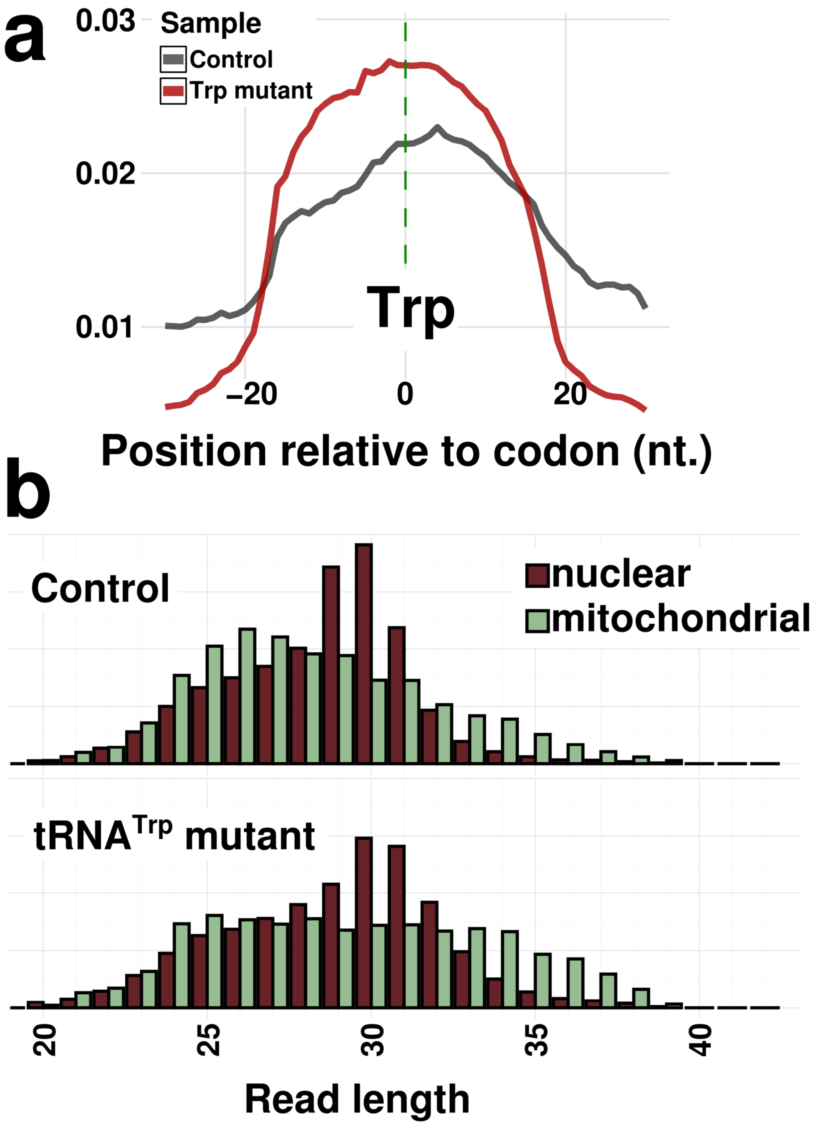


Supplemental Figure S1: Translational efficiency (TE) versus abundance (measured by RNAseq) in two recently published datasets from (Liu et al., 2013) and (Shalgi, 2013).

Since the data from Liu et al. was supplied without matching RNAseq of the samples, we used RNA sequencing data from matching cell line and control condition from (Xiao et al., 2009). GEO sample identifiers of the used data are given in Supplemental Table S1. For the data from Liu et al. the same pipeline was used as the data described in the article. For the data from Shalgi et al. an adapted version of the protocol was used using the Mus Musculus NCBIM37 reference sequence for alignment and Ensembl v67 annotations for analysis.

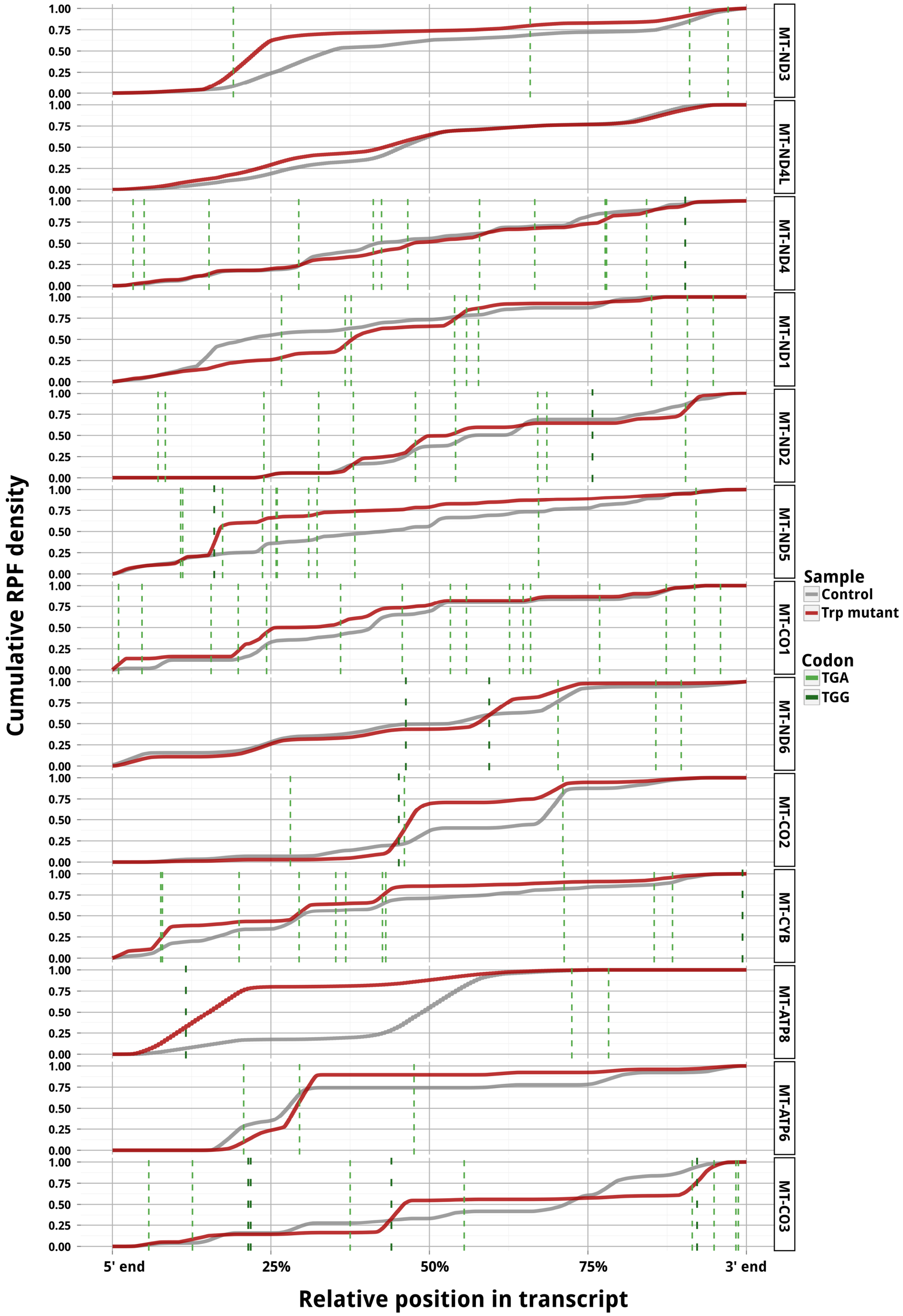

a
Control
mt-tRNA(Trp)5556G>A
b
Control
mt-tRNA(Trp)5556G>A

Supplemental Figure S2: Impact of tRNA(Trp)$^{5556G>A}$ on phenotype.
(a) Immunofluorescence microscopy of Cybrids with wt and tRNA(Trp)$^{5556G>A}$ mutated mitochondria transfected with a mitochondria-targeted YFP vector. The scales measure 6.45 and 7.81 micrometer respectively. **(b)** Image of wells containing control and tRNA(Trp)$^{5556G>A}$ mutant cells.

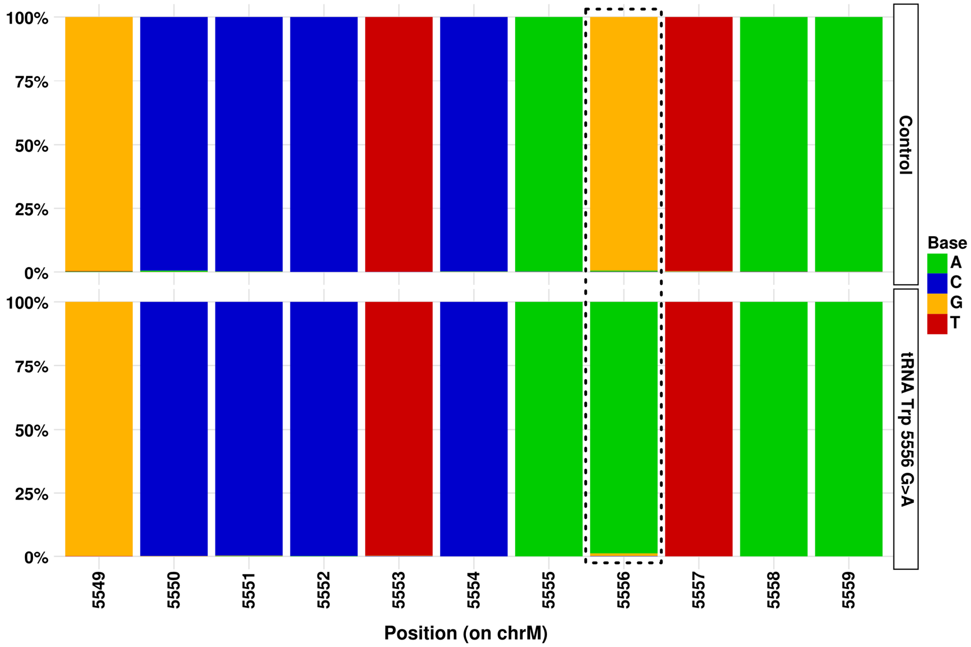


Supplemental Figure S3: Analysis of RP-sequencing data on an independent biological replicate pair of control and tRNA(Trp)$^{5556G>A}$ mutant cybrids.

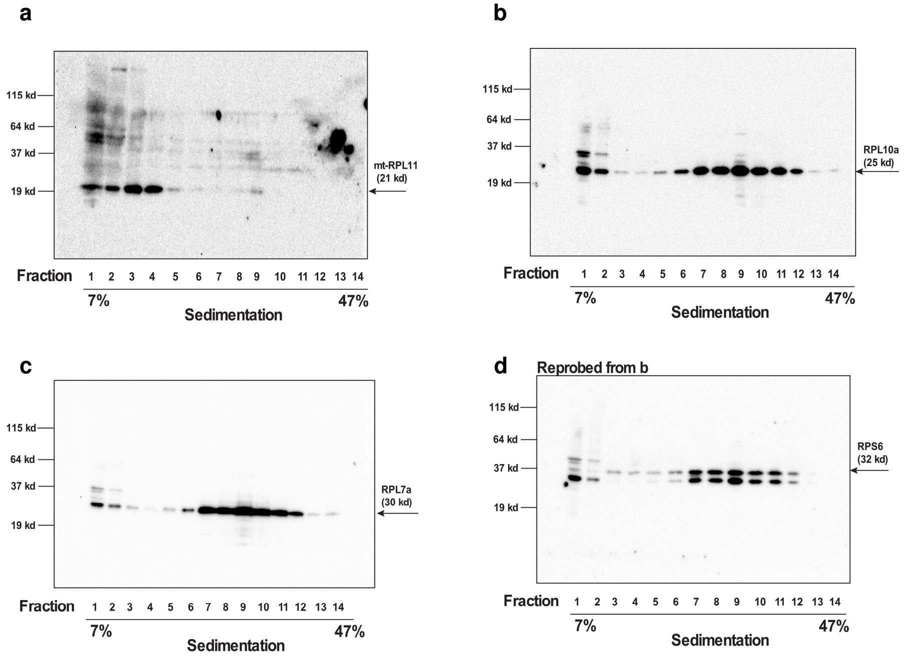
(a) Normalized RPF density plot at tryptophan codon regions in the control and tRNA(Trp)$^{5556G>A}$ mutant sample, analogous to Fig. 3e. **(b)** RPF length analysis analogous to Fig. 4D.



Supplemental Figure S4: Cumulative density plots for all 13 mitochondrial protein-coding genes in the wild-type (gray line) and tRNA(Trp)$^{5556G>A}$ mutant cybrid (red line) samples. Dashed lines indicate the positions of TGA (light green) and TGG (dark green) codons.



Supplemental Figure S5: Genotyping of the control and mutant cells based on reads covering the mitochondrial tRNA(Trp)$^{5556G>A}$ gene mutation. Y-axis shows the fraction of occurrence of each nucleotide for the positions 5549-5559.


a
115 kd
64 kd
37 kd
19 kd
mt-RPL11
(21 kd)
Fraction
1 2 3 4 5 6 7 8 9 10 11 12 13 14
7%
47%
Sedimentation
b
115 kd
64 kd
37 kd
19 kd
RPL10a
(25 kd)
Fraction
1 2 3 4 5 6 7 8 9 10 11 12 13 14
7%
47%
Sedimentation
c
115 kd
64 kd
37 kd
19 kd
RPL7a
(30 kd)
Fraction
1 2 3 4 5 6 7 8 9 10 11 12 13 14
7%
47%
Sedimentation
d
Reprobed from b
115 kd
64 kd
37 kd
19 kd
RPS6
(32 kd)
Fraction
1 2 3 4 5 6 7 8 9 10 11 12 13 14
7%
47%
Sedimentation

Supplemental Figure S6: Full western blots from Fig. 1b.

Supplemental Table S1: Datasets used in the analysis of Supplemental Figure S1.

	Sample	SRA ID	GEO ID	RP/RNAseq
Liu et al.	HEK293_CON_Rep1	SRR619082	N/A	RP
	HEK293_CON_Rep2	SRR619083	N/A	RP
Xiao et al.	Control-KD	SRR029285	GSM417716	RNAseq
Shalgi et al.	3T3-Control-RNAseq	SRR649749	GSM794848	RNAseq
	3T3-Control-Riboseq	SRR649752	GSM794854	RP
	3T3-Control-Riboseq-rep2	SRR649755	GSM794857	RP

CHAPTER 3

Sensing amino acid deficiencies through differential ribosome codon occupancies

Fabricio Loayza-Puch[1,*], Koos Rooijers[1,*], Jelle Zijlstra[1], Joachim F. Oude Vrielink[1], Rui Lopes[1], Alejandro Pineiro Ugalde[1], Ingrid Hofland[2], Jelle Wesseling[2], Axel Bex[3], and Reuven Agami[1,4]

1 Division of Biological Stress Response, The Netherlands Cancer Institute
Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands
2 Molecular Pathology and Biobank, The Netherlands Cancer Institute
Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands
3 Division of Surgical Oncology, Department of Urology, The Netherlands Cancer Institute
Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands
4 Erasmus MC, Rotterdam University, The Netherlands
* These authors contributed equally to this work

Based on a manuscript submitted

Abstract

Tumor growth and metabolic adaptation may restrict the availability of certain amino-acids for protein synthesis. Such information can be utilized for diagnosis and therapy; however, genetic and environmental conditions mandate a tailored detection system for measuring restrictive amino-acids in each tumor. Here, we harnessed ribosome profiling to sense restrictive amino-acids, and developed *diricore*, a procedure for differential ribosome measurements of codon reading. We first demonstrate *diricore's* functionality and constraints using metabolic inhibitors and nutrient deprivation assays. Then, we uncovered leucine restrictions following TGF-beta, ascribing a role for its transporter. Finally, we applied *diricore* to kidney cancer and discovered restrictive proline. Interestingly, proline deficiency was linked to high level of PYCR1, a key enzyme in proline production, suggesting a compensatory mechanism allowing tumor expansion. Indeed, PYCR1 is induced by shortage of proline precursors, and its suppression attenuates kidney cancer cell proliferation when proline is limiting, thus demonstrating the potential therapeutic application of *diricore*.

Introduction

The identification of amino acid biosynthesis deficiencies in cancer cells, as determined by gene expression studies, has already led to one innovative cancer therapeutic treatment. Pioneering research conducted more than five decades ago paved the way for the idea that certain cancers may be auxotrophic for a particular amino acid, and that amino acid deprivation can sensitize these tumors to conventional cancer therapy. So far, this was put to practice by targeting the amino acid asparagine in acute lymphoblastic leukemia (ALL), a common type of leukemia in children and young adults. Lack of the biosynthesis enzyme asparagine synthetase (ASNS) in ALL, but not in most other cell types in the body, causes their addiction to extracellular supply of asparagine (reviewed in (Richards and Kilberg, 2006)). The combination of chemotherapy regiments with a PEGylated form of L-asparaginase treatment (an enzyme that deaminates asparagine to aspartic acid and ammonia) led to a remarkable increase in cure-rates of leukemia from 5% to about 90% (Offman et al., 2011). PEGylation is used to protect protein-drugs from autoimmune reactivity and to increase sustained circulation in the blood.

Similar to L-asparaginase, recent reports have indicated the dependency of certain tumors to other amino acids. Lack of expression of either ASS1 (argininosuccinate synthase) or ASL (argininosuccinate lyase), two enzymes that convert citrulline to arginine, has been noted in several types of cancer (e.g. human melanoma and hepatocellular carcinoma (Delage et al., 2010; Phillips et al., 2013)). Indeed, arginine deprivation in cancer cell lines and tumor models suggested increased cell death that correlates with ASS1 expression, and augmentation of chemo-sensitivity (for recent reference (Liu et al., 2014)). Moreover, clinical trials showed that tumor response to arginine deprivation, using PEGylated arginine deaminase (ADI-PEG20), seems effective and specific to ASS1-deficient tumors (Feun and Savaraj, 2006; Takaku et al., 1992). In addition to asparagine and arginine, it has recently been shown that certain types of cancer cells depend on glycine, glutamine, leucine and serine metabolism to proliferate and survive (Jain et al., 2012; Maddocks et al., 2013; Sheen et al., 2011; Son et al., 2013).

Though gene expression analysis in cancer cells has already been successfully exploited to identify deficiencies in amino acid production in cancer, progress was limited in recent years. An essential limiting aspect is that amino acid demand depends on many genetic and environmental factors of the growing tumor in the organism. Most influential factors are extracellular free amino acid levels, alterations in cellular amino acid uptake, production and catabolism, the use of amino acid for energy, as well as tRNA levels and their availability for protein synthesis. Therefore, to identify restrictive amino acids of a growing tumor, a novel measurement way is required. With this type of information, better diagnosis and assignment of new combinations of therapies can be developed and adapted for each patient. Moreover, new tumor-restrictive amino acids can be discovered.

In contrast to gene expression regulation at the transcriptome level, systematic exploration of the modulation of mRNA translation significantly lagged behind due to the lack of genomic techniques that probe this regulatory layer. This has been changed in the past five years with the development of a deep-sequencing based technique called ribosome profiling (Ribo-Seq) (Ingolia et al., 2009). Ribo-Seq allows the study of changes in rate of protein translation on a truly global scale. It maps the positions of ribosomes on transcripts by nuclease footprinting, generating nuclease-protected ribosome-bound mRNA fragments (RPFs) that are converted into a DNA library suitable for deep sequencing. The prime utility of Ribo-Seq is transcriptome-wide determination of translation rates. In this way Ribo-Seq was used in recent years to explore key functions of translation programs in microRNA-mediating gene suppression, mTOR inhibition, heat shock response, and p53 and senescent induction (Bazzini et al., 2014; Hsieh et al., 2012; Ingolia, 2014; Liu et al., 2013; Loayza-Puch et al., 2013; Thoreen et al., 2012). Thus, Ribo-Seq is a powerful technology that can uncover cellular regulatory translational programs.

Beyond measurement of protein synthesis rates, a second and not less important utility of Ribo-Seq is emerging from the fact that it maps ribosome positions at a nucleotide resolution. This allows global analysis of changes related to codon usage (Ingolia et al., 2009). Our realization of this feature evolved when we interrogated protein translation of human mitochondria (Rooijers et al., 2013). We adapted the Ribo-Seq protocol specifically to examine mitochondrial translation and used it to demonstrate the effect of mitochondrial tRNA mutations on ribosome progression. A strong and reproducible ribosome pausing precisely at the codons corresponding to the affected tRNA was observed. Beyond the biological significance of these findings, these experiments indicated the capability of Ribo-Seq to identify tRNA availability shortages for protein synthesis without any prior genetic knowledge. In other words, by examination of differential ribosome occupancy of normal and patient-derived mitochondria by Ribo-Seq one can pinpoint the causal disease agent. In a similar way, Ishimura R et al., have recently unmasked the disease potential of mutations in nuclear-encoded tRNA genes (Ishimura et al., 2014). However, whether the Ribo-Seq approach can be applied to global identification of amino acid deficiencies was not yet explored.

Here, we hypothesized that Ribo-Seq could be utilized for defining amino acid deficiencies in cancer, and developed *diricore* a tool for this purpose. We validated, evaluated and investigated the functionality of this tool to detect restrictive amino acids in a variety of cell lines and conditions. Moreover, we applied *diricore* to cancer samples and explored amino acid deficiencies in kidney tumors.

Results

Differential ribosome codon reading method

In previous work we used ribosome profiling (Ribo-Seq) to map changes in translation of wild type and disease-triggering mutant mitochondria (Rooijers et al., 2013). A mitochondrial disease-promoting tryptophan (Trp)-tRNA mutation elicited specific pausing of mitochondrial ribosomes at tryptophan codons. Consequently, ribosome occupancy was increased at mitochondrial tryptophan codons, resulting in enrichment of ribosome-protected fragments (RPFs) covering these sites. We therefore hypothesized that also in the cytosol, global occupancy measurements of ribosome positions could be utilized as a cellular detector of alterations in the availability of amino acid for protein synthesis. To examine this issue we constructed a protocol to measure differential ribosome codon reading (*diricore*) (Figure 1a). In particular, we performed two complementary analyses: One, subsequence analysis; where we examined the codons at 9, 12 and 15 nucleotides (nt) from the 5'-ends of RPFs. And two, RPF density analysis; where we analyzed every instance of a codon throughout the transcriptome to examine the local RPF 5'-end density surrounding that codon. While the latter makes use of all RPFs, subsequence analysis takes into account only those that are in frame with the coding sequence. The positions 9, 12 and 15 from the 5'-end of the RPFs correspond to the E (exit), P (peptide bond), and A (tRNA recruitment and reading codon) sites of the ribosome respectively (Bazzini et al., 2014; Guo et al., 2010; Ingolia et al., 2009).

Extent and sensitivity of diricore

To validate our approach we first examined a dataset of Ribo-Seq performed on yeast treated with 3-Amino-1,2,4-triazole (3-AT), an inhibitor of histidine synthesis (Guydosh and Green, 2014), and measured the occupancy of all codons at the different positions. As expected, this resulted in a strong signal in the two histidine-codons at the reading codon (position 15 of RPFs, Figure 1b, left panel). Moreover, examination of histidine codons, but not control phenylalanine, showed a vast increase of 5'-ends of RPFs mainly at position 15 nt upstream of the codon (Figure 1b, right panels). Second, we tested harringtonine treatment, an agent that immobilizes initiating ribosomes and depletes elongating ones (Fresno et al., 1976; Ingolia et al., 2011), in the human SUM1315 breast cancer cell line. As expected, we obtained a clear increased signal at methionines, but only at the initiator ATG codon (Figure 1c). Here, however, the signal was detected at position 12, corresponding to ribosomes arresting with the initiation codon at the P-site and in line with harringtonine's function in blocking elongation. As a final proof-of-principle we analysed the response of cancer cells to L-asparaginase, a currently used therapeutic agent that targets asparagine in acute lymphoblastic leukemia (ALL) (Richards and Kilberg, 2006), by *diricore*. We revealed a specific signal at asparagine codons in cells treated with L-asparaginase (Figure 1d), indicative of asparagine shortage of the treated cells.

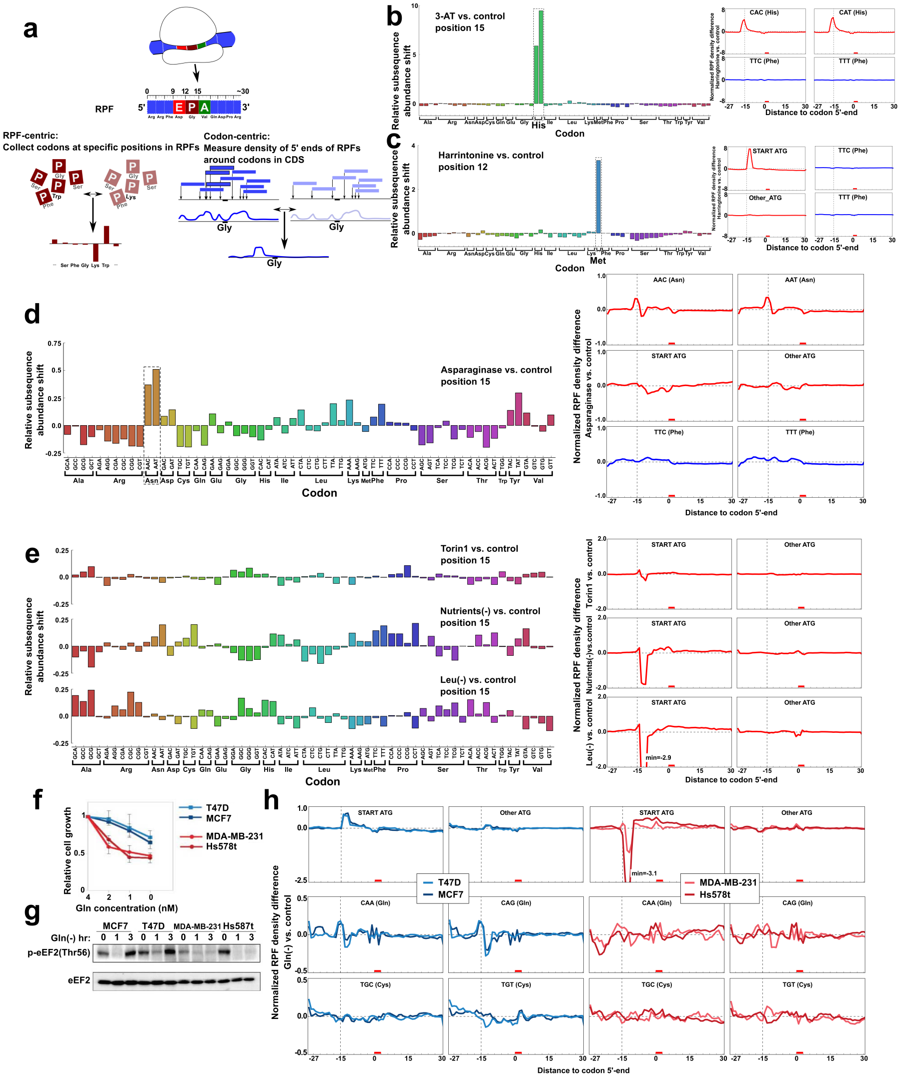


Figure 1: Ribosome positions as readout for amino acid availability for protein synthesis. **(a)** Scheme of the protocol we used for differential ribosome codon reading (*diricore*). In subsequence analysis (left) subsequences (in the case where the RPF is aligned to the CDS these subsequences are mRNA codons) are counted, normalized, averaged and compared between conditions. In RPF density analysis (right) the density of RPF 5'-ends is measured, normalized, averaged and compared between conditions. **(b)** *Diricore* analysis of data obtained from (Guydosh and Green, 2014) of yeast cells treated with 3-Amino-1,2,4-triazole (3-AT), an inhibitor of His production, shows signal at His codons by increased occurrence of His codons in the 15th nt. from the 5'end of the RPFs (left panel), and corresponding increased density of RPFs 15 nt. upstream of His codons (right panels). **(c)** *Diricore* analysis of SUM1315 cells treated with harringtonine, an inhibitor of initiation of translation (Fresno et al., 1976; Ingolia et al., 2011), reveals a strong and specific signal at the initiator Met at position 12 of RPFs. **(d)** *Diricore* analysis of the PC3 cells treated for 48hrs with L-asparaginase reveals a specific signal at asparagine codons at position 15 in RPFs. **(e)** *Diricore* analysis of MCF10A cells either treated with Torin1 (an inhibitor of mTOR (Thoreen et al., 2009)) for 2 hours, or deprived of nutrients for 2 hours, or deprived of leucine for 24 hours. Left panels show subsequence analyses, and right panels RPF density analyses of the initiator-ATG and all other ATG codons (marked "START" and "Others", respectively).

Next, we wished to evaluate *diricore* in other biological settings. The kinase mTOR complex is a cellular amino acid sensing hub (Efeyan and Sabatini, 2013; Kim and Guan, 2011). Nutrient deprivation of the human non-transformed MCF10A epithelial cell line leads to rapid mTOR inhibition, global reduction in ribosomal protein translation, and acute suppression of initiation of protein synthesis that is followed by the induction of autophagy (Supplemental Figure S1). A similar principle response is obtained by treatment with Torin1, a specific inhibitor of mTOR (Thoreen et al., 2012). Moreover, also deprivation of leucine, a key amino acid of the mTOR sensing machinery, is known to provoke such global responses (Efeyan and Sabatini, 2013). We therefore tested whether amino acid deprivation, either global or specific, would lead to ribosome pausing and specific signals in *diricore*. We used Torin1 as control since it causes a halt of translation without any shortage of amino acids. Figure 1e, top panel shows that no specific *diricore* signal was provoked by 2 hours Torin1 treatment. A very similar *diricore* profile was observed by 2 hours nutrient deprivation (Figure 1e, middle panel). However, here we observe a strong depletion of RPFs 12 nt. upstream of the initiator-ATG codon. We attribute this depletion to the strong halt of translation initiation following treatment. Lastly, 24 hours leucine deprivation also triggered a *diricore* signal of initiator-ATG codon depletion at position 12, accompanied by no specific signal at other codons (Figure 1e lower panels). Thus, as all treatments were demonstrated to globally impact protein synthesis, our results correlate the strong depletion of the initiator-ATG codon in *diricore* with global inhibition of the protein translation machinery. Moreover, these results also suggest that strong global translation inhibition puts constrains on the identification of codon-specific signals, as no specific signal in the coding sequence was observed neither when essential amino acids or only leucine were acutely deprived.

To further test *diricore* we used basal and luminal breast cancer cell lines and deprived them from glutamine. It has been recently demonstrated that while basal breast cancer cell lines are highly sensitive to a glutamine deprivation treatment, luminal breast cancer cell lines are relatively resistant (Singh et al., 2012). Indeed, glutamine deprivation greatly affected the survival of the basal MDA-MB-231 and Hs587T cell lines, while it only mildly affected T47D and MCF7 luminal cells (Figure 1f). As previously reported (Leprivier et al., 2013), the resistance of the luminal cells to glutamine deprivation was associated with sustained control of translation elongation by maintaining the inhibitory phosphorylation on Threonine 56 of the key elongation factor eEF2 (Figure 1g). In contrast, the basal cell lines

Figure 1 (cont.)
(f) Normalized cell growth of luminal (T47D and MCF7, blue shades) and basal (MDA-MB-231 and Hs578t, red shades) breast cancer cell lines at different concentration of glutamine. Measurements were taken 48hrs after plating. **(g)** Immunoblot analysis of indicated breast cancer cells grown in complete media or under glutamine starvation for the indicated times. **(h)** *Diricore* analysis of luminal (blue shades) and basal (red shades) breast cancer cell lines deprived of glutamine for 48 hrs.

do not exhibit this inhibitory phosphorylation, allowing translation elongation to proceed. This suggested to us that glutamine deprivation would greatly impact global translation in the basal cells, preventing detection of specific signals by *diricore*, while luminal cells will respond in a more mild and specific way. We therefore treated the basal and luminal cell lines with a culture medium deprived of glutamine, profiled ribosomes, and analyzed RPFs by *diricore*. In line with our assumption, *diricore* revealed great reduction of the initiator-ATG codon at position 12 in the basal cells (indicating acute translational inhibitory response), while very little signal was observed at this position in the luminal cell lines (Figure 1h upper panels). In contrast, we observed *diricore* signals at glutamine codons clearly appearing only in the luminal cells (Figure 1h middle panels). As control we used two cysteine codons, which showed no specific signal in any of the conditions (Figure 1h lower panels). Thus, our observations here support the use of *diricore* analysis of the initiator-ATG codon to estimate global changes in protein translation processes, and also indicate that *diricore*-signals of specific codons within the coding sequence appear in conditions where strong global inhibition of translation has not (yet) occurred.

***Diricore* uncovers restrictive amino acids following TGF-beta treatment**

Following *diricore* development and validation, we searched for apparent changes in a number of cell treatments, including TGF-beta, an inducer of epithelial to mesenchymal transition (EMT). Interestingly, 48 hours TGF-beta treatment of MCF10A cells resulted in a strong EMT signature (Supplemental Figure S2a) while simultaneously causing increased leucine occupancies at the reading codon (position 15) and decreased occupancies in position 12 (Figure 2a top panel). We used three biological replicates to strengthen our confidence and allow statistical analysis to be performed. This indicated a highly significant effect of TGF-beta on RPF density at leucine codons (adjusted p-value < 0.001). In line with our observation so far, we observed no decrease in *diricore*-signal at position 12 of the initiator-ATG (Figure 2a bottom panel), indicating that TGF-beta treatment did not elicit global inhibitory effects on protein translation, likely permitting *diricore*-signals at specific codons to be detected. We corroborated this result with *in vivo* protein incorporation assays (Liu et al., 2012), which consistently revealed no significant change in global protein synthesis rate following TGF-beta treatment, while control Torin1 treatment reduced it as expected (Figure 2b). Additionally, by western blotting we did not observe a reduction in phosphorylated 4E-BP1, a downstream key translation initiation factor of mTOR activity, in contrast to the marked expected effect of Torin1 (Supplemental Figure S2b). Our data so far suggest that TGF-beta treatment of MCF10A cells provokes leucine-codon signals in *diricore,* which may indicate mild leucine limitations for protein synthesis. Our analysis of the *diricore* signals of the initiator-ATG suggests that even though leucine is a signaling molecule that is required for mTOR activation (Efeyan and Sabatini, 2013), the presumed leucine limitation following TGF-beta treatment did not affect global translation, permitting its detection by *diricore*.

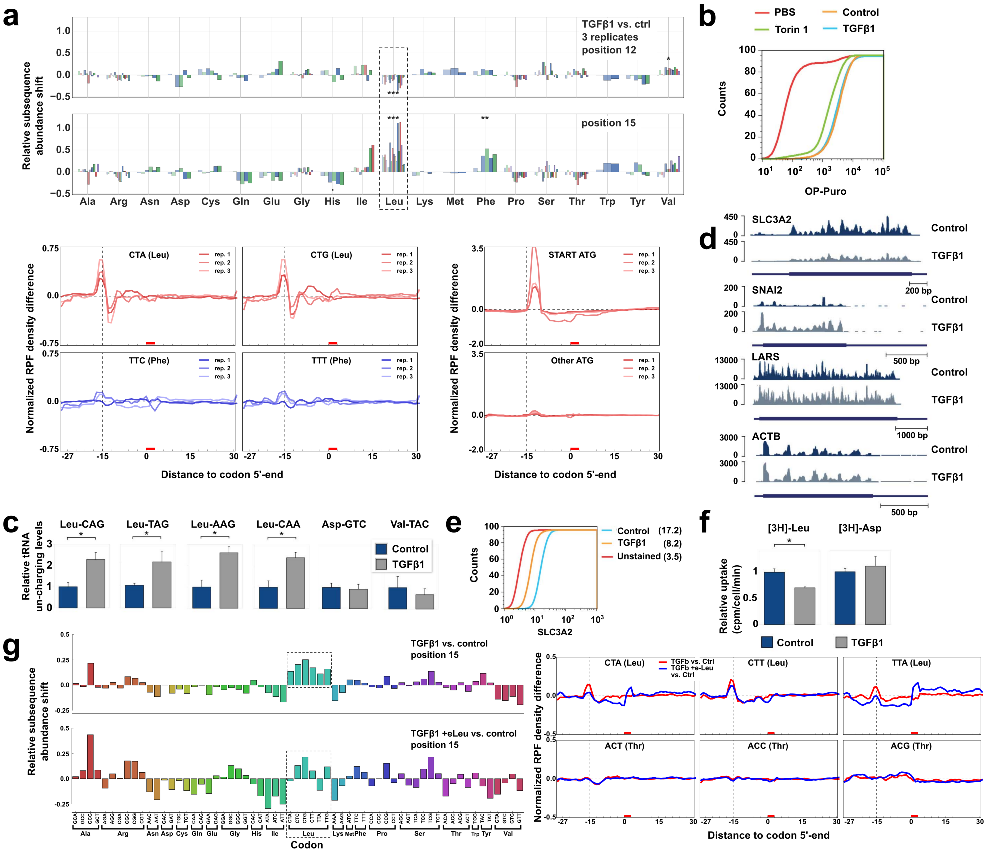


Figure 2: *Diricore* links regulation of leucine transporter levels to TGF-beta signaling. **(a)** *Diricore* analysis of MCF-10A cells treated with TGF-beta vs. untreated controls, in biological triplicates. Top panels show subsequence analysis of codons at the 12th and 15th positions of RPFs. Bottom panels show 5' RPF density analyses at leucine and phenylalanine codons (left) and methionine codons, split out by initiator codon (right). Asterisks at the subsequence plot indicate statistical significance for deviating occupancy of codons of a particular amino acid, compared to the controls. Statistical significance was determined by a linear mixed model in which the amino acids constitute the fixed effects and codons constitute random effects. A separate model for each position (12 or 15 in this plot) is fit. Significance thresholds are: * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$ and are adjusted for multiple testing. **(b)** OP-Puro incorporation in MCF10A cells treated either with TGF-beta for 48 hours, Torin-1 for two hours, or left untreated, assessed by flow cytometry. **(c)** MCF10A cells were treated as in panel (a), and subjected to tRNA aminoacylation analysis (Zaborske et al., 2009) for the indicated tRNAs. Error bars represent standard deviations. n=3, and '*' indicates p<0.01 by a two-tailed Student t-test. **(d)** Ribo-Seq-based expression analysis of the indicated genes shows suppression of SLC3A2, an essential component of the leucine transporter in the cell. ACTB is shown as negative control for differential expression, while SNAI2 expression is shown as positive control for TGF-beta-induced EMT.

Figure 2 (cont.)
(e) SLC3A2 expression determined by FACS in control and 48h TGF-beta-treated MCF10A cells. **(f)** Uptake of [^{3}H]-Leu and control [^{3}H]-Asp measured in MCF10A cells that were treated as in panel (a). Error bars display standard deviations; n = 3;* indicates $p < 0.05$ by two-tailed Student's t test. **(g)** *Diricore* analysis of MCF10A cells treated with TGF-beta and esterified leucine (eLeu) or control vehicle. Left panels show subsequence shifts while right panels show RPF density differences, where in red the response to TGF-beta and in blue the response to TGF-beta and eLeu is shown.

To unravel the chain of events causing the leucine signal, we initially examined changes in gene expression, as a dramatic increase in the amount of leucine codons following TGF-beta treatment may indicate increase in demand. However, no indication for increased leucine demand was obtained when we either inspected it by RNAseq (Supplemental Figure S2c) or when a correlation was made between the fraction of leucine codons and TGF-beta-induced translation efficiency changes (Supplemental Figure S2d). Then, we examined the total levels of Leu-tRNAs, but observed no change either (Supplemental Figure S2e). We proceeded to examine the tRNA-aminoacylation levels (Zaborske et al., 2009). This analysis indicated a significant increase of non-aminoacylated levels of all tested Leu-tRNAs, but not of the controls (Asp-tRNA and Val-tRNA) upon treatment (Figure 2c). This pinpoints to cellular limitation in leucine availability for translation following TGF-beta treatment.

To uncover the molecular mechanisms leading to TGF-beta-induced limitation in leucine we examined the expression levels of genes related to its cellular uptake, synthesis, and tRNA aminoacylation. Figure 2d shows that while TGF-beta induced no change in the level of LARS (the only cytosolic Leu-tRNA synthetase), the level of SLC3A2 (also known as CD98, the major leucine transporter) was down ~2 fold. We used snail2 (SNAI2), a key TGF-beta target, and beta-Actin (ActB) as positive and negative controls, respectively. Using flow cytometry analysis we validated SLC3A2 down-regulation following TGF-beta treatment at the level of the protein (Figure 2e). If the leucine amino acid transporter is involved in generating the leucine *diricore* signal, reduced cellular uptake of leucine is predicted. Indeed, the uptake of leucine, but not control aspartic acid, was reduced in MCF10A treated with TGF-beta (Figure 2f). Thus, reduced SLC3A2 leucine-transporter level coincides with the increase in *diricore* signal at leucine codons following TGF-beta treatment.

Next, we postulated that if TGF-beta affects the uptake of leucine, artificially increasing the cellular uptake of leucine should weaken the *diricore* signal in a specific manner. To examine this assumption we reduced restraints in leucine transportation by adding to the medium esterified-leucine (eLeu, a leucine analogue that can cross cellular membranes and be used for protein synthesis, (Thiele and Lipsky, 1990)). We therefore induced MCF10A cells with TGF-beta for 48 hours in the presence of eLeucine or control vehicle and first tested Leu-tRNA aminoacylation. Supplemental Figure S2f shows the expected increase in Leu-tRNA-aminoacylation following TGF-beta, and reveals a significant attenuation of this effect by eLeucine. We use Val-tRNA aminoacylation as control, and this remained unaffected. Then we used the same cell populations, profiled ribosomes, and examined RPFs by *diricore*. Figure 2g shows that while mock-treated cells elicited the expected leucine *diricore* signal at position 15, the addition of eLeucine abrogated this signal in a specific manner. Furthermore, we monitored the global cellular response to TGF-beta by examining SNAI2 and SLC3A2 expression levels. This indicated comparable changes in both eLeucine and control conditions (Supplemental Figures S2g and S2h). These results strongly support the notion that TGF-beta affects leucine uptake in MCF10A cells. Moreover, our observations thus far underscore the robustness of *diricore* to unravel biological events occurring following changes in cellular states.

Down-regulation of SLC3A2 is required for the leucine-*diricore* signal following TGF-beta treatment

Next, we questioned whether SLC3A2 is causal to the *diricore* signals following TGF-beta treatment using SLC3A2 knockdown and over-expression tools. We first knocked down SLC3A2 and confirmed suppression of expression and functionality by qRT-PCR (Figure 3a) and flow cytometry (Figure 3b). We confirmed that the knockdown led to reduced uptake of leucine (Figure 3c). Furthermore, tRNA-aminoacylation assays showed a significant increase in non-aminoacylated Leu-tRNAs, but not in control Asp- and arginine (Arg)-tRNAs, following SLC3A2 knockdown, suggesting a specific effect of SLC3A2 on Leu-tRNAs (Figure 3d). Finally, *diricore* analysis identified a signal at leucine codons elicited by SLC3A2 siRNAs, which was comparable to TGF-beta in control siRNAs transfected cells (Figure 3e). In contrast, the major TGF-beta target, SNAI2, was not induced by SLC3A2 knockdown (Supplemental Figure S3a), indicating that the leucine-*diricore* signal in the SLC3A2 siRNA-transfected cells was not due to aberrant activation of TGF-beta signaling. To further examine the role of reduced SLC3A2 expression in TGF-beta-induced leucine stalling, we stably over-expressed SLC3A2. We validated the over-expression by FACS (Supplemental Figure S3b). TGF-beta treatment in SLC3A2 over-expressing cells still causes down-regulation of SLC3A2, but its levels are comparable to non-overexpressing untreated cells (compare blue and dark green lines in Supplemental Figure S3b). We found that SLC3A2 over-expression stimulated leucine uptake (Supplemental Figure S3c) and prevented the decrease in aminoacylation of Leu-tRNA following TGF-beta treatment (Supplemental Figure S3d). SLC3A2 over-expression did not alter TGF-beta-induced EMT as indicated by SNAI2 induction (Supplemental Figure S3e). Finally, *diricore* analysis showed that SLC3A2 over-expression attenuated the signal at leucine codons (Supplemental Figures S3f). Taken together, using *diricore* we uncovered a causal key role of SLC3A2 in limiting intracellular levels of leucine following TGF-beta treatment.

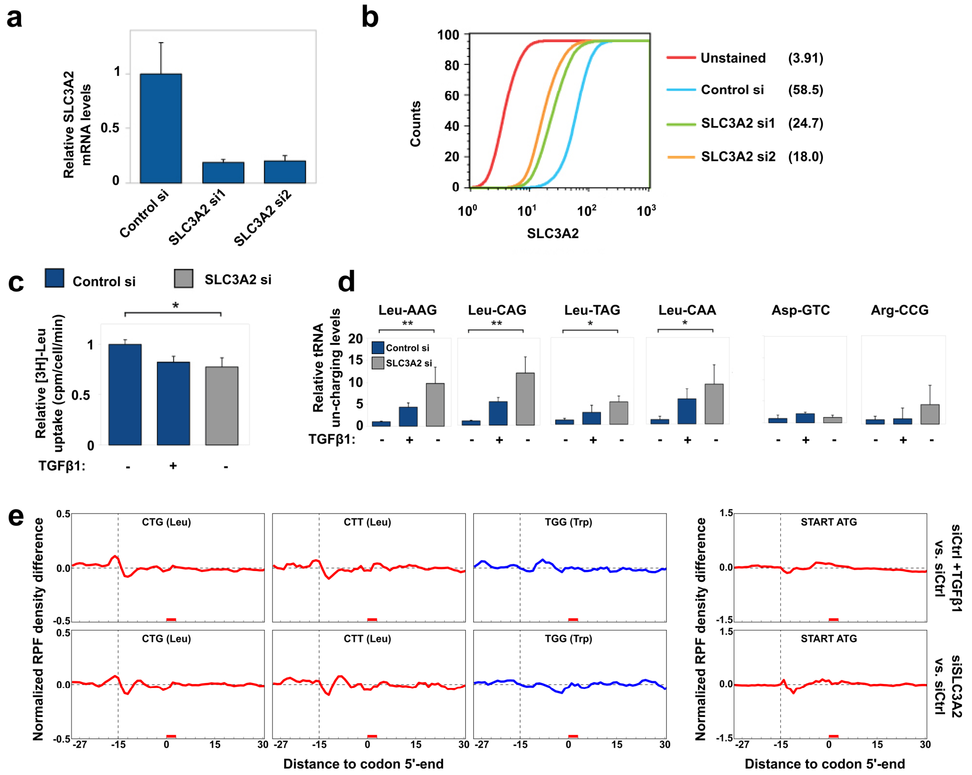


Figure 3: SLC3A2 partially mediates the differential ribosomal leucine occupancy elicited by TGF-beta

(a) SLC3A2 expression as determined by qRT-PCR analysis in MCF10A cells transfected with siRNAs against SLC3A2 or a control siRNA. Expressions were normalized to GAPDH. **(b)** SLC3A2 expression as determined by flow cytometry analysis in MCF10A cells transfected with siRNAs against SLC3A2 or a control siRNA. **(c)** The uptake of [^{3}H]-Leu measured in MCF10A cells transfected with SLC3A2 or control siRNA in either TGF-beta-treated or control condition. **(d)** Aminoacylation of indicated tRNAs in MCF10A cells treated as in panel (c). **(e)** RPF 5'-end density analysis of MCF10A cells treated as in panel (c).

Diricore uncovers Proline deficiency in kidney cancer

Following the application of *diricore* to cell lines, we examined whether *diricore* could also detect restrictive amino acids in cancer. We therefore obtained fresh clear cell renal cell carcinoma (ccRCC) kidney tumor samples, and sampled two normal and four tumor tissues for Ribo-Seq (Supplemental Figure S4a). Quality control of the Ribo-Seq data indicated high and comparable quality of all samples (determined by the percentage of mapped reads to coding sequences, frame of RPFs, and the high correlation within the two normal samples, Supplemental Figure S4b). Furthermore, gene expression analysis verified high expression of the ccRCC marker genes: EGFR, vimentin (VIM) and CAIX (CA9) in all tumor tissues (The Cancer Genome Atlas Research Network, 2013) (Supplemental Figure S4c).

Next, we performed *diricore* analysis comparing each of the normal and tumor tissues to both normal tissues. While no signal was detected between the two normal tissues (Supplemental Figure S4d), the comparisons of each of the tumor tissue to either normal revealed two consistent signals (Figure 4a). We identified a strong peak in methionine codons at position 12, which correspond to the P site of the ribosome, as well as a signal in proline codons at both positions 12 and 15, corresponding to ribosome's P and A sites respectively. Importantly, all tumor-versus-control samples showed these two signals, indicating first that *diricore* is a robust technology, and second, that at least for this tumor common signals are detected from samples retrieved from different regions of the tumor and normal tissues. Intriguingly, the strong signal in methionine was solely originating from the initiator-ATG codon and was not observed in other ATG codons, indicating an alteration in translation initiation rather than methionine (Met)-tRNA levels (see Figure 4a right panels). As also the position of the peak is at the P-site, and not the A-site where the specific tRNA enters the ribosome, we propose that increased translation initiation rate in the tumor is the underling mechanism. To inspect this issue in more depth, we examined the phosphorylation level of 4E-BP1, a member of a family of translation initiation repressors that by binding to the eukaryotic translation initiation factor 4E (eIF4E) limits the recruitment of 40S ribosomal subunit to the ribosome. Phosphorylation at threonines 37/46 commonly occurs in cancer cells resulting in the inhibition of 4E-BP1 translation-suppressive function and induction of translation initiation (Zoncu et al., 2011). By immunohistochemistry (IHC) analysis with specific anti-phosphorylated 4E-BP1 antibodies (phospho-Thr 37/46) we observed a strong elevation in p4E-BP1 level in the tumor compared to the normal tissue (Figure 4b). Altogether, *diricore* identified a signal at the initiator methionine codon that in all likelihood is a result of enhanced global translation initiation in the tumor.

In contrast to the initiator-ATG, the signal in proline codons appeared at both positions 15 and 12 (P and A sites), suggesting limiting availability of proline (Pro)-tRNA for protein synthesis (Figure 4a right panels). In line with this conclusion, the tumor tissue showed increased non-aminoacylated Pro-tRNA levels compared to the normal tissue (Figure 4c) while Leu-tRNA controls did not show changes. Surprisingly, gene expression analysis of the proline metabolic enzymes showed vast up-regulation in the expression of PYCR1, a mitochondrially-localized protein that catalyzes the last step in the proline synthesis pathway, while the proline catabolic enzymes PRODH2 and ALDH4A1 were strongly down-regulated (Figure 4d and Supplemental Figure 4c). While loss of PRODH2 expression is common to the vast majority, if not all, of kidney tumors (Dalgliesh et al., 2010), PYCR1 over-expression is infrequently observed (around 4% of tumors have high PYCR1; Figure 4e). We confirmed up-regulation of PYCR1 in this tumor by IHC (Figure

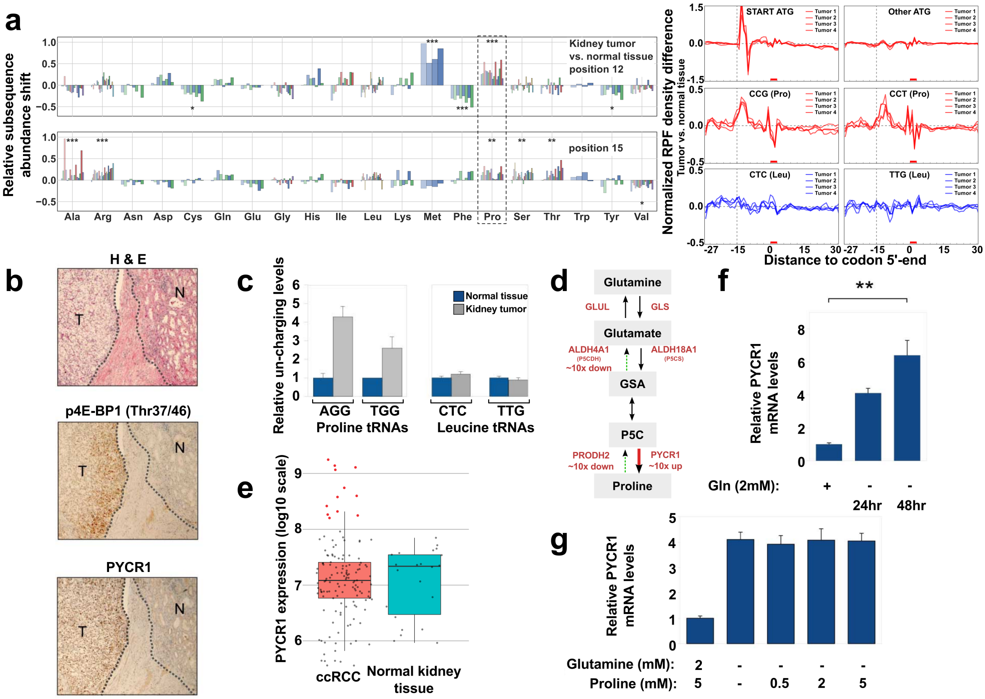


Figure 4: *Diricore* detects limitations in proline availability in ccRCC.
(a) *Diricore* analysis of ccRCC tumor samples (4 replicates) vs. normal tissue. Statistical significance is indicated in the subsequence analysis as in Figure 2a. Right panels show RPF density at ATG codons (top), proline codons (middle) and leucine codons (bottom, negative control). **(b)** IHC performed on a section containing tumor and normal tissues (T and N, respectively) using hematoxylin and eosin (top), anti-p4E-BP1 threonine 37/46 (middle) and anti-PYCR1 (bottom) antibodies. **(c)** Tumor and normal kidney tissue were subjected to Pro-tRNA aminoacylation analysis (left panel). Leu-tRNAs (right panel) are used as control. **(d)** Scheme and summary of gene expression changes observed in the data in the metabolic pathway from glutamine to proline, common to all tumor samples. Presented in detail in Supplemental Figure S4c. In particular, activation of PYCR1 and down regulation of PRODH2 and ALDH4A1 were noted. **(e)** Expression analysis of PYCR1 in a panel of ccRCC tumor and normal tissue samples (dataset GSE17895 (Dalgliesh et al., 2010)). **(f)** PYCR1 levels as determined by qRT-PCR in response to glutamine starvation in A498 cells. **(g)** As in panel (f), A498 cells were depleted from glutamine, and then supplemented with proline in the indicated concentrations. PYCR1 levels were measured by qRT-PCR.

4b, bottom panel). The proline signal in *diricore* and the activation of the proline production pathway indicate a regulatory compensatory attempt mechanism in the kidney cancer cells to shortage of proline for protein synthesis.

Proline deficiency in the tumors can be a result of increased proline demand or decreased in proline production, for example due to lack of proline precursors in the proline metabolic pathway. To investigate this point we first measured codon demand differences caused by gene-expression changes, by calculating the relative number of proline codons in transcriptomes of the normal and tumor sections by RNASeq. Supplemental Figure S4e shows that no global increase in proline demand is observed in the tumors. Moreover, no increase in proline demand was detected, even when we restricted our analysis to the genes whose expression significantly changed (up or down) in the tumor compared to the normal tissue (Supplemental Figure S4f). To take into account any changes at the translational level that might introduce a codon demand, we analyzed how translational efficiencies relate to the fraction of proline codons per gene, but observed no correlation (Supplemental Figure S4g). This further indicated that no changes in proline demand were underlying our observations. Second, we investigated the regulatory events leading to PYCR1 activation, rationalizing that this can be a clue to the proline deficiency. Particularly, we asked whether limiting amount of glutamine, the precursor of proline (see Figure 4d), or low proline levels, are causal to activation of PYCR1. For this purpose we selected a ccRCC cell line (A498) with moderate levels of PYCR1 (Supplemental Figure S4h), depleted glutamine for 24 hours, and measured the relative level of PYCR1 by q-RT-PCR. This analysis revealed a strong and significant activation of PYCR1 expression when glutamine level was reduced (Figure 4f). Moreover, when glutamine deprivation was extended to 48 hours, PYCR1 activation was further increased suggesting a dynamic control (Figure 4f). To examine the role of proline in this effect, we added proline to the glutamine-depleted cells but observed no reduction in the elevated levels of PYCR1 (Figure 4g). Thus, these results do not support an auto-regulatory circuit between proline levels and PYCR1, but rather indicate that PYCR1 responds to reduced levels of a precursor substrate in the metabolic pathway from glutamine to proline. Altogether, these results suggest that the proline deficiency in PYCR1-high ccRCC tumors is a result of shortage in building blocks for proline rather than increase in proline demand.

Interestingly, when we examined a second ccRCC tumor by *diricore*, a clear signal was identified in the initiator-ATG, which resembled the signal observed in the first tumor, but no signal was observed at proline codons (Supplemental Figure S4i). Also in the second tumor, the signal in the initiator-ATG was in line with a global activation of translation initiation, as measured by IHC with anti-p4E-BP1 Threonine 37/46 antibodies (Supplemental Figure S4j). In contrast, the lack of proline signal suggested to us that this second tumor did not experience deficiency in proline availability. Indeed, examining the levels of aminoacylation of Pro-tRNA and Leu-tRNAs showed no increase in uncharged tRNAs (Supplemental Figure S4k), supporting the conclusion that this tumor has sufficient level of proline supply for protein synthesis. Interestingly, examining the proline production pathway of this tumor revealed no up-regulation of PYCR1 or any other related genes (Supplemental Figure S4l), suggesting that this tumor had sufficient precursors for proline production.

Interference with the proline auto-regulator circuit attenuates proliferation of ccRCC cell lines

Thus far, by examining two ccRCC tumors using *diricore* we identified proline deficiency and linked it to shortage in proline production that induces the expression of the proline-producing enzyme PYCR1. To expand on this, we examined whether interference with the proline/PYCR1 regulatory pathway can affect cellular proliferation. For this purpose we used 786-O, a ccRCC cell line expressing relatively high levels of PYCR1 (Supplemental Figure S4h). We used RNA interference to knockdown PYCR1, and cultured cells in a medium with dialyzed serum, either containing or lacking proline to mimic proline shortage. Figure 5a shows potent PYCR1 knockdown in all tested conditions. Figure 5b demonstrates that cells with PYCR1 knockdown present the expected increased level of non-aminoacylated Pro-tRNAs, and that the addition of proline attenuated this effect. Last, we measured the effect of these treatments on cell proliferation. Figure 5c shows that knocking down PYCR1 inhibited the proliferation of the 786-O kidney cancer cell line when incubated in a medium lacking proline. Moreover, this proliferation inhibition by PYCR1 knockdown was almost completely negated when proline was added to the medium. Altogether, these results indicate that at least in the case of kidney cancer cells with high PYCR1 expression, interfering with the intracellular proline production by knocking down PYCR1 attenuates cellular proliferation when proline availability is limiting (Figure 5d). In more general terms, our observations indicate that *diricore* examination of human tumors can identify novel amino acid deficiencies, which can lead to the exploration of novel regulatory pathways, and can be used to investigate novel cancer treatment strategies.

Discussion

We constructed and successfully evaluated *diricore* as a broad and sensitive novel method for detecting deficiencies in amino acid availabilities for protein synthesis. *Diricore* allows systematic and sensitive detection of cellular amino acid deficiencies by comparing global ribosome positions between any biological sample and a suitable control. Particularly, we demonstrated here that in conditions where global protein synthesis is not acutely affected, *diricore* can potentially sense critical changes in extracellular free amino acid levels, detect signals originating from alterations in cellular amino acid uptake and amino acid production, tRNA levels and their availability for protein synthesis, as well as differential metabolic rewiring. Sensing amino acid deficiencies in a growing tumor is likely to be beneficial for cancer diagnosis and can be explored for therapy. Using *diricore* we detected deficiencies in proline in kidney cancer, and linked it to tumors with high expression level of PYCR1, a key enzyme in the proline-producing pathway. By measuring the impact of PYCR1 inhibition in a proline-restrictive environment, our experiments provide evidence

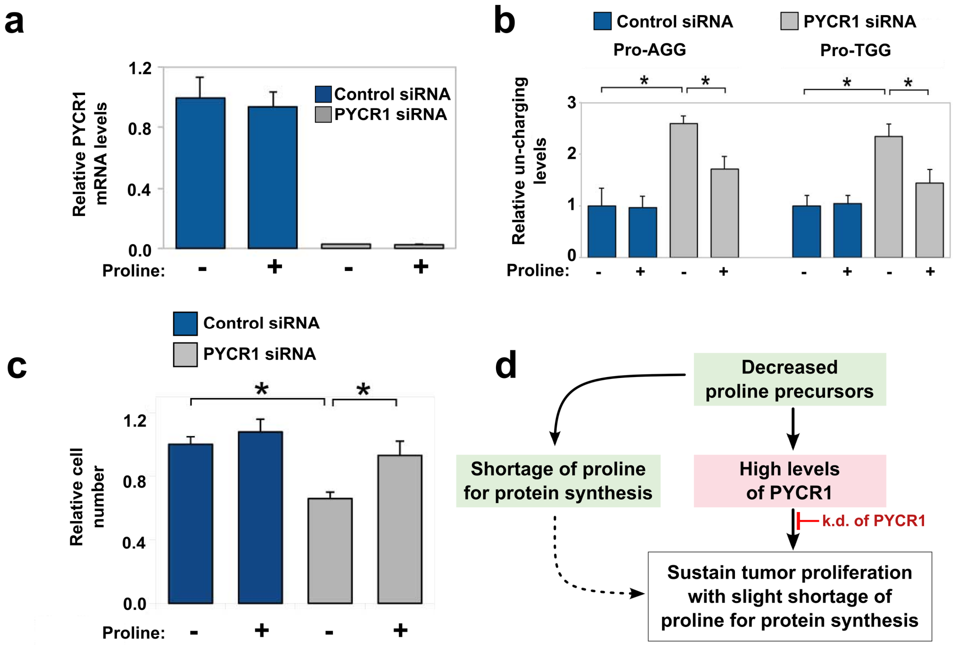


Figure 5: High levels of PYCR1 or sufficient proline are required to sustain proliferation of a ccRCC cell line with high PYCR1 levels.
(a) PYCR1 levels as determined by qRT-PCR upon knockdown of PYCR1 in 786-O cells, expressing high levels of PYCR1 (Supplemental Figure S4h). **(b)** Pro-tRNA aminoacylation analysis was performed on 786-O ccRCC cells subjected to proline deprivation and treated with either PYCR1 or control siRNAs. **(c)** Proliferation assay by cell counting, 72h after treatment. Treatments were as in panel (b). **(d)** A scheme showing how negative effects of decreased levels of proline on proliferation can be negated by high PYCR1 expression. Also indicated is the therapeutic window to inhibit cellular proliferation and tumor growth in cells experiencing shortage of proline for protein synthesis, by reduction of their high PYCR1 expression.

for the potential of *diricore* in diagnosing tumor-specific requirements.

The expanding toolbox of ribosome profiling (Ribo-Seq)

Our findings here expand the utility of Ribo-Seq. Since its invention (Ingolia et al., 2009), Ribo-Seq was extensively used to explore protein translation programs at a genomic scale. Here, Ribo-Seq was utilized as a cellular biological sensor for amino acid deficiencies. We compared ribosome positioning of test and a control conditions by two different ways, “subsequence” and “RPF density”. These two analyses complement each other, as their sensitivity and robustness are different with respect to different aspects of the ribosome profiling data. Comparing subsequences (e.g. codons protected by ribosome’s E, P and A sites) is a measurement of directly interpretable signal. Furthermore, much of the available data can be used, since usually thousands of genes will have many reads in their coding sequence, thus allowing a reasonable gene-wise cutoff to be set. Then, a distribution over the different codons can be taken over many coding sequences of genes, yielding a robust signal. However, this method can use only in-frame reads (without resorting to modeling of overhang due to incomplete RNase digestion), which can range from >90% in high-quality samples to <50% in lower quality samples. In contrast, plotting RPF densities in windows surrounding codons gives visual insight into how RPFs are (differentially) distributed surrounding a codon of interest (most interestingly upstream of the codon). This gives additional insights that are not readily obtained from the subsequence analysis. Moreover, RPF density analysis is relatively insensitive to incomplete digestion of RPFs, even if the degree of digestion is different between samples. However, only when many windows are overlapped and averaged over, the true (differential) distribution becomes apparent. Furthermore, the distribution within a single window is only well estimated if many RPFs are present within that window, thus requiring a cutoff on which windows to average over. This cutoff is generally more stringent than a gene-wise cutoff, and thus less RPFs contribute to the signal, reducing the sensitivity compared to the subsequence analysis. This becomes readily apparent in low-quality datasets where few RPFs covering coding sequences are obtained, which limits the number of windows to average over.

Diricore signals at the initiator-ATG as a measure of alterations in global protein synthesis

By using defined inhibitors of protein production, such as Torin1, nutrient and leucine deprivation, we uncovered a correlation between global reduction in initiation of translation and strong reduction in *diricore* signal at position 12 of the initiator-ATG (Figure 1e). In contrast, treatment with harringtonine, which blocks ribosome elongation, resulted in a strong increase in *diricore* signal at position 12 of the initiator-ATG (Figure 1c). Thus, *diricore* can pinpoint global alterations in protein translation rates, but its interpretation must be taken with care as similar *diricore* signals result from different processes, for example an increased diricore signal at position 12 of the initiator-ATG may result from either accelerating initiation of protein synthesis or blocking of translation elongation. Nevertheless, we used this feature of *diricore* to indicate increase in global initiation of protein synthesis of tumors (Figure 4a). Our experiments also uncover a limitation of *diricore*. Acute global inhibition of protein translation overshadows specific amino acids signals (for example in Figure 1h). However, for the purpose of uncovering amino acid deficiencies in cancer, this feature is not predicted to be restrictive as tumor growth is associated with increase in protein translation. Supporting this notion is our analysis of kidney tumors, which was successful in uncovering at least one restrictive amino acid.

SLC3A2 and TGF-beta

Following TGF-beta treatment and induction of EMT, *diricore* has identified shortage in leucine. At least part of this effect can be explained by reduced expression of the leucine transporter SLC3A2 (also called CD98, which together with SLC7A5 forms the large neutral amino acid transporter LAT1), triggering a reduction in cellular uptake rate of leucine. The mechanism of SLC3A2 inhibition by TGF-beta and its breadth need to be yet resolved. Nevertheless, leucine is an important regulator of mTOR (Jewell and Guan, 2013), and SLC3A2 expression is up-regulated in a wide variety of cancers and correlates with poor prognosis (Cantor and Ginsberg, 2012). How exactly SLC3A2 contributes to tumorigenesis is unclear. In particular, whether the regulation of leucine uptake following TGF-beta bears functional consequences to the EMT process and cancer remains to be explored. However, the elucidation of the causal events leading to reduced leucine uptake following TGF-beta treatment corroborates the finding of leucine deficiency following TGF-beta-induced EMT by *diricore*.

Sensing amino acid deficiencies for cancer diagnosis

At present, one amino acid deprivation treatment is put to practice in ALL using L-asparaginase, and clinical trials are preformed to assess whether arginine deprivation-based treatments can sensitize tumors to conventional and/or targeted therapies. We envision that *diricore* measurements will be able to provide the opportunity to expand the use of L-asparaginase beyond ALL to other tumor types, and also assign individual cases to arginine deprivation treatments for a better outcome. Stratifying patients for such therapy will require testing *diricore* in a large cohort of various types of tumors. Our findings that defined proline deficiency in kidney cancer are important as pioneering case in this area. It has been recently shown that proline generates systematic biases in Ribo-Seq (Artieri and Fraser, 2014). However, as *diricore* bases its analysis on the comparison between two different samples, any systematic bias in Ribo-Seq is normalized. Second, the proline-*diricore* signal in kidney tumor samples did not correlate with the induction of translation, as both analyzed tumors activated translation (as indicated the initiator-ATG diricore analysis, and confirmed by p4E-BP1 staining) but only one of the analyzed tumors showed a proline deficiency signature. Last, examining states of cells with strong difference in global translation rates also showed no proline signal (e.g. Torin1 and nutrient starvation treatments, Figure 1e). Altogether, indicating that the proline signature in the kidney tumor samples is not due to Ribo-Seq biases or changes in global translation rates.

The link between proline deficiency and elevated levels of PYCR1 indicated a regulatory mechanism for proline production to permit sustained growth of the tumor. This information may pave the way for a novel type of therapy. Suppression of PYCR1 expression, for example by PEGylated-PRODH treatment (Proline Dehydrogenase 1; which catalyzes the first step in proline degradation) or through RNA interference tools, may greatly sensitize tumors marked with high PYCR1 and proline deficiencies. In kidney cancer only a small subset of about 4% of the tumors has high PYCR1 expression. However, PYCR1 is observed to be over-expressed in other cancers (e.g. invasive breast carcinoma), suggesting that such a treatment scheme can be beneficial in these types of tumors as well. Intriguingly, Possemato et al have identified PYCR1 in a functional genomic test designed to screen for targets that negatively affect *in vivo* tumorigenesis of human breast cancer cell lines (Possemato et al., 2011). This suggests to us that proline deficiency may impact on the growing tumor following transplantation in mice. A combined shortage of proline for protein synthesis with reduced ability to produce proline may restrict growth of the tumor cells *in vivo*. Here too, screening a cohort of breast cancer tumors with *diricore* should allow the assessment of proline deficiency and its correlation with PYCR1 levels. Altogether, *diricore* is opening new avenues for tumor diagnosis and the discovery of novel amino acid deficiencies in cancer.

Material and methods

Cell culture

MCF10A cells were cultured in DMEM/F12 1:1 medium supplemented with 5% horse serum, EGF (10 ng/ml), insulin (10 µg/ml), cholera toxin (100 ng/ml), and hydrocortisone (500 ng/ml) in 5% CO_2 at 37°C. Small interfering RNAs (siRNAs) against *SLC3A2* were purchased from Life Technologies (Grand Island, NY, USA). MCF10a cells were transfected using Dharmafect I reagent (Dharmacon) following the manufacturer's instructions. For inhibition of mTOR and nutrient starvation experiments, MCF10a cells were treated either with 250 nM of Torin 1 (Tocris Bioscience, Bristol, UK) or with EBSS medium (Sigma) for 2 hours, respectively. For TGFβ1 treatment, MCF10a cells were treated with human recombinant TGFβ1 (10ng/ml) for 48 hours (R&D Systems).

Ribosome profiling (Ribo-Seq)

Libraries from cultured cells were prepared as described previously (Loayza-Puch et al., 2013). For ribosome profiling from tissue, samples were snap frozen after surgical removal and lysed mechanically using a tissue homogenizer in the presence of ice- cold lysis buffer (20 mM Tris-HCl, pH 7.8, 100 mM KCl, 10 mM MgCl2, 1% Triton X-100, 2 mM DTT, 100 µg/ml cycloheximide, 1X EDTA-free Complete protease inhibitors). Lysates were centrifuged at 5,000 rpm and the supernatant was digested with 2 U/µl of RNase I (Life Technologies, Grand Island, NY, USA) for 45 min at room temperature. Resulting monosomes were purified, RNA was isolated, and RP libraries were prepared as described previously. Primers and linkers used in the preparation of libraries are listed in Table S1.

tRNA aminoacylation

Cells were harvested in cold PBS, centrifuged, and cell pellets were resuspended in 0.3 M NaOAc/HOAc (pH 4.5). Total RNA was isolated using acetate-saturated phenol/$CHCl_3$ (pH 4.5). Precipitated RNA was resuspended in 10 mM NaOAc/HOAc (pH 4.5). Samples were split in two, one half (2 µg) was oxidized with 50 mM $NaIO_4$ in 100 mM NaOAc/HOAc (pH 4.5) for 30 min and the other half (2 µg) was incubated in 50 mM NaCl in 100 mM NaOAc/HOAc (pH 4.5) for 30 min. Samples were quenched with glucose 100 mM for 5 minutes at room temperature, purified in G50 columns (GE Healthcare), and then precipitated with ethanol. tRNAs were deacylated in 50 mM Tris-HCl (pH 9) for 30 min at 37 °C. RNA was precipitated and then was ligated to 3'adaptor using T4 RNA ligase 2 (NEB) for 2 hours at 37 °C. Relative aminoacylation levels were calculated by qRT-PCR using tRNA specific primers (Table S1).

Measurement of protein synthesis

$2x10^5$ MCF10a cells were treated with OP-Puro (50 µM, Medchem Source) for 1 hour, cells were washed twice with PBS and then fixed in 1% paraformaldehyde in PBS for 15 min on ice. Cells were washed with cold PBS and permeabilized in 0.1% Triton-X in PBS supplemented with 2% fetal calf serum (FCS) for 5 min at room temperature. The azide-alkyne cycloaddition was performed using the Click-iT Cell Reaction Buffer Kit (Life Technologies) and azide conjugated to Alexa Fluor 488 at 5 µM final concentration. Cells were incubated for 30 min at room temperature, then cells were washed three times in PBS with 2% FCS and analysed by FACS.

Leucine and aspartic acid uptake

Cells were plated in 12-well plates ($1x10^5$ cells/well). When the cells reached 80% confluence, they were washed twice in PBS and incubated for 5 min either in sodium-free uptake buffer (4.8 mM KCl, 1.3 mM CaCl2, 1.2 mM MgSO4, 25 mM HEPES, 1.2 mM KH2PO4, 5.6 mM glucose, pH 7.4) for Leu or in PBS for Asp.

Cells were incubated for 5 min with either [3H] L-leucine or [3H] L-Aspartic acid. Uptake was finished by washing the cells three times with ice-cold uptake solution or PBS. Cells were solubilized in 0.1 N NaOH, and radioactivity was counted by liquid scintillation.

Northern blot

4 µg of total RNA was separated on a 10% denaturing polyacrylamide gel and transferred onto a nylon membrane (Hybond-N+; Amersham Biosciences). The membrane was UV-crosslinked, dried, and hybridized using radiolabeled RNA probes in NorthernMax buffer (Ambion). Riboprobes were synthesized using the miRNA probe construction kit (Ambion) following the manufacturer's instructions. Oligonucleotide sequences are listed in Table S1.

Real time PCR

1 µg of total RNA was reverse transcribed using the SuperScript III first-strand synthesis system (Life Technologies) following the manufacturer's instructions. Real time PCR was performed using the SensiFAST SYBR real time PCR kit (Bioline). Primers used in this study are listed in Table S1.

Deep sequencing data preprocessing and alignment

Adapter sequences were trimmed from raw RP data using cutadapt (Martin, 2011). Sequences shorter than 20bp after adapter trimming were discarded. rRNA and tRNA sequences were filtered by alignment to indices of rRNA and tRNA sequences respectively, using bowtie2 (Langmead and Salzberg, 2012) and default parameters. A rRNA index was constructed from GENCODE v19 annotations, transcript types "rRNA", "Mt_rRNA" and "rRNA_pseudogene", supplemented with UCSC repeats of class "rRNA" (table "rmsk"). The tRNA index was constructed from sequences obtained from GtRNAdb (Chan and Lowe, 2009) at 5th of September 2012. The remaining (unmapped) reads were aligned to GRCh37/hg19 using TopHat2 (Kim et al., 2013) and GENCODE v19/BASIC (Harrow et al., 2012) transcript coordinates (parameters “-N 2 -m 1 --no-novel-juncs --no-novel-index --no-coverage-search --segment-length 25”). In subsequent analyses, only primary alignments with mapping quality of 10 or greater were considered. RP and RNAseq data followed the same pipeline, with the exception of adapter trimming, which was only performed on RP data. Data presented in this manuscript is available with GEO accession number GSE59821.

Subsequence analysis

In summary, subsequence shift analysis compares RPF codon occupancy frequencies between samples, in a gene-level normalized manner (to exclude gene expression differences as cause for differences in observed codon frequencies).

Specifically, gene IDs and reading frames were assigned to RPFs, using GENCODE v19/BASIC. RPFs not within a valid CDS (taking into account the 15nt. 5'-overhang of RPFs), RPFs with ambiguous gene ID and RPFs with ambiguous reading frame (determined by the aligned position of 5'-end of the RPF) were excluded. The remaining RPFs were used to count, for the different positions (12 and 15nt. from 5'-end), the frequencies of all codons, for all genes in the transcriptome. Within-gene frequencies were calculated by dividing the observed counts by the total counts for that gene. Normalized codon frequencies were averaged over all genes that had at least *n* counts in both the condition and control samples (n is 100 for all figures). Shifts were calculated from these normalized and averaged frequencies per codon as: (condition – reference) / reference. I.e. per codon shifts were calculated relative to the control. To test for significant differential codon occupancy at the amino acid level, we used the subsequence shifts between control and conditions over the replicates. We used a linear mixed model (R package 'lme4') with fixed effects over the 20 amino acids and random effects over the codons. This supplied us with t-values and p-values. We used Benjamini-Hochberg multiple testing correction (R function ‘p.adjust’, method “fdr”) to extract adjusted p-values.

RPF density analysis

Codon-regions of 61 nucleotides (nt) width around designated codons along the transcriptome (as annotated by GENCODE v19/BASIC) were identified (using transcript coordinates; i.e. codon-regions were all supported by exons). Overlapping transcript annotations were used, and overlapping regions were retained, but regions with identical genomic coordinates were collapsed. Codon-regions that could not be extended to 61 nt (due to being near the 5' or 3' ends of transcripts) were discarded.

5' ends of RPFs were counted for each codon-region. For a comparison between two samples, only codon-regions where the total number of counts was at least 100 in both samples were taken into consideration. For each sample and each codon-region, the normalized 5' end RPF density was calculated by dividing over the total number of counts within that region, and multiplying by the width of the codon-region, so the average density within each codon region equaled 1. The normalized densities were convolved using a rectangular window of width 3 and height 1/3. The mean density over the codon-regions was taken and the difference in the mean densities between the samples yielded to the density shift.

Normalized abundances and translational efficiencies

Gene-wise counts were performed using annotations from GENCODE v19/BASIC and HTSeq. Counts were normalized by correction for the total library sizes using trimmed-mean of M-values (TMM) normalization, supplied by the R-package ‘edgeR’. Translational efficiencies were calculated as the ratio of (normalized abundance determined by RP) / (normalized abundance determined by RNAseq).

GSEA

To obtain a RankedList as input for GSEA (Subramanian et al., 2005), gene-wise counts (generated using GENCODE v19/BASIC and HTseq) were normalized using edgeR's (Robinson et al., 2010) TMM normalization. Predictive fold-changes were obtained using edgeR's predFC routine with a prior count of 3. Ensembl/GENCODE gene IDs were converted to Entrez gene IDs using a look-up table obtained from Ensembl MartView. Fold-changes of genes that mapped to a single Entrez gene ID were averaged. In GSEA, the predictive fold-changes were used a scores. The gene sets analyzed contained all MSigDB/C2 (v4.0) gene sets. Default settings were used, with the exception that 10000 permutations were used to estimate the significance of enrichments.

Codon demand calculations

Codon demand was calculated by calculation of the absolute abundances of each codon in the transcriptome, as determined by RNAseq. Reads per gene ID were counted using GENCODE v19/BASIC annotations and HTSeq. Gene-wise abundances were normalized by the total library sizes. Per-gene isoform abundances were estimated using MISO (Katz et al., 2010). The total numbers of codons present per transcript were calculated using the CDS annotations. Shifts in codon demand were calculated relative to the reference, i.e. the shifts were calculated as (condition - reference) / reference.

Processing of yeast data

Yeast ribosome profiling data of cells treated with 3-AT (Guydosh and Green, 2014) was obtained from GEO (samples GSM1279568 and GSM1279579). Adapter sequences were trimmed using cutadapt (Martin, 2011). Sequences shorter than 20bp after adapter trimming were discarded. rRNA and tRNA sequences were filtered by alignment to indices of rRNA and tRNA sequences respectively. A rRNA index was constructed from Ensembl annotations (EF4.69), transcript type "rRNA"). Similarly, a tRNA index was constructed from Ensembl annotations (EF4.69), transcript type "tRNA"). Remaining (unmapped) reads were aligned to the yeast genome (EF4.69) using corresponding Ensembl transcript annotations and TopHat2 (Kim et al., 2013) (using parameters "-N 2 -m 1 --no-novel-juncs --no-novel-index --no-coverage-search --segment-length 25"). In subsequent analyses, only primary alignments with mapping quality of 10 or greater were considered.

Acknowledgments

We thank Petur Snaebjornsson for valuable help with the pathology studies, Roy Zent for kindly providing the SLC3A2 expression vectors, Ron Kerkhoven, Roel Kluin and Marja Nieuwland from the NKI Genomics Core Facility for assistance with deep sequencing experiments, the NKI Radionuclides Centre and the NKI flow cytometry facility for assistance with experiments, and all the members of the Agami group for valuable discussion. This work was supported by funds from EMBO long-term fellowships to A.P.U, and the Netherlands Organization for Scientific Research (NWO-VICI) and the Dutch cancer society (KWF) to R.A.

References

Artieri, C.G., and Fraser, H.B. (2014). Accounting for biases in riboprofiling data indicates a major role for proline in stalling translation. Genome Res. *24*, 2011–2021.

Bazzini, A.A., Johnstone, T.G., Christiano, R., Mackowiak, S.D., Obermayer, B., Fleming, E.S., Vejnar, C.E., Lee, M.T., Rajewsky, N., Walther, T.C., et al. (2014). Identification of small ORFs in vertebrates using ribosome footprinting and evolutionary conservation. EMBO J. *33*, 981–993.

Cantor, J.M., and Ginsberg, M.H. (2012). CD98 at the crossroads of adaptive immunity and cancer. J. Cell. Sci. *125*, 1373–1382.

Chan, P.P., and Lowe, T.M. (2009). GtRNAdb: a database of transfer RNA genes detected in genomic sequence. Nucleic Acids Res. *37*, D93–D97.

Dalgliesh, G.L., Furge, K., Greenman, C., Chen, L., Bignell, G., Butler, A., Davies, H., Edkins, S., Hardy, C., Latimer, C., et al. (2010). Systematic sequencing of renal carcinoma reveals inactivation of histone modifying genes. Nature *463*, 360–363.

Delage, B., Fennell, D.A., Nicholson, L., McNeish, I., Lemoine, N.R., Crook, T., and Szlosarek, P.W. (2010). Arginine deprivation and argininosuccinate synthetase expression in the treatment of cancer. Int. J. Cancer *126*, 2762–2772.

Efeyan, A., and Sabatini, D.M. (2013). Nutrients and growth factors in mTORC1 activation. Biochem. Soc. Trans. *41*, 902–905.

Feun, L., and Savaraj, N. (2006). Pegylated arginine deiminase: a novel anticancer enzyme agent. Expert Opin Investig Drugs *15*, 815–822.

Fresno, M., Carrasco, L., and Vazquez, D. (1976). Initiation of the polypeptide chain by reticulocyte cell-free systems. Survey of different inhibitors of translation. Eur. J. Biochem. *68*, 355–364.

Guo, H., Ingolia, N.T., Weissman, J.S., and Bartel, D.P. (2010). Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature *466*, 835–840.

Guydosh, N.R., and Green, R. (2014). Dom34 rescues ribosomes in 3’ untranslated regions. Cell *156*, 950–962.

Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Zadissa, A., Searle, S., et al. (2012). GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res. *22*, 1760–1774.

Hsieh, A.C., Liu, Y., Edlind, M.P., Ingolia, N.T., Janes, M.R., Sher, A., Shi, E.Y., Stumpf, C.R., Christensen, C., Bonham, M.J., et al. (2012). The translational landscape of mTOR signalling steers cancer initiation and metastasis. Nature *485*, 55–61.

Ingolia, N.T. (2014). Ribosome profiling: new views of translation, from single codons to genome scale. Nat. Rev. Genet. *15*, 205–213.

Ingolia, N.T., Ghaemmaghami, S., Newman, J.R.S., and Weissman, J.S. (2009). Genome-wide analysis in vivo of translation with nucleotide resolution using ribosome profiling. Science *324*, 218–223.

Ingolia, N.T., Lareau, L.F., and Weissman, J.S. (2011). Ribosome profiling of mouse embryonic stem cells reveals the complexity and dynamics of mammalian proteomes. Cell *147*, 789–802.

Ishimura, R., Nagy, G., Dotu, I., Zhou, H., Yang, X.-L., Schimmel, P., Senju, S., Nishimura, Y., Chuang, J.H., and Ackerman, S.L. (2014). RNA function. Ribosome stalling induced by mutation of a CNS-specific tRNA causes neurodegeneration. Science *345*, 455–459.

Jain, M., Nilsson, R., Sharma, S., Madhusudhan, N., Kitami, T., Souza, A.L., Kafri, R., Kirschner, M.W., Clish, C.B., and Mootha, V.K. (2012). Metabolite profiling identifies a key role for glycine in rapid cancer cell proliferation. Science *336*, 1040–1044.

Jewell, J.L., and Guan, K.-L. (2013). Nutrient signaling to mTOR and cell growth. Trends Biochem. Sci. *38*, 233–242.

Katz, Y., Wang, E.T., Airoldi, E.M., and Burge, C.B. (2010). Analysis and design of RNA sequencing experiments for identifying isoform regulation. Nat. Methods *7*, 1009–1015.

Kim, J., and Guan, K.-L. (2011). Amino acid signaling in TOR activation. Annu. Rev. Biochem. *80*, 1001–1032.

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*, R36.

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

Leprivier, G., Remke, M., Rotblat, B., Dubuc, A., Mateo, A.-R.F., Kool, M., Agnihotri, S., El-Naggar, A., Yu, B., Somasekharan, S.P., et al. (2013). The eEF2 kinase confers resistance to nutrient deprivation by blocking translation elongation. Cell *153*, 1064–1079.

Liu, B., Han, Y., and Qian, S.-B. (2013). Cotranslational response to proteotoxic stress by elongation pausing of ribosomes. Mol. Cell *49*, 453–463.

Liu, J., Xu, Y., Stoleru, D., and Salic, A. (2012). Imaging protein synthesis in cells and tissues with an alkyne analog of puromycin. Proc. Natl. Acad. Sci. U.S.A. *109*, 413–418.

Liu, J., Ma, J., Wu, Z., Li, W., Zhang, D., Han, L., Wang, F., Reindl, K.M., Wu, E., and Ma, Q. (2014). Arginine deiminase augments the chemosensitivity of argininosuccinate synthetase-deficient pancreatic cancer cells to gemcitabine via inhibition of NF-κB signaling. BMC Cancer *14*, 686.

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

Maddocks, O.D.K., Berkers, C.R., Mason, S.M., Zheng, L., Blyth, K., Gottlieb, E., and Vousden, K.H. (2013). Serine starvation induces stress and p53-dependent metabolic remodelling in cancer cells. Nature *493*, 542–546.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal *17*, 10.

Offman, M.N., Krol, M., Patel, N., Krishnan, S., Liu, J., Saha, V., and Bates, P.A. (2011). Rational engineering of L-asparaginase reveals importance of dual activity for cancer cell toxicity. Blood *117*, 1614–1621.

Phillips, M.M., Sheaff, M.T., and Szlosarek, P.W. (2013). Targeting arginine-dependent cancers with arginine-degrading enzymes: opportunities and challenges. Cancer Res Treat *45*, 251–262.

Possemato, R., Marks, K.M., Shaul, Y.D., Pacold, M.E., Kim, D., Birsoy, K., Sethumadhavan, S., Woo, H.-K., Jang, H.G., Jha, A.K., et al. (2011). Functional genomics reveal that the serine synthesis pathway is essential in breast cancer. Nature *476*, 346–350.

Richards, N.G.J., and Kilberg, M.S. (2006). Asparagine synthetase chemotherapy. Annu. Rev. Biochem. *75*, 629–654.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Rooijers, K., Loayza-Puch, F., Nijtmans, L.G., and Agami, R. (2013). Ribosome profiling reveals features of normal and disease-associated mitochondrial translation. Nat Commun *4*, 2886.

Sheen, J.-H., Zoncu, R., Kim, D., and Sabatini, D.M. (2011). Defective regulation of autophagy upon leucine deprivation reveals a targetable liability of human melanoma cells in vitro and in vivo. Cancer Cell *19*, 613–628.

Singh, B., Tai, K., Madan, S., Raythatha, M.R., Cady, A.M., Braunlin, M., Irving, L.R., Bajaj, A., and Lucci, A. (2012). Selection of metastatic breast cancer cells based on adaptability of their metabolic state. PLoS ONE *7*, e36510.

Son, J., Lyssiotis, C.A., Ying, H., Wang, X., Hua, S., Ligorio, M., Perera, R.M., Ferrone, C.R., Mullarky, E., Shyh-Chang, N., et al. (2013). Glutamine supports pancreatic cancer growth through a KRAS-regulated metabolic pathway. Nature *496*, 101–105.

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. U.S.A. *102*, 15545–15550.

Takaku, H., Takase, M., Abe, S., Hayashi, H., and Miyazaki, K. (1992). In vivo anti-tumor activity of arginine deiminase purified from Mycoplasma arginini. Int. J. Cancer *51*, 244–249.

The Cancer Genome Atlas Research Network (2013). Comprehensive molecular characterization of clear cell renal cell carcinoma. Nature *499*, 43–49.

Thiele, D.L., and Lipsky, P.E. (1990). The action of leucyl-leucine methyl ester on cytotoxic lymphocytes requires uptake by a novel dipeptide-specific facilitated transport system and dipeptidyl peptidase I-mediated conversion to membranolytic products. J. Exp. Med. *172*, 183–194.

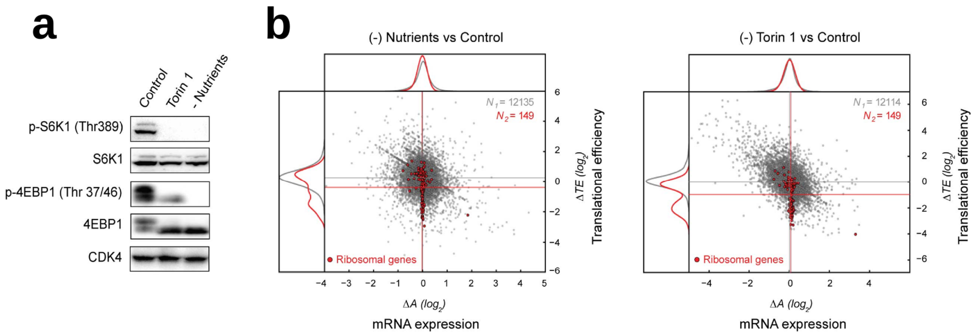
Thoreen, C.C., Kang, S.A., Chang, J.W., Liu, Q., Zhang, J., Gao, Y., Reichling, L.J., Sim, T., Sabatini, D.M., and Gray, N.S. (2009). An ATP-competitive mammalian target of rapamycin inhibitor reveals rapamycin-resistant functions of mTORC1. J. Biol. Chem. *284*, 8023–8032.

Thoreen, C.C., Chantranupong, L., Keys, H.R., Wang, T., Gray, N.S., and Sabatini, D.M. (2012). A unifying model for mTORC1-mediated regulation of mRNA translation. Nature *485*, 109–113.

Zaborske, J.M., Narasimhan, J., Jiang, L., Wek, S.A., Dittmar, K.A., Freimoser, F., Pan, T., and Wek, R.C. (2009). Genome-wide analysis of tRNA charging and activation of the eIF2 kinase Gcn2p. J. Biol. Chem. *284*, 25254–25267.

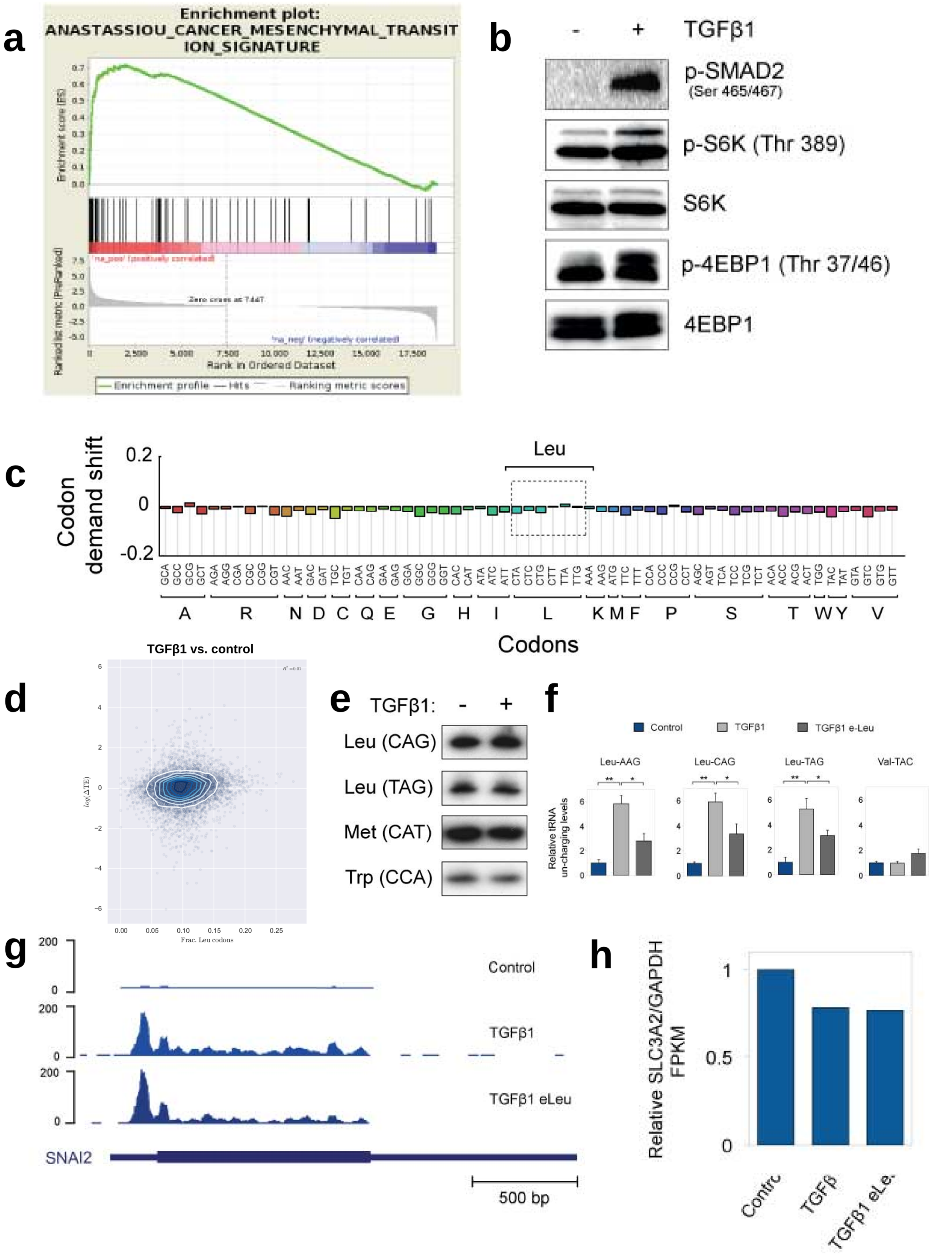
Zoncu, R., Efeyan, A., and Sabatini, D.M. (2011). mTOR: from growth signal integration to cancer, diabetes and ageing. Nat. Rev. Mol. Cell Biol. *12*, 21–35.

Supplemental figures



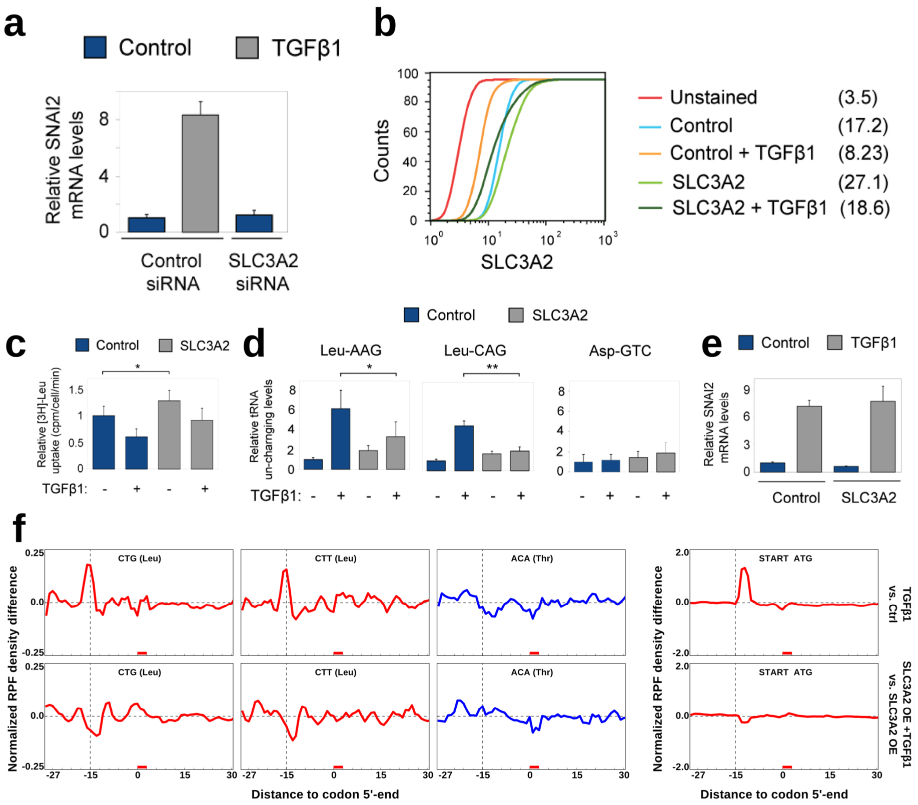
Supplemental Figure S1: Suppression of initiation of protein synthesis upon nutrient starvation and Torin1 treatment.

(a) Immunoblots from MCF10A cells that were treated either with DMSO (control), 250 nM Torin 1, or starved for nutrients for 2h. CDK4 is used as loading control. **(b)** Changes in translational efficiencies of ribosomal proteins in response to Torin 1 and nutrient starvation.



Supplemental Figure S2: Transcriptional and translational response of MCF10A cells to TGF-beta treatment.

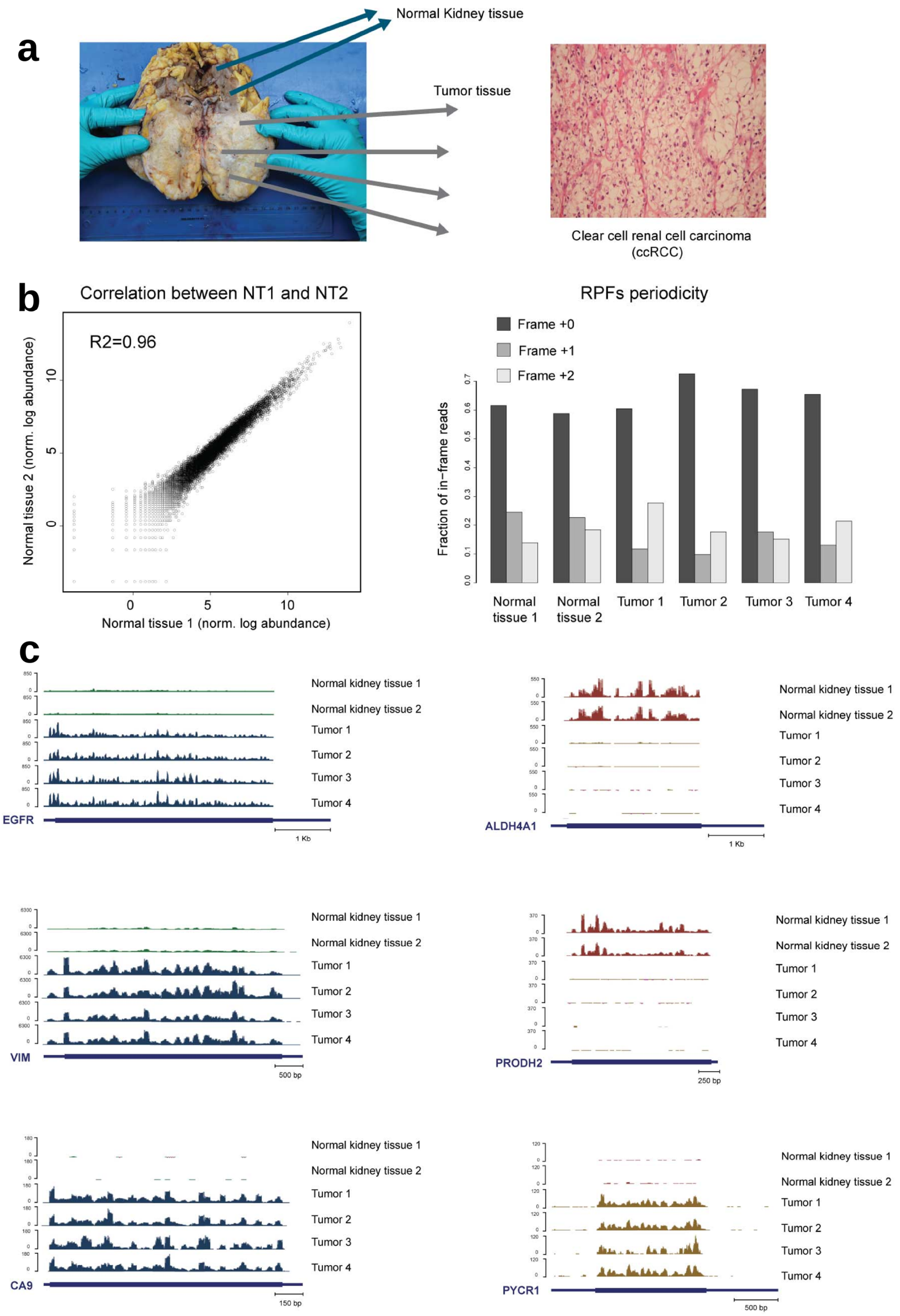
(a) The human non-transformed epithelial MCF10A cells were treated with TGF-beta for 48 hours. Global gene expression analysis verifies TGF-beta pathway induction and reveals expression patterns indicative of epithelial-to-mesenchymal transition (EMT). **(b)** Western blot analysis of the indicated mTOR target proteins. **(c)** Codon demand analysis based on transcriptome profiling data (measured by RNA-seq) from one replicate of Figure 2a. **(d)** Gene-wise translational efficiency changes upon TGF-beta treatment calculated from one replicate of Figure 2a, plotted against the gene-wise fraction of leucine codons. RNA-seq of the control condition was used to determine the most abundant isoform per gene, which was used to calculate the fraction of leucine codons per gene. **(e)** Northern blot analysis for the indicated tRNAs of MCF10A cells treated as in Figure 2a. **(f)** tRNA aminoacylation assay on MCF10A cells treated with TGF-beta and either esterified-leucine (e-Leu) or control vehicle. **(g)** Ribo-Seq-based expression analysis of SNAI2 in MCF10A cells either untreated, TGF-beta treated or TGF-beta and eLeu treated. **(h)** Relative expression of SLC3A2 in MCF10A cells treated as in panel (g), determined by Ribo-Seq.

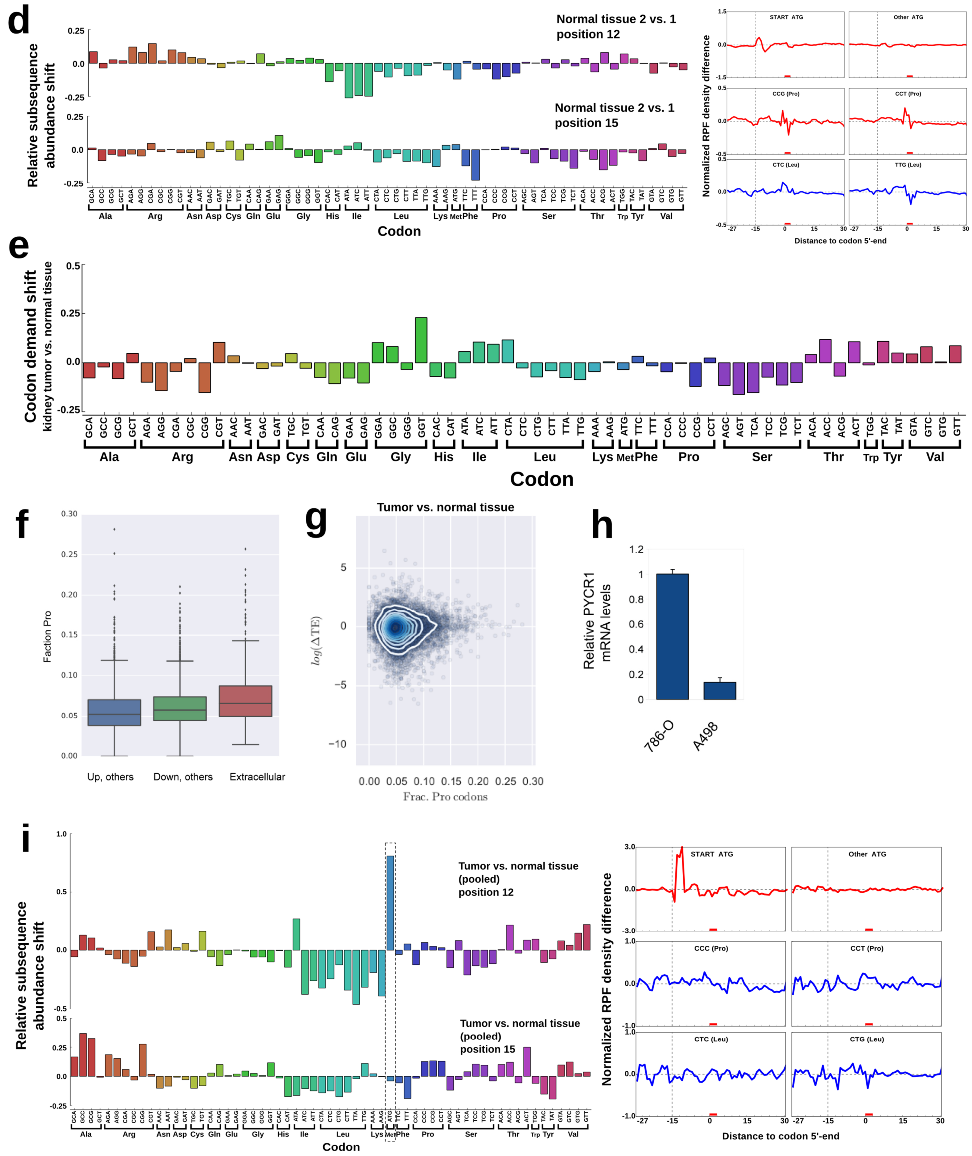


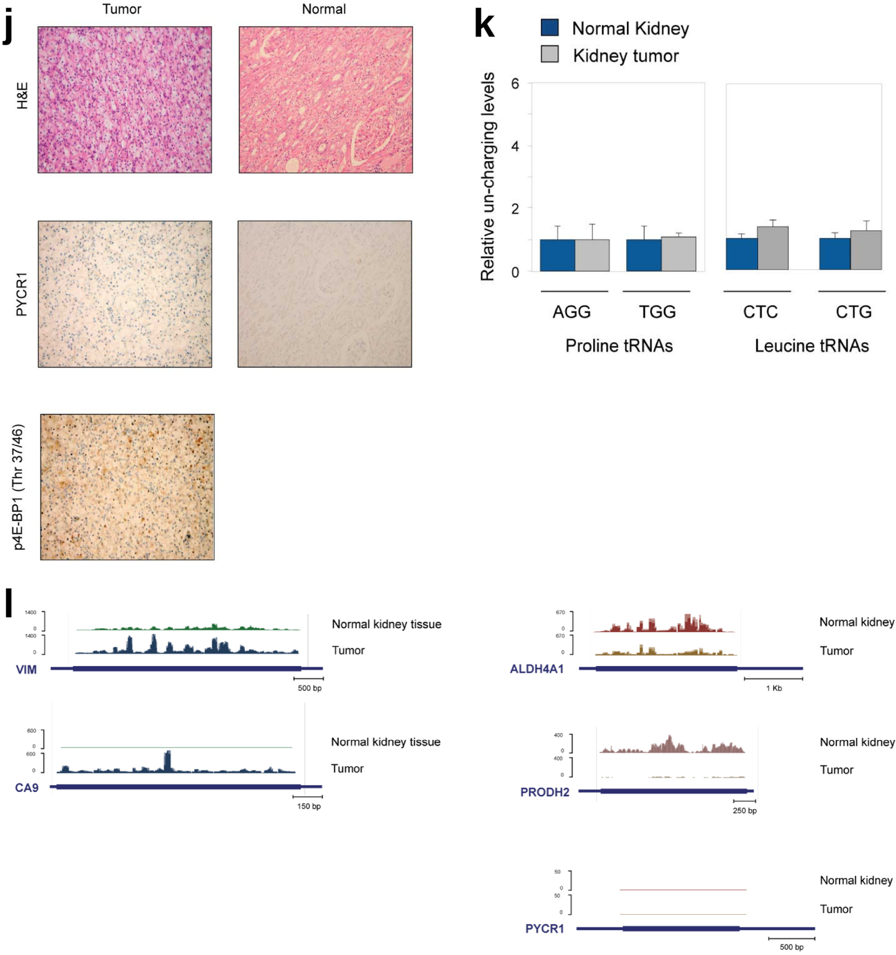
Supplemental Figure S3: SLC3A2 knockdown and over-expression reveal role in TGF-beta-induced ribosomal stalling on leucine codons.

(a) Relative levels of SNAI2 mRNA, a major target of TGF-beta signaling, as determined by qRT-PCR in cells treated as in Figure 3c. Expression is normalized to GAPDH mRNA. **(b)** SLC3A2 expression determined by FACS of stable control and SLC3A2-expressing polyclonal MCF10A populations, in either control or 48h TGF-beta treated conditions. **(c)** [^{3}H]-Leu uptake analysis of the same MCF10A populations and conditions as in panel (b). **(d)** Aminoacylation assay of the indicated tRNAs in the same MCF10A populations and conditions as in panel (b). **(e)** Relative levels of SNAI2 mRNA as determined by qRT-PCR in the same MCF10A populations and conditions as in panel (b). Expression is normalized to GAPDH mRNA. **(f)** RPF density analysis of the same MCF10A populations and conditions as in panel (b) at leucine and threonine (negative control) codons.

NB: For panels (c), (d), and (e), error bars represent SD; $n = 3$; $^{**}p < 0.01$, $^{*}p < 0.05$ by two-tailed Student's t test.


a
Normal Kidney tissue
Tumor tissue
Clear cell renal cell carcinoma (ccRCC)
b
Correlation between NT1 and NT2
R2=0.96
Normal tissue 2 (norm. log abundance)
Normal tissue 1 (norm. log abundance)
0
5
10
RPFs periodicity
Frame +0
Frame +1
Frame +2
Fraction of in−frame reads
0.0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
Normal tissue 1
Normal tissue 2
Tumor 1
Tumor 2
Tumor 3
Tumor 4
c
Normal kidney tissue 1
Normal kidney tissue 2
Tumor 1
Tumor 2
Tumor 3
Tumor 4
EGFR
1 Kb
850
ALDH4A1
1 Kb
550
VIM
500 bp
6300
PRODH2
250 bp
370
CA9
150 bp
180
PYCR1
500 bp
120


d
Relative subsequence abundance shift
Normal tissue 2 vs. 1 position 12
Normal tissue 2 vs. 1 position 15
0.25
0.0
-0.25
GCA GCC GCG GCT AGA AGG CGA CGC CGG CGT AAC AAT GAC GAT TGC TGT CAA CAG GAA GAG GGA GGC GGG GGT CAC CAT ATA ATC ATT CTA CTC CTG CTT TTA TTG AAA AAG ATG TTC TTT CCA CCC CCG CCT AGC AGT TCA TCC TCG TCT ACA ACC ACG ACT TGG TAC TAT GTA GTC GTG GTT
Ala Arg Asn Asp Cys Gln Glu Gly His Ile Leu Lys Met Phe Pro Ser Thr Trp Tyr Val
Codon
Normalized RPF density difference
START ATG
Other ATG
CCG (Pro)
CCT (Pro)
CTC (Leu)
TTG (Leu)
1.5
0.0
-1.5
0.5
-0.5
-27
-15
0
15
30
Distance to codon 5'-end
e
Codon demand shift
kidney tumor vs. normal tissue
0.5
0.25
0.0
-0.25
Codon
f
Faction Pro
0.30
0.25
0.20
0.15
0.10
0.05
0.00
Up, others
Down, others
Extracellular
g
Tumor vs. normal tissue
log(ΔTE)
5
0
−5
−10
0.00 0.05 0.10 0.15 0.20 0.25 0.30
Frac. Pro codons
h
Relative PYCR1 mRNA levels
1.2
1
0.8
0.6
0.4
0.2
0
786-O
A498
i
Relative subsequence abundance shift
Tumor vs. normal tissue (pooled) position 12
Tumor vs. normal tissue (pooled) position 15
1.0
0.5
0.0
-0.5
0.25
-0.25
Codon
Normalized RPF density difference
START ATG
Other ATG
CCC (Pro)
CCT (Pro)
CTC (Leu)
CTG (Leu)
3.0
0.0
-3.0
1.0
-1.0
-27
-15
0
15
30
Distance to codon 5'-end



Supplemental Figure S4: Quality control on ccRCC tumor sample #1 and data generated from a second ccRCC tumor sample.

(a) Left: overall presentation of clear cell renal cell carcinoma (ccRCC). Arrows indicate regions where the samples of Figure 4a were taken. Right: Representative microscopic images of tumor sections stained with hematoxylin and eosin (H&E). Pathological analysis confirmed a clear cell renal cell carcinoma (ccRCC). **(b)** Expression correlation between normal tissues (left) and RPF periodicity analysis of the Ribo-Seq data obtained from normal and tumor tissues (right). **(c)** Gene expression analysis by Ribo-Seq confirms high level of EGFR, VIM, and CA9 in the tumor samples compared with the normal tissues, as was reported for renal cell carcinomas (The Cancer Genome Atlas Research Network, 2013). Changes in gene expression levels of the proline metabolic pathway are shown in support of Figure 4d. **(d)** *Diricore* analysis comparing the two normal tissue samples. **(e)** Codon demand analysis based on transcriptome profiling data of the samples of Figure 4a (measured by RNA-seq). **(f)** Proline content (y-axis) of genes either up-regulated, down-regulated or annotated to be extracellular. Differential expression was determined by RNA-seq comparing tumor samples to normal tissue samples, using the R package 'edgeR' and an FDR of 0.05 as cutoff. **(g)** Gene-wise translational efficiency changes comparing a representative tumor sample to a normal tissue sample of Figure 4a, plotted against the gene-wise fraction of proline codons. RNA-seq of the pooled normal tissue samples was used to determine the most abundant isoform per gene, which was used to calculate the fraction of proline codons per gene. **(h)** qRT-PCR analysis of the relative expression of PYCR1 in 786-O and A498 ccRCC cell lines, normalized to GAPDH. **(i)** *Diricore* analysis of a second ccRCC tumor. **(j)** IHC analysis of p4E-BP1 threonine 37/46 and PYCR1 on samples of a second ccRCC tumor (samples from panel (i)) as in Figure 4b. **(k)** Aminoacylation assay of the indicated tRNAs in the samples of the second ccRCC tumor (panel (i)). **(l)** Gene expression as determined by Ribo-Seq in the second ccRCC tumor (panel (i)) of ccRCC marker genes (left) and genes in the proline metabolic pathway (right).

Supplemental Table 1: Oligos used in this chapter.

qRT-PCR		
Target gene	**Forward primer**	**Reverse primer**
GAPDH	ACCCAGAAGACTGTGGATGG	TCTAGACGGCAGGTCAGGTC
SLC3A2	CGTGGTTCTCCACTCAGGTT	GTTAGTCCCCGCAATCAAGA
SNAI2	GGGGAGAAGCCTTTTTCTTG	TCCTCATGTTTGTGCAGGAG
PYCR1	CATGACCAACACTCCAGTCG	CCTTGGAAGTCCCATCTTCA
tRNA--Leu (AAG)	GGTAGCGTGGCCGAGCGGTCTA	GCCTTGGCACCCGAGAATTCCA
tRNA--Leu (CAA)	GTCAGGATGGCCGAGTGGTCTA	GCCTTGGCACCCGAGAATTCCA
tRNA--Leu (CAG)	GTCAGGATGGCCGAGCGGTCTA	GCCTTGGCACCCGAGAATTCCA
tRNA--Leu (TAA)	ACCAGAATGGCCGAGTGGTTAA	GCCTTGGCACCCGAGAATTCCA
tRNA--Leu (TAG)	GGTAGCGTGGCCGAGTGGTCTA	GCCTTGGCACCCGAGAATTCCA
tRNA--Val (AAC)	GTTTCCGTAGTGTAGTGGTCA	GCCTTGGCACCCGAGAATTCCA
tRNA--Val (CAC)	GTTTCCGTAGTGTAGCGGTTATC	GCCTTGGCACCCGAGAATTCCA
tRNA--Val (TAC)	GGTTCCATAGTGTAGTGGTTAT	GCCTTGGCACCCGAGAATTCCA
tRNA--Leu (GTC)	TCCTCGTTAGTATAGTGGTGAG	GCCTTGGCACCCGAGAATTCCA
tRNA--Arg (ACG)	GGGCCAGTGGCGCAATGGATAAC	GCCTTGGCACCCGAGAATTCCA
tRNA--Arg (CCG)	GGCCGCGTGGCCTAATGGATAA	GCCTTGGCACCCGAGAATTCCA
t--RNA--Pro (AGG)	GGCTCGTTGGTCTAGGGGTATG	GCCTTGGCACCCGAGAATTCCA
t--RNA--Pro (TGG)	CAGAACTATAGGAATTGAACCTA	GCCTTGGCACCCGAGAATTCCA
Ribosome profiling libraries		
5' RNA Adaptor	GUUCAGAGUUCUACAGUCCGACGAUC	
RNA 3'Adapter	/5rApp/TGGAATTCTCGGGTGCCAAGG/3ddC/	
Reverse transcription Primer	GCCTTGGCACCCGAGAATTCCA	
PCR Primer	AATGATACGGCGACCACCGAGATCTACACGTTCAGAGTTCTACAGTCCGA	
Index 1	CAAGCAGAAGACGGCATACGAGATCGTGATGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA	
Index 2	CAAGCAGAAGACGGCATACGAGATACATCGGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA	
Index 3	CAAGCAGAAGACGGCATACGAGATGCCTAAGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA	
Index 4	CAAGCAGAAGACGGCATACGAGATTGGTCAGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA	
Index 5	CAAGCAGAAGACGGCATACGAGATCACTGTGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA	
Index 6	CAAGCAGAAGACGGCATACGAGATATTGGCGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA	
Index 7	CAAGCAGAAGACGGCATACGAGATGATCTGGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA	
Index 8	CAAGCAGAAGACGGCATACGAGATTCAAGTGTGACTGGAGTTCCTTGGCACCCGAGAATTCCA	

CHAPTER 4

Specific changes of ribosome distribution over messenger-RNAs during mitotic arrest

Koos Rooijers[1], Fabricio Loayza-Puch[1], Lenno Krenning[2], René Medema[2] and Reuven Agami[1,3]

1 Division of Gene Regulation, The Netherlands Cancer Institute
Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands
2 Division of Cell Biology, The Netherlands Cancer Institute
Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands
3 Erasmus MC, Rotterdam University,
Dr Molewaterplein 50, 3015 GE Rotterdam, The Netherlands

Introduction

Mitosis is a crucial step in cellular reproduction at which the replicated DNA is divided equally in order to be deposited in two daughter cells. Even though mitosis is generally the shortest phase of the mammalian cell cycle, it is accompanied by drastic changes to the cellular organization. Importantly, to segregate the duplicated chromosomes the nuclear membrane is broken down, allowing the condensed genetic material to be divided. The division of condensed chromosomes is orchestrated after intricate attachment to and alignment by a tailored part of the cytoskeleton.

Because of its obvious involvement in cellular proliferation mitosis has been well studied in the context of cancer. Many regulators of the various steps have been long known, such as the mitotic cyclins and the anaphase-promoting complex (Peters, 2006). Regulation of mRNA translation has also been investigated, and for example phosphorylation of eEF2 was shown to be modulated during mitosis (Sivan et al., 2007), as well as 4E-BP (Pyronnet et al., 2001), thereby putting a halt to cap-dependent translation. Importantly, genes involved in the mitotic process have been hypothesized to escape this global repression of translation (Pyronnet et al., 2001). In order to study regulation at the translational level, ribosome profiling has become the de facto standard technique, as it has shown great sensitivity combined with the ability to comprehensively profile the translation of the transcriptome of a sample (Ingolia et al., 2009). Using ribosome profiling, several translational programs have been uncovered (Gonzalez et al., 2014; Stumpf et al., 2013), the speed of translation elongation has been estimated (Ingolia et al., 2011) and recently we used ribosome profiling to study the impact of a mitochondrial tRNA mutation on translation and ribosome progression (Rooijers et al., 2013). Moreover, we used ribosome profiling as a tool to look “further upstream” at metabolic causes that cause differential tRNA availabilities (Chapter 3 of this thesis). At its core, the commonly performed ribosome profiling protocol is similar to RNA sequencing. However, by isolation of ribosomes attached to mRNAs, nuclease digestion to remove unprotected bases and obtain monosomes, and isolation of ribosome protected mRNA fragments, before proceeding with generation of sequencing libraries, the location of ribosomes on the mRNA can be mapped with nucleotide precision (Ingolia et al., 2012). This resolution has been invaluable in interpretation of differential ribosome codon occupancies, which allows analysis of regulation at the level of tRNAs and further upstream. The obtained ribosome protected fragments (RPFs) have an average length of ~30nt, which corresponds with the dimensions of ribosomes. As core component of the translational machinery, ribosomes take up activated (aminoacylated) tRNAs, match the anti-codon of the tRNA to the codon in the mRNA, facilitate the incorporation of the tRNAs' amino acid moiety into the peptide chain, and finally ejects the unloaded tRNA. The progression of these steps take place while tRNAs move from one of three binding pockets in the ribosome to the next binding pocket, until finally being released. The first pocket the tRNA encounters is the A-site, which functions as accepting site in the ribosome for the next (correct) tRNA. Movement to the next pocket, the P-site, results in peptidyltransferase activity, and the tRNA in the P-site holds the full nascent peptide chain. Finally, the last pocket in the ribosome, the E-site, allows exit of the tRNA from the ribosome complex. Uptake of (correct) tRNAs into the ribosome is generally the rate-limiting step, and inferring the position of the A-site within RPFs allowed for instance identification of reduced tRNA availability (Rooijers et al., 2013).

To investigate the level of translational regulation during mitosis we made use of ribosome profiling, combined with RNA-sequencing to identify transcript abundances. We used U2OS cells as model system, which when treated with cytotoxic drugs such as nocodazole exhibit complete mitotic arrest. We also investigated mitotic arrest using cytotoxic-independent methods. We show that during mitosis, a strong global repression of translational elongation takes place. We also show that ribosomes occupy codons in their A-sites differentially, and this is unrelated to the global repression of elongation. Lastly, we identify several functional groups of genes that escape the global repression of elongation, or are otherwise translationally induced, indicative of the existence of the hypothesized translational programs during mitosis.

Results

Treatment of U2OS cells with nocodazole leads to alterations in ribosomal codon occupancies

To identify novel regulatory modes of mRNA translation we performed ribosome profiling (RP) and transcriptome profiling (RNA-seq) on control and nocodazole-treated U2OS cells (Figure 1a). After processing of the sequencing data, we validated the quality and observed high correlation between independent biological replicates (Supplemental Figure S1a) and good correlation between RP and RNA-seq on the same samples (Supplemental Figure S1b). Furthermore, by characterizing the differentially expressed genes upon nocodazole treatment we could confirm a transcriptional response typical for a cell population enriched for mitotic cells (Figure 1b and 1c).

Due to the nucleotide resolution of the ribosome profiling data (supported by the disequilibrium of proportion of reads in the three reading frames, see Supplemental Figure S1c) the majority of ribosome protected fragments (RPFs) are in-frame with the CDS of the transcript they occupy. Since the dimensions and tRNA-binding pockets (E, P and A-sites) at RPFs can be inferred from the data and are well-documented (O’Connor et al., 2013; Rooijers et al., 2013) we were able to extract, count and analyze the codons occupied in the different ribosomal tRNA-binding sites. We found that when treated with nocodazole, ribosomes occupy several codons differentially, especially at the A- and P-site (Figure 1d). These differences were strong enough to robustly discriminate between control and nocodazole-treated samples based on the ribosomal codon occupancies alone (Figure 1e). Note that these differences are not due to differentially expressed genes since we included only genes in our analyses with at least 100 RPFs across all samples, and applied intra-gene count normalization of occupied codons prior to averaging over the genes, thereby negating any effect of changes in gene expression levels.

To investigate whether the differences in codon occupancies could be contributed by differential amino acid availability, we calculated codon occupancy log-fold-differences (wrt. the control condition) upon nocodazole treatment and arranged the codons by amino acid (Figure 1f). We noticed several amino acids that exhibit significant differential occupancy over their corresponding codons at both the A- and P-site, namely alanine, arginine, aspartic acid, leucine and proline. To further investigate the behavior of ribosomes at the codons coding for these amino acids we generated RPF density profiles in windows of 61 nt. surrounding the codons of these amino acids. We found that the RPF density profiles were consistently changing in Asp codons GAC and GAT, where RPFs shift from being 15nt. upstream in control condition (Figure 1g, blue) to 12nt. in nocodazole treated condition (Figure 1g, green). On the other hand, RPF density shifts were inconsistent for the codons of the amino acids arginine and proline (see Supplemental Figures S1d and S1e).

The RPF density averages at aspartic acid codons suggest that a rate-limiting step is imposed in the control condition when aspartic codons occupy the A-sites of ribosomes. However, following nocodazole treatment this rate-limiting step seems exchanged for a rate-limiting step where the codons occupy the P-site.

Codon occupancy shift is microtubule-disruption independent and unrelated to overall RPF positioning

To test whether the distinct occupancy of codons upon nocodazole treatment is due to the enrichment of mitotic cells or a consequence of the interference with microtubules, we performed ribosome profiling on samples enriched for mitotic cells by treatments which do not cause microtubule disruption. Using S-trityl-L-cysteine (STLC, an Eg5 inhibitor), BI2536 (an inhibitor of Plk1) and Taxol (like nocodazole a microtubule toxin), we observed a codon occupancy fold-change pattern comparable to nocodazole treatment (Figure 2a) and analysis of RPF 5'-end densities at e.g. aspartic acid codons again showed a shift of the codon from ribosomal A-site to P-site (Figure 2b).

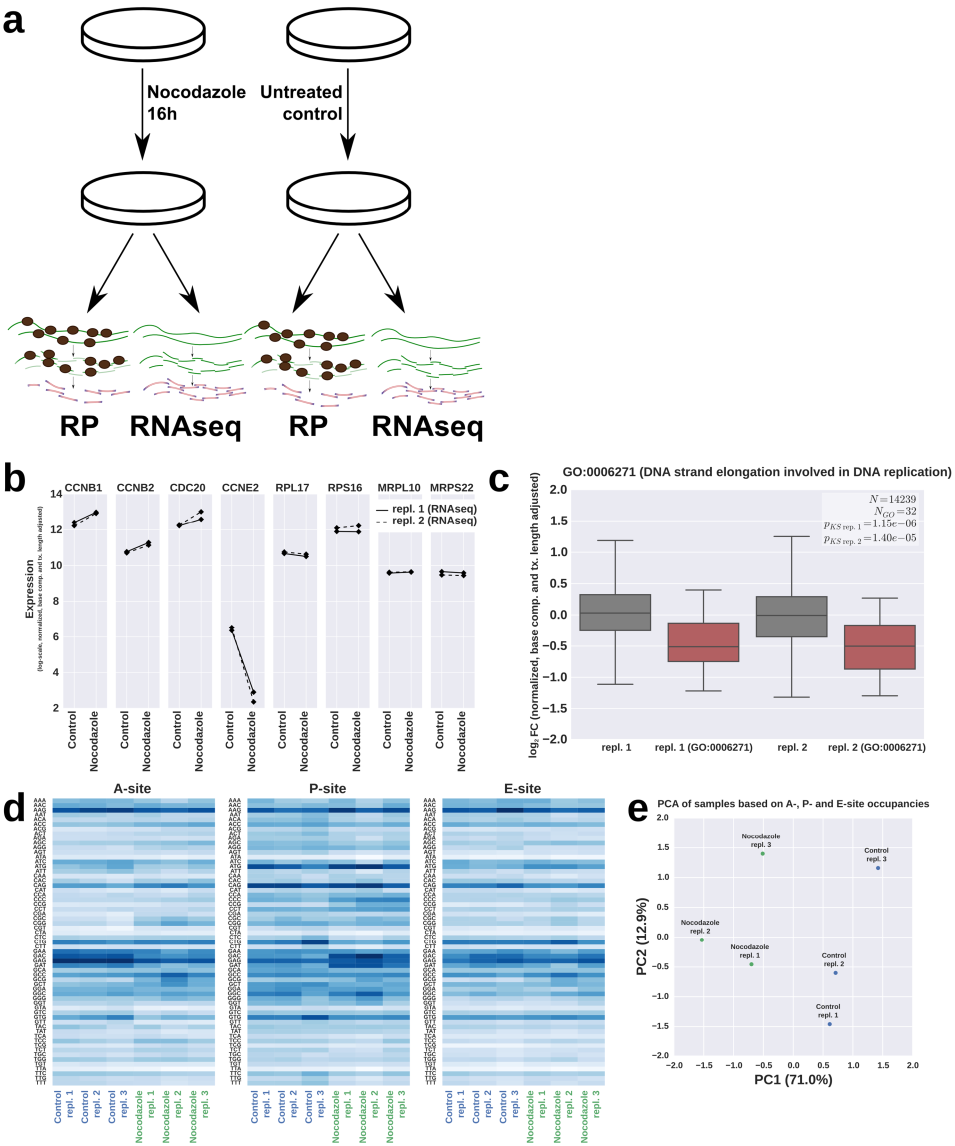


Figure 1: treatment of U2OS cells with nocodazole leads to reproducible changes in ribosome codon occupancies.

(a) Schematic overview of experimental setup. Cells are either treated with nocodazole, or grown as untreated controls. Of each treatment, the transcriptome is profiled using both ribosomal profiling and RNA sequencing. **(b)** mRNA levels (y-axis, log scale) determined by RNA sequencing for various mitotic marker genes. Solid and dotted lines indicate biological replicates. RPL17, RPS16, MRPL10 and MRPS22 are ribosomal protein genes that serve as controls. **(c)** Expression log fold-changes (nocodazole vs. control) of genes in GO term “DNA strand elongation involved in DNA replication” for two biological replicates. **(d)** Heatmap of normalized and intra-gene averaged counts of codon frequencies in the inferred ribosomal A-, P- and E-site for control and nocodazole-treated samples in 3 independent biological replicates. **(e)** Principal component analysis of nocodazole and control samples in 3 independent biological repeats based on the ribosomal codon occupancies (see panel (d)). The percentages in the parenthesis indicate the variance explained by the first two principal components.

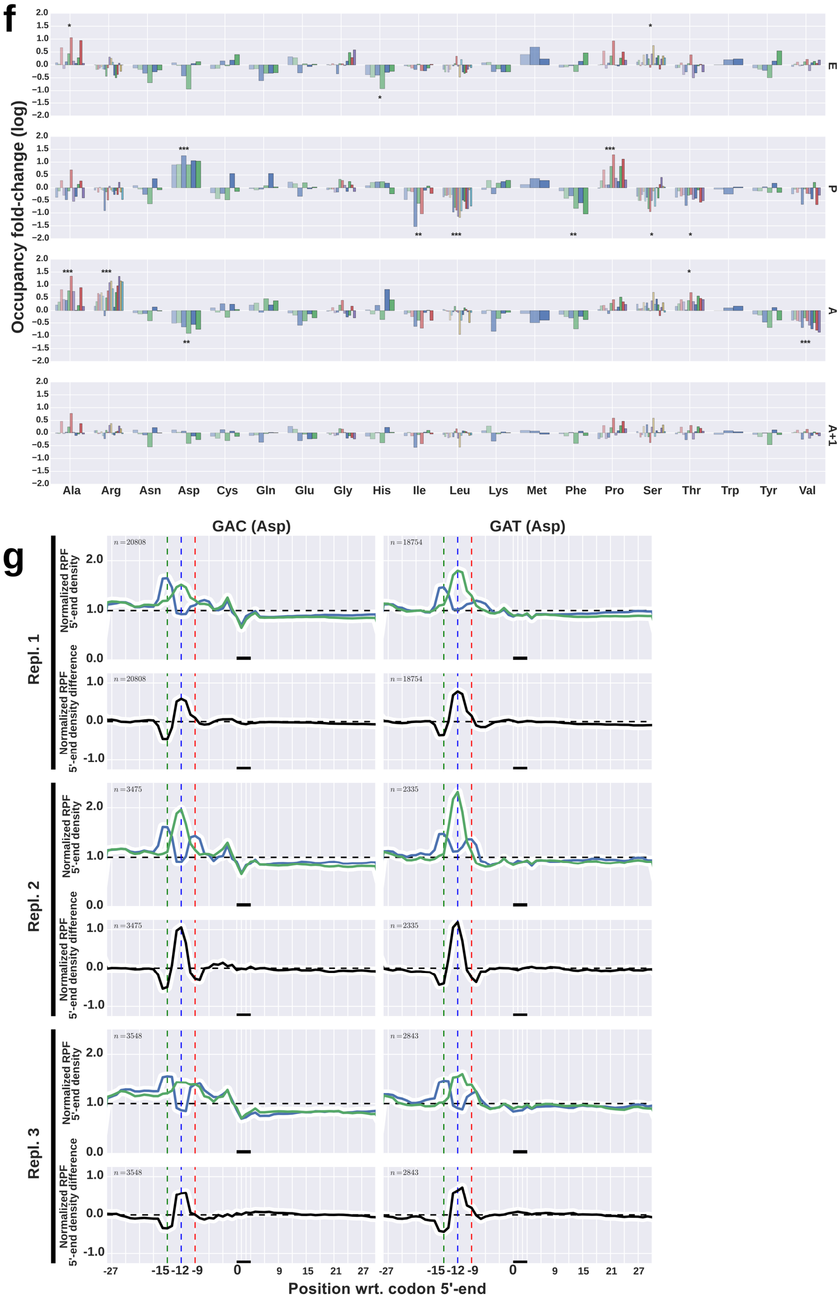


Figure 1 (cont.)

(f) Fold changes of codon occupancies arranged by amino acid. Different colors indicate different codons. 3 different independent biological repeats are displayed in 3 different shades. Statistical significance of differential occupancy at the amino acid level is determined by a mixed linear model with fixed effects over the 20 amino acids and random effects over the codons. Significance thresholds are: * $p < 0.05$, ** $p < 0.01$ and *** $p < 0.001$ and are adjusted for multiple testing. **(g)** RPF 5'-end density profiles at the two aspartic acid (Asp) codons, GAC and GAT, in 3 independent biological replicates. Blue lines display the density in control samples, green lines display the density in nocodazole-treated samples, black lines display the difference between nocodazole and control samples.

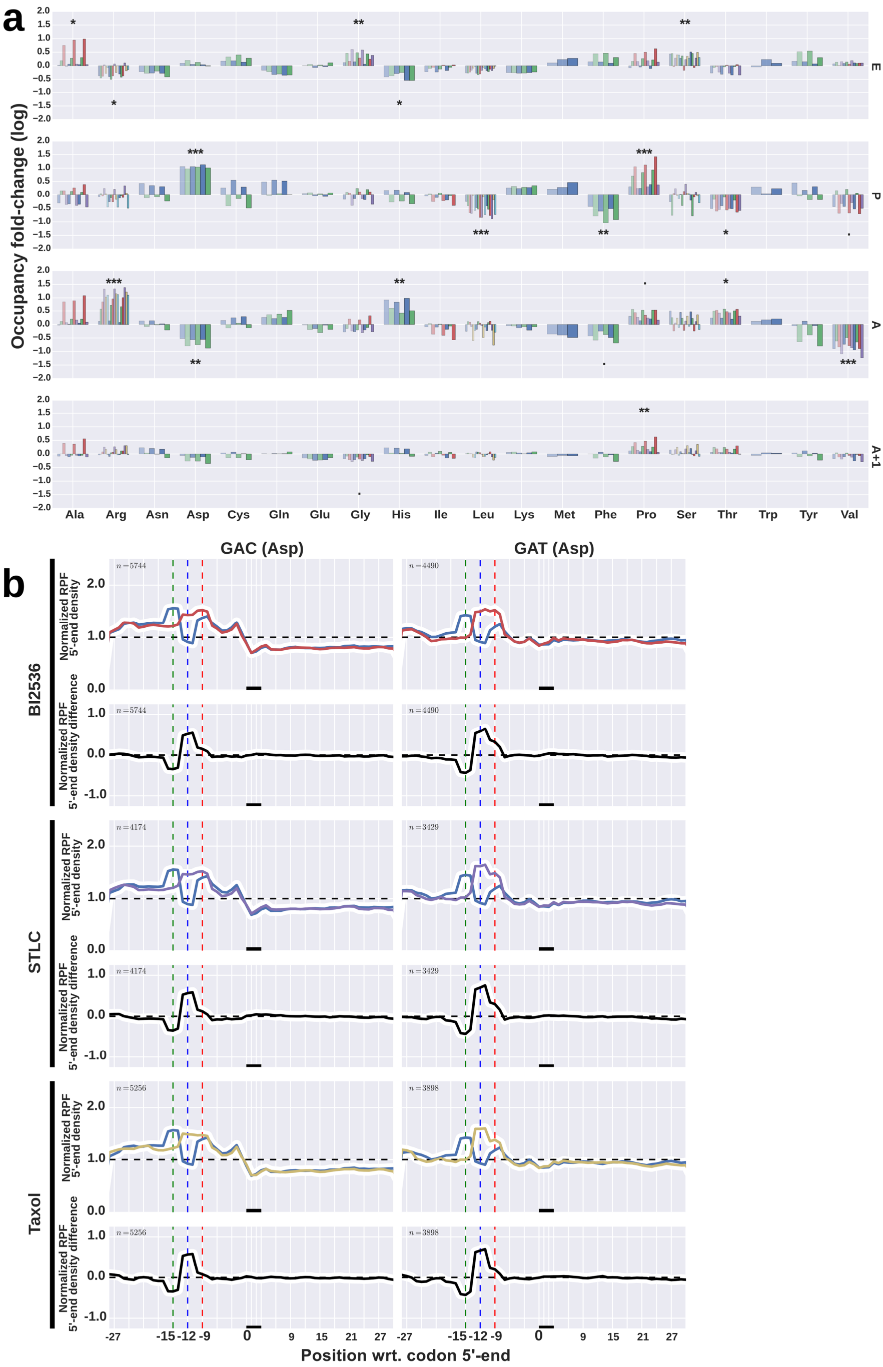
Readily observed in the RP data is a shift of RPFs towards the 5'-end of the CDS in mitotic samples, see Figure 2c for ribosomal coverage of *GAPDH* as an example. We quantified this shift transcriptome-wide in Figures 2d and 2e, indicating that this is a widespread and microtubule disruption-independent phenomenon. This observation corresponds to an inhibition of elongation (Sivan et al., 2007). To test whether the distinct occupancy of codons of mitotic cells is caused by a general shift of RPFs to the 5'-ends of CDSs, we made use of a published dataset on heatshock-treated 293T cells (Shalgi, 2013). Heatshock treatment was shown to cause a similar shift of ribosomes to 5'-ends of CDSs, a consequence of reduced availability of protein chaperones (see Figure 2f, analyzed analogously to Figures 2d and 2e). However, codon occupancy analysis shows that despite a large global shift of ribosomes to 5'-ends of CDSs, heatshock causes very little changes in codon occupancy (Figures 2g and 2h).

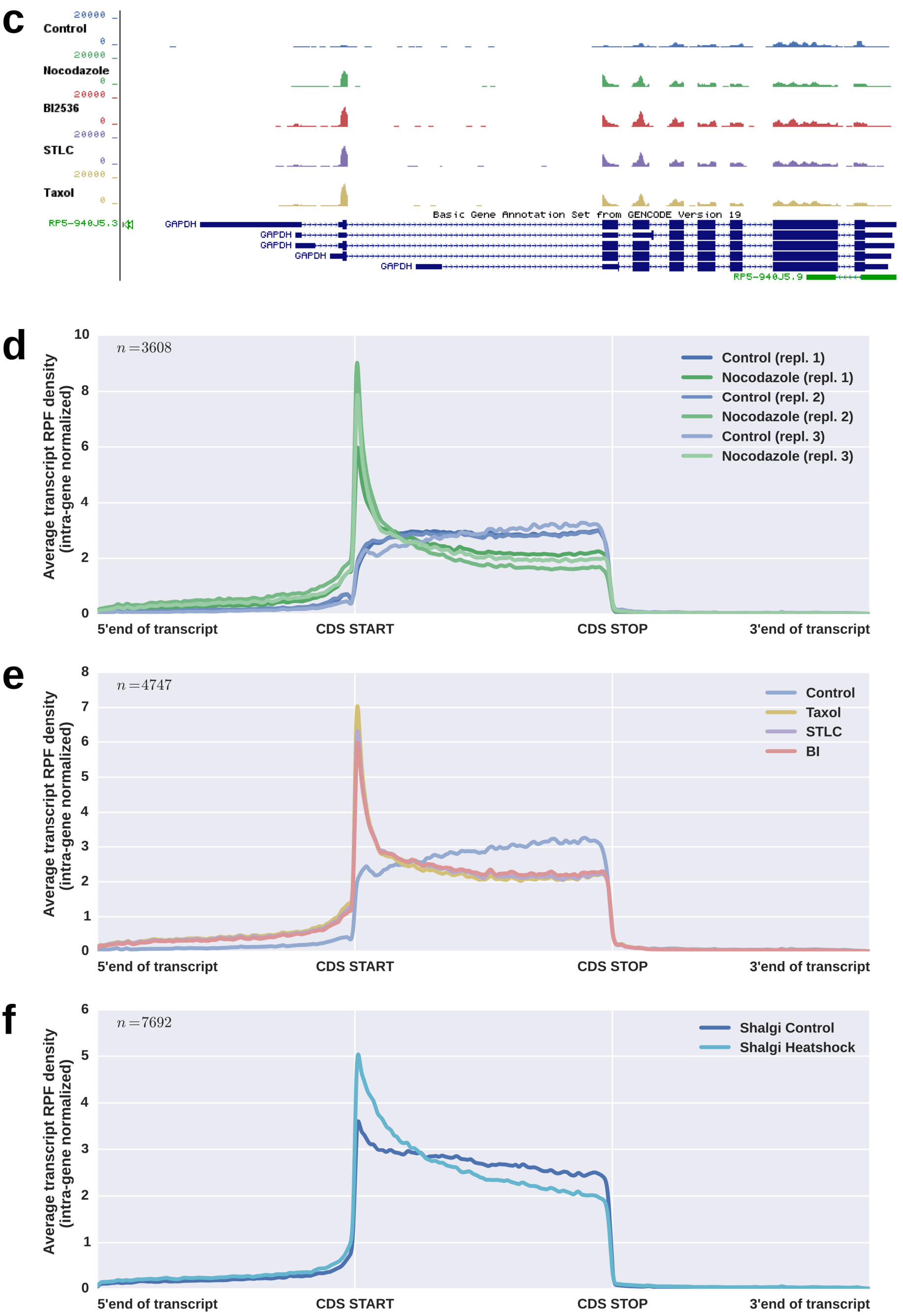
To further test whether the global shift of ribosomes to 5'-ends of CDSs is related to the aspartic acid codons A-to-P-shift, we partitioned CDSs in quartiles and measured RPF density at aspartic acid codons in these quartiles. We found that RPFs surrounding aspartic codons were distributed independently of the codons' position in the CDS (Supplemental Figures S2a and S2b). Specifically, the stalling of ribosomes with Asp codons in their A-site in the control condition is independent of position in the CDS, as well as the Asp A-to-P-shift that is observed upon induction of mitosis.

tRNA abundances are modulated in mitotic cells

Next, we set out to identify the role of tRNA abundance in the differential codon occupancy in mitotic cells. While some signals were consistent at all codons coding for the same amino acid (e.g. aspartic acid codons GAC and GAT), many amino acids showed signals that were reproducible in our biological replicates, but were not consistent across the different codons coding for that amino acid (e.g. arginine codons). Consistency might result from a differential availability of conjugatable amino acid moieties. Additionally, we recognized that for the case of aspartic acid codons the consistency can be derived from the two codons being translated by the same tRNA (due to the wobble position).

Profiling of the tRNA abundances in control and mitotic cells was carried out by smallRNAseq, where the transcriptome of RNA molecules smaller than approx. 200nt is subjected to deep-sequencing. Analysis of the full-length tRNA sequences in this data indicated widespread differential expression of tRNAs that was consistent across two different treatments (Figure 3a, comparing nocodazole and STLC treatments log-fold changes to control). It did not, however, provide an explanation for the differential ribosomal codon occupancies. As can be seen in Figure 3a, tRNA-Asp-GTC is relatively enriched in both nocodazole and STLC-treated conditions compared to control cells, which could explain the ribosomal A-to-P-shift on Asp codons.


a
Occupancy fold-change (log)
2.0
1.5
1.0
0.5
0.0
−0.5
−1.0
−1.5
−2.0
E
P
A
A+1
*
**
***
Ala
Arg
Asn
Asp
Cys
Gln
Glu
Gly
His
Ile
Leu
Lys
Met
Phe
Pro
Ser
Thr
Trp
Tyr
Val
b
GAC (Asp)
GAT (Asp)
BI2536
STLC
Taxol
Normalized RPF 5'-end density
Normalized RPF 5'-end density difference
n = 5744
n = 4490
n = 4174
n = 3429
n = 5256
n = 3898
2.0
1.0
0.0
-1.0
-27
-15
-12
-9
0
9
15
21
27
Position wrt. codon 5'-end


c
Control
20000
0
Nocodazole
20000
0
BI2536
20000
0
STLC
20000
0
Taxol
20000
0
Basic Gene Annotation Set from GENCODE Version 19
RP5-940J5.3
GAPDH
GAPDH
GAPDH
GAPDH
GAPDH
RP5-940J5.9
d
$n = 3608$
Average transcript RPF density (intra-gene normalized)
10
8
6
4
2
0
Control (repl. 1)
Nocodazole (repl. 1)
Control (repl. 2)
Nocodazole (repl. 2)
Control (repl. 3)
Nocodazole (repl. 3)
5'end of transcript
CDS START
CDS STOP
3'end of transcript
e
$n = 4747$
Average transcript RPF density (intra-gene normalized)
8
7
6
5
4
3
2
1
0
Control
Taxol
STLC
BI
5'end of transcript
CDS START
CDS STOP
3'end of transcript
f
$n = 7692$
Average transcript RPF density (intra-gene normalized)
6
5
4
3
2
1
0
Shalgi Control
Shalgi Heatshock
5'end of transcript
CDS START
CDS STOP
3'end of transcript

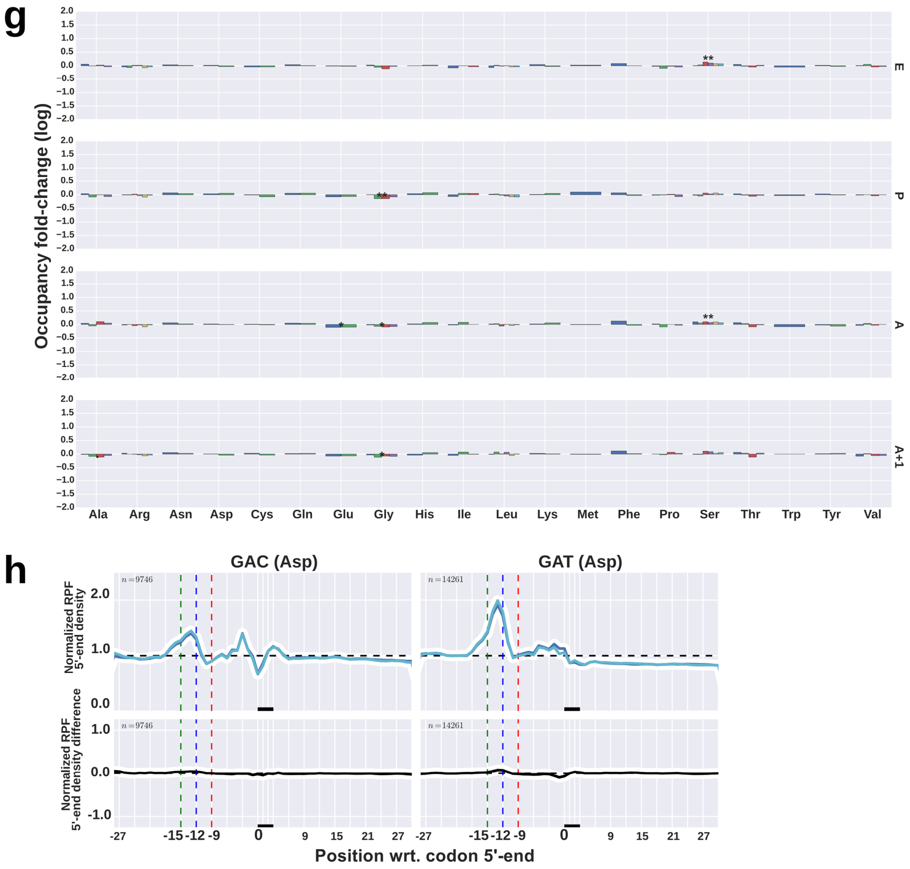


Figure 2: Codon occupancy shifts are microtubule-disruption independent and unrelated to overall RPF positioning.
(a) Fold changes of codon occupancies arranged by amino acid, as in Figure 1f. Shades indicate BI2536-, STLC- and taxol-treatments vs. control, in order of increasing opacity. **(b)** RPF 5'-end density density profiles for aspartic acid codons as in Figure 1g for control and BI2536-, STLC- and taxol-treated samples. **(c)** Genomic view of RPF density on GAPDH in control and nocodazole-, BI2536-, STLC- and taxol-treated samples. **(d)** Transcriptome-wide average of RPF density along transcripts for control and nocodazole-treated samples for 3 independent biological replicates. **(e)** Transcriptome-wide average of RPF density, as in panel (d), for control and BI2536-, STLC- and taxol-treated samples. **(f)** Transcriptome-wide average of RPF density, as in panel (d), for control and heatshock-treated samples. **(g)** Fold changes of codon occupancies of heatshock-treated sample vs. control, arranged by amino acid, as in Figure 1F. **(h)** RPF 5'-end density density profiles for aspartic acid codons as in Figure 1g for control and heatshock treated samples.

However, we observed tRNA-Glu-CTC and tRNA-Glu-TTC molecules to be enriched to a higher degree, while the codons translated by these tRNAs do not show an A-to-P shift (Supplemental Figure S3).

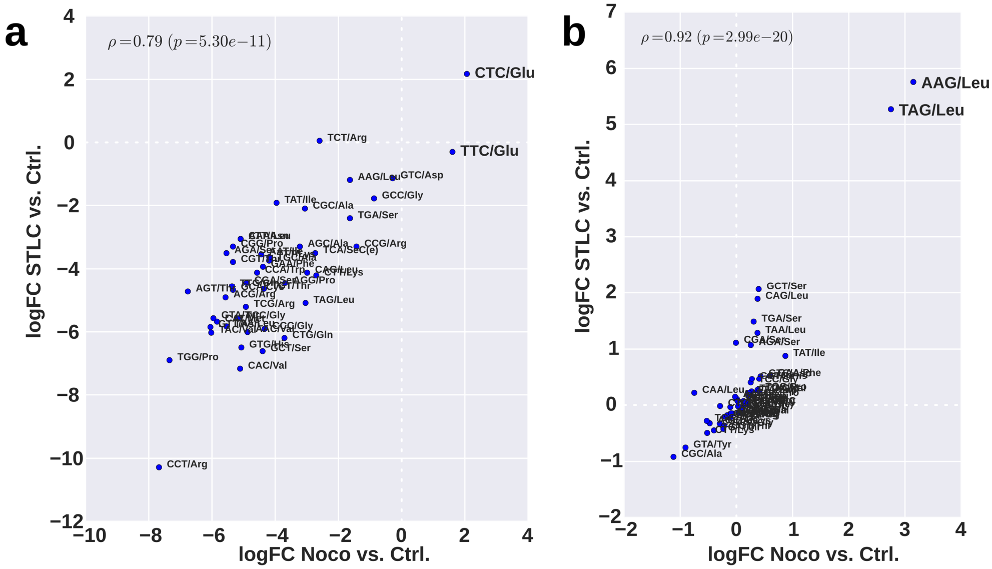


Figure 3: tRNA abundances are modulated in mitosis.
(a) Full-length tRNA abundance log-fold changes as determined by small RNA sequencing, where log-fold changes between nocodazole and control samples are plotted on the horizontal axis, and log-fold changes between STLC and control samples are plotted on the vertical axis. **(b)** tRNA abundance log-fold changes as determined by deep sequencing protocol tailored to tRNA profiling. As in panel (a), nocodazole and control samples are plotted on the horizontal axis, and log-fold changes between STLC and control samples are plotted on the vertical axis.

Importantly, smallRNA sequencing provided us with a relatively small fraction of reads of full-length tRNAs (less than 2% in each sample), while most reads were derived from snRNAs, and to a lesser degree from miRNAs and rRNAs.

In order to increase sensitivity of our profiling of tRNA sequences, we adopted a protocol where RNA was first size-selected for the range typical for tRNAs (70-100 nt.), followed by fragmentation in order to reduce the negative effects that the secondary and tertiary structure of tRNAs might have on amplification and sequencing. The dataset we generated was strongly enriched for tRNAs (at least 90% of the reads in the data were identified as tRNA fragments in each sample) and provided a much greater depth of profiling. Furthermore, the mitotic samples again showed good correlation (see Figure 3b). However, this dataset also did not show correlation with the differential ribosomal codon occupancies.

Mitosis-induced RPF positioning shift and translational efficiency changes are associated, suggesting programs of translational regulation

Having performed RNAseq alongside RP on two of our replicate nocodazole treatment experiments, we were able to uncover levels of translational regulation by analysis of gene-wise translational efficiencies (TEs). Conventional analysis of TEs (Ingolia et al., 2009;

Stumpf et al., 2013) involves calculating a ratio of reads obtained by RP to reads obtained by RNAseq (both normalized for sequencing depth). We used this definition of TE and calculated TE changes upon treatment with nocodazole, compared to control cells. In our replicates, the changes in TE show high correlation (Pearsons R 0.69, Figure 4a). Furthermore, we quantified the shift of RPFs to the 5'-end by calculating a center-of-mass of RPFs per gene (calculated using only the CDS, excluding 15nt from the START codon to prevent differential occupancy of the START site alone). This quantity was also found to correlate well between the replicates (Pearsons R 0.57, Figure 4b). Importantly, we found that both the shift in TE and shift in center of mass correlate with the length of CDS (see Supplemental Figure S4a). We adjusted both metrics (change in TE and change in center of mass) for CDS length, and after this correction the correlation between the two metrics was still apparent (Supplemental Figure S4b). These data indicate that genes that exhibit little shift of RPFs towards the 5'-end (i.e. Δcenter-of-mass near 0) show a tendency to decrease in translational efficiency more strongly than those that have a large shift of RPFs towards the 5'-end. We further investigated whether certain functional gene groups show a differential change in RPF center of mass. Using the CDS length-controlled center-of-mass metric we found GO terms relating to translation and ribosomal subunits to be associated with reduced shift in RPF density. For example the term "structural constituent of ribosome" (GO:0003735) shows a markedly reduced shift of center of mass, compared to the rest of the genes (Figure 4c). Interestingly, genes in this group showed a greater TE in nocodazole treated samples compared to the rest of the genes, even when changes in TE

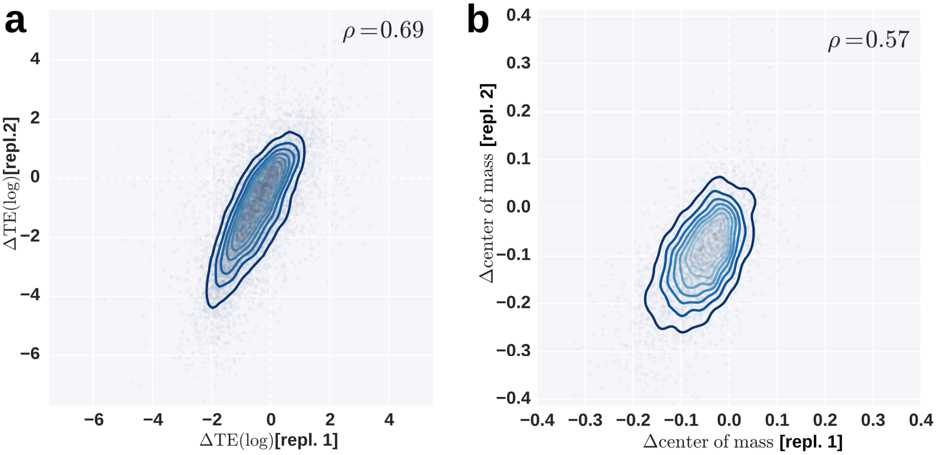


Figure 4: Analysis of gene-wise RPF positioning and translational efficiency reveals translational regulation of ribosomal protein genes and genes involved in the mitotic cell cycle.
(a) Correlation between changes in translational efficiencies (TEs) in two independent sets of control and nocodazole-treated sample pairs. **(b)** Correlation between changes in RPF center of mass in two independent sets of control and nocodazole-treated sample pairs.

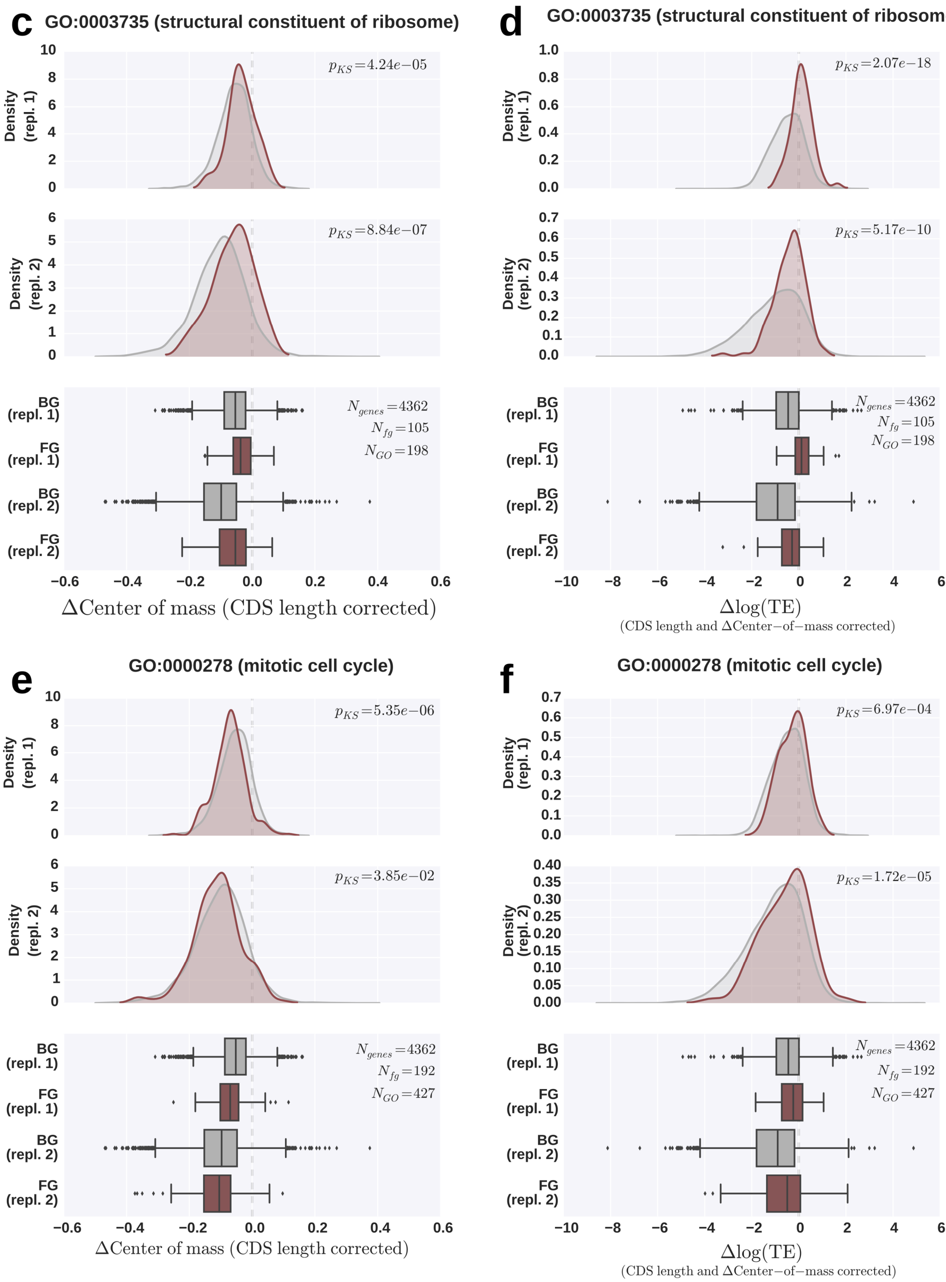


Figure 4 (cont.) (c) Change in center of mass of genes part of GO term “structural constituent of ribosome” (in red) and other genes (in gray) for two independent biological replicates. **(d)** Change in TE of genes part of GO term “structural constituent of ribosome” (in red) and other genes (in gray) for two independent biological replicates. **(e)** Change in center of mass of genes part of GO term “mitotic cell cycle” (in red) and other genes (in gray) for two independent biological replicates. **(f)** Change in TE of genes part of GO term “mitotic cell cycle” (in red) and other genes (in gray) for two independent biological replicates.

were controlled for CDS length and changes in center of mass (Figure 4d). Conversely, we found genes associated with the term “mitotic cell cycle” (GO:0000278) to be associated with an increased shift in center of mass (Figure 4e), while also showing a greater TE in nocodazole treated samples (Figure 4f). Finally, we found that the aspartic acid content of CDSs of genes had a positive effect on the ΔTE upon nocodazole treatment, however observed correlation was weak and the effect size was minimal (Supplemental Figure S4c). Furthermore it did not appear to be specific to the aspartic acid codons (GAC and GAT) as also glutamic acid (codons GAA and GAG) showed a negative correlation between codon content of the CDS and ΔTE, while RPF density analyses did not show any differential ribosomal occupancy at these codons.

Discussion

Translational processes are controlled during mitosis, most notably elongation as indicated by previous work and indicated by our results. Ribosomal profiling displays this effect as a shift of ribosomal protected fragments towards the 5'-end of transcripts. It is likely the result of reduced rates of elongation compared to the rate of translational initiation. Additionally, during mitosis ribosomes display an altered pattern of stalling on certain codons. We show (amongst other codons) that ribosomes shift aspartic acid codons from A-position to P-position. This phenomenon is reproducible and independent of the microtubule-disrupting effects of nocodazole, as it is reproduced with non-cytoskeletal treatments as well. Note that a stalling with aspartic acid codons in the A-site seems predisposed in the control cells, indicating a limitation at the step of charged tRNA-Asp interaction with the ribosome, which is abrogated in mitotic cells. We investigated whether differential expression levels of the tRNA themselves underlie this differential stalling pattern. To that end we profiled tRNA expression levels by two divergent extraction and sequencing methods. However, we found that differential expression levels of tRNAs were unable to explain the differences in stalling patterns observed in the ribosome profiling data. The cause for the A-to-P shift remains elusive, but it seems likely that either a differential availability of aspartic acid or a tRNA-specific regulation (e.g. modification) plays a role in this phenomenon.

Analysis of translational efficiency (TE) changes during mitosis illustrated the length of the CDS as an important confounding variable. The influence of CDS length on the observed translational efficiency is likely due to the overall shift of ribosomes towards 5'ends of transcripts. It does however highlight the importance of identification of confounding variables, since functional analyses (such as analysis of enriched GO terms) will give hugely deviating results between uncorrected and corrected metrics (such as TE). After correction for CDS length and RPF center of mass shift we found few functional groups associated with significantly different efficiency. The strongest effects are seen at ribosomal protein genes. This group of genes is known to be regulated at the translational level and is sensitive to activation of mTOR, through alleviation of the inhibitory phosphorylation of 4E-BP (Ma and Blenis, 2009; Thoreen et al., 2012). The activation of the ribosomal protein mRNAs at the translational levels suggests that mTOR is activated during mitosis. A recent report suggests that during mitosis not mTOR but CDK1 is responsible for the alleviation of the inhibitory phosphorylation of 4E-BP1 (Shuda et al., 2015). Another functional group which we found to be translationally induced were genes of the term “mitotic cell cycle”. This suggests that translational regulation occurs, but to strengthen this observation it would be helpful if the factors conveying this translational regulation could be identified. It should also be noted that the effect size is small in comparison to the effect on ribosomal protein genes; however this could be a result from the relatively late timepoint at which the mitotic samples were taken (16h into mitosis). We think our work provides valuable insights in the translational levels of regulation that take place during mitosis, and we expect that our observations and data can act as a starting point and provide pointers to find novel regulators of the mitotic cell cycle.

Materials and methods

Cell culture and enrichment of mitotic cells

U2OS cells were cultured in DMEM supplemented with 10% heat-inactivated fetal calf serum in 5% CO_2 at 37°C. Enrichment of mitotic cells by cell cycle arrest was achieved by treatment of cells with nocodazole (50 ng/ml), BI2536 (100 nM), S-Trityl-L-cysteine (STLC, 20 µM) or taxol (1 µM) for 18 hours.

Preparation of deep-sequencing libraries

Libraries from cultured cells were prepared as described previously (Loayza-Puch, 2013). Lysates were centrifuged at 5,000 rpm and the supernatant was digested with 2 U/µl of RNase I (Life Technologies, Grand Island, NY, USA) for 45 min at room temperature. Resulting monosomes were purified, RNA was isolated, and RP libraries were prepared as described previously. Primers and linkers used in the preparation of libraries are listed in Chapter 3, Supplemental Table S1.

Preparation of tRNA sequencing libraries

For total small RNA-seq libraries, the TruSeq Small RNA Library Preparation Kit (Illumina) was used according to the manufacturer's instructions. For the targeted sequencing of tRNAs, total RNA was gel-purified on a denaturing 10% polyacrylamide urea (7M) gel. A section corresponding to the range of 70-100 nucleotides was excised, eluted and ethanol precipitated. RNA was 3′-dephosphorylated using T4 polynucleotide kinase (New England Biolabs Inc. Beverly, MA, USA) for 6 h at 37°C in 2-(N-morpholino)ethanesulfonic acid (MES) buffer (100 mM MES-NaOH, pH 5.5, 10 mM $MgCl_2$, 10 mM β-mercaptoethanol, 300 mM NaCl). 3′ adaptors were added with T4 RNA ligase 1 (New England Biolabs Inc. Beverly, MA, USA) for 2.5 h at 37°C. Ligation products were 5′-phosphorylated with T4 polynucleotide kinase for 30 min at 37°C. 5′ adaptors were added with T4 RNA ligase 1 for 2.5 hrs at 37°C.

Preprocessing and alignment of deep-sequencing data

Adapter sequences were trimmed from raw RP data using cutadapt (Martin, 2011). Sequences shorter than 20bp after adapter trimming were discarded. rRNA and tRNA sequences were filtered by alignment to indices of rRNA and tRNA sequences respectively, using bowtie2 (Langmead and Salzberg, 2012) and default parameters. The rRNA index was constructed from GENCODE v19 annotations, transcript types "rRNA", "Mt_rRNA" and "rRNA_pseudogene", supplemented with UCSC repeats of class "rRNA" (table "rmsk"). The tRNA index was constructed from sequences obtained from GtRNAdb (Chan and Lowe, 2009) at 5th of September 2012. The remaining (unmapped) reads were aligned to GRCh37/hg19 using TopHat2 (Kim et al., 2013) and GENCODE v19/BASIC (Harrow et al., 2012) transcript coordinates (parameters “-N 2 -m 1 --no-novel-juncs --no-novel-index --no-coverage-search --segment-length 25”). In subsequent analyses, only primary alignments with mapping quality of 10 or greater were considered. RP and RNAseq data followed the same pipeline, with the exception of adapter trimming, which was only performed on RP data.

Codon occupancy analysis

Generally, codon occupancy analysis compares RPF codon occupancy frequencies between samples, in a gene-level normalized manner (to exclude gene expression differences as cause for differences in observed codon frequencies).

Specifically, gene IDs and reading frames were assigned to RPFs, using GENCODE v19/BASIC. RPFs not within a valid CDS (taking into account the 15nt. 5'-overhang of RPFs), RPFs with ambiguous gene ID and RPFs with ambiguous frame were excluded. The remaining RPFs were used to count, for the different positions (12 and 15nt. from 5'-end), the frequencies of all codons, for all genes in the transcriptome. Within-gene frequencies were calculated by dividing the observed counts by the total counts for that gene. Normalized codon frequencies were averaged over all genes that had at least *n* counts in both the condition and control samples (*n* is 100 for all figures). Codon occupancy shifts were calculated from these normalized and averaged frequencies per codon as: log(normalized occupancy in condition / normalized occupancy in control).

RPF density analysis

Codon-regions of 61 nucleotides in width around designated codons along the transcriptome (as annotated by GENCODE v19/BASIC) were identified (using transcript coordinates; i.e. codon-regions were all supported by exons). Overlapping transcript annotations were used, and overlapping regions were retained, but regions with identical genomic coordinates were considered redundant and collapsed. Codon-regions that could not be extended to 61 nt. (due to being near the 5' or 3' ends of transcripts) were discarded.

5' ends of RPFs were counted within each codon-region. For a comparison between two samples, only codon-regions where the RPF count was at least 50 in both samples, were taken into consideration. For each sample and each codon-region, the normalized 5' end RPF density was calculated by dividing over the total number of counts within that region, and multiplying by the width of the codon-region, making the average density within each codon region equal to 1. The normalized densities were convolved using a rectangular window of width 3 and height 1/3. The mean density over the codon-regions was taken and the difference in the mean densities between the samples yielded to the density shift.

Translational efficiency and center of mass analyses

For calculation of translational efficiencies, gene-wise counts were performed using annotations from GENCODE v19/BASIC and HTSeq, on CDSs only, for both the RNAseq and RP data. Counts were normalized by correction for the total library sizes, using the trimmed mean of M-values routine found in edgeR (Robinson et al., 2010), and nucleotide composition biases in the RNAseq and RP data were corrected prior to calculating translational efficiencies by linear regression of abundances using mono- and dinucleotide compositions of CDSs of genes as regressors. Translational efficiencies were calculated as the ratio of [normalized abundance determined by RP] to [normalized abundance determined by RNAseq]. Centers of mass were calculated per gene by taking the 5'-ends of RPFs at the CDS (excluding the first 15 nt.) and the following formula:

$$\text{Center of mass} = \frac{1}{NL}\sum_{i=1}^{n} d_i$$

where d_i is the position of the 5'-end of the i-th RPF minus 15, N is the total number of RPFs on the CDS and L is the length of the CDS (minus 15 nt.). For genes that had more than one single valid CDS (a CDS annotated by CDS, start_codon and stop_codon features, where the sequence of the first codon is a valid start codon and the stop codon is a valid stop codon), the longest CDS was chosen, and when tied, the transcript with the least number of exons was chosen as representative. For the calculation of TEs, a cutoff of at least 25 reads per gene was applied to both the RP and RNAseq data. Similarly, for the calculation of RPF centers of mass, a cutoff of at least 25 RPFs per gene was applied. For all analyses, cutoffs were applied across all samples/conditions that were compared, thus ensuring that the same set of genes was compared.

Differential translational efficiency and center of mass analysis of gene sets

The gene ontology flat file was obtained from geneontology.org (dd. 15 May 2014).

GENCODE/Ensembl gene ID mapping to GO ID was obtained from BioMart. Gene annotations were extended upwards in the gene ontology hierarchy, i.e. genes annotated with an ontology child term were considered also annotated with any parent term ("is_a" was considered the parent-child defining relation in the gene ontology flat file). For the different metrics (ΔTE, Δcenter of mass) statistical significance was tested by considering the values for genes annotated with a GO term vs. those not annotated with that term, using both the Kolmogorov-Smirnov and Mann-Whitney-U non-parametric tests. We adjusted the two-sided p-values for multiple testing using the Benjamini-Hochberg correction (the default method in R's *p.adjust* routine). We tested all replicates individually, and considered only GO terms where both test yielded (multiple testing-corrected) p-values < 0.1 in all replicates.

Analysis of small RNA and tRNA sequencing libraries

Adapter sequences were trimmed from raw sequencing data using cutadapt (Martin, 2011), requiring an adapter overlap length of at least 6 nt. and an overall matching rate of 20%. Sequences shorter than 20bp after adapter trimming were discarded. Trimmed reads were aligned using bowtie2 (Langmead and Salzberg, 2012) with parameters '-D 20 -R 3 -L 20 -N 1 -i C,1 -k 50' to the human genome (build hg19). The reads were assigned features, based on tRNA, miRNA, rRNA, snRNA and snoRNA annotations from GENCODE/v19. Each read that aligned was allowed to yield multiple alignments. Of these multiple alignments, only those with the highest score (sorted by BAM alignment score and BAM 'NM' tag) were considered. All those high-scoring alignments were considered in order to assign a feature to the read. For the paired-end smallRNA sequencing data, alignments for reads from both ends were jointly considered, and the alignments of the ends were not allowed to be further than 300bp. apart (and were not allowed to be dovetailed). For the remaining valid high-quality alignments, annotated features were considered, and reads with ambiguous features (i.e. with alignments indicating different features, such as miRNA vs. rRNA or different tRNA genes) were considered ambiguous reads and not considered in further analyses.

References

Chan, P.P., and Lowe, T.M. (2009). GtRNAdb: a database of transfer RNA genes detected in genomic sequence. Nucleic Acids Res. *37*, D93–D97.

Gonzalez, C., Sims, J.S., Hornstein, N., Mela, A., Garcia, F., Lei, L., Gass, D.A., Amendolara, B., Bruce, J.N., Canoll, P., et al. (2014). Ribosome profiling reveals a cell-type-specific translational landscape in brain tumors. J. Neurosci. *34*, 10924–10936.

Harrow, J., Frankish, A., Gonzalez, J.M., Tapanari, E., Diekhans, M., Kokocinski, F., Aken, B.L., Barrell, D., Zadissa, A., Searle, S., et al. (2012). GENCODE: the reference human genome annotation for The ENCODE Project. Genome Res. *22*, 1760–1774.

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

Ingolia, N.T., Lareau, L.F., and Weissman, J.S. (2011). Ribosome Profiling of Mouse Embryonic Stem Cells Reveals the Complexity and Dynamics of Mammalian Proteomes. Cell *147*, 789–802.

Ingolia, N.T., Brar, G.A., Rouskin, S., McGeachy, A.M., and Weissman, J.S. (2012). The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. Nature Protocols *7*, 1534–1550.

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 Biology *14*, R36.

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

Loayza-Puch, F. (2013). p53 induces transcriptional and translational programs to suppress cell proliferation and growth. Genome Biol. *14*, R32.

Ma, X.M., and Blenis, J. (2009). Molecular mechanisms of mTOR-mediated translational control. Nat Rev Mol Cell Biol *10*, 307–318.

Martin, M. (2011). Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet.journal *17*, pp. 10–12.

O'Connor, P.B.F., Li, G.-W., Weissman, J.S., Atkins, J.F., and Baranov, P.V. (2013). rRNA:mRNA pairing alters the length and the symmetry of mRNA-protected fragments in ribosome profiling experiments. Bioinformatics *29*, 1488–1491.

Peters, J.-M. (2006). The anaphase promoting complex/cyclosome: a machine designed to destroy. Nat Rev Mol Cell Biol *7*, 644–656.

Pyronnet, S., Dostie, J., and Sonenberg, N. (2001). Suppression of cap-dependent translation in mitosis. Genes Dev. *15*, 2083–2093.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Rooijers, K., Loayza-Puch, F., Nijtmans, L.G., and Agami, R. (2013). Ribosome profiling reveals features of normal and disease-associated mitochondrial translation. Nat Commun *4*, 2886.

Shalgi, R. (2013). Widespread regulation of translation by elongation pausing in heat shock. Mol. Cell *49*, 439–452.

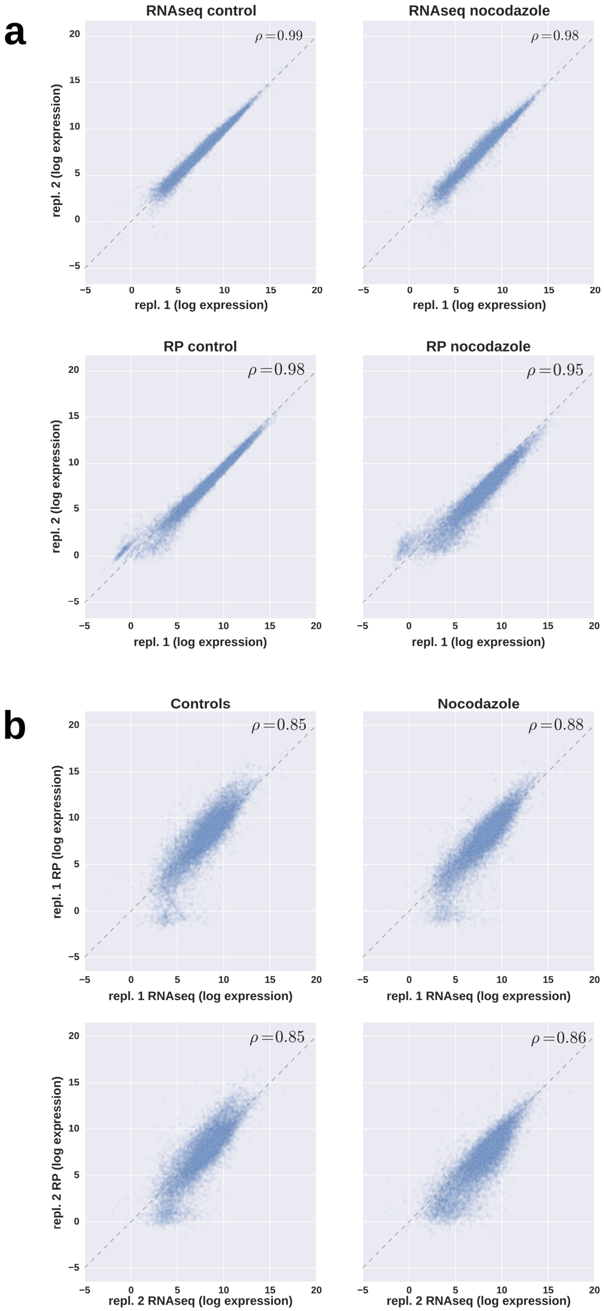
Shuda, M., Velásquez, C., Cheng, E., Cordek, D.G., Kwun, H.J., Chang, Y., and Moore, P.S. (2015). CDK1 substitutes for mTOR kinase to activate mitotic cap-dependent protein translation. PNAS *112*, 5875–5882.

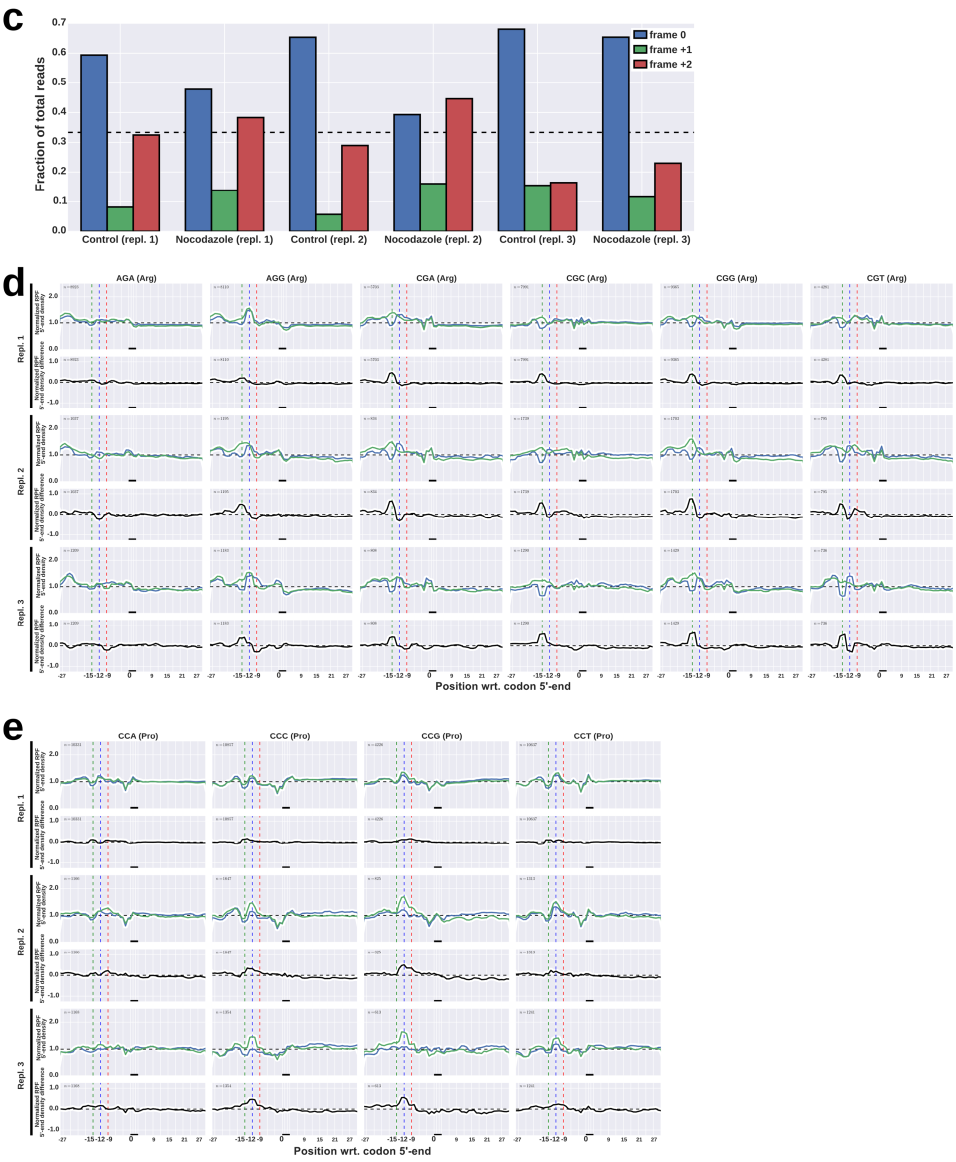
Sivan, G., Kedersha, N., and Elroy-Stein, O. (2007). Ribosomal Slowdown Mediates Translational Arrest during Cellular Division. Mol. Cell. Biol. *27*, 6639–6646.

Stumpf, C.R., Moreno, M.V., Olshen, A.B., Taylor, B.S., and Ruggero, D. (2013). The Translational Landscape of the Mammalian Cell Cycle. Molecular Cell *52*, 574–582.

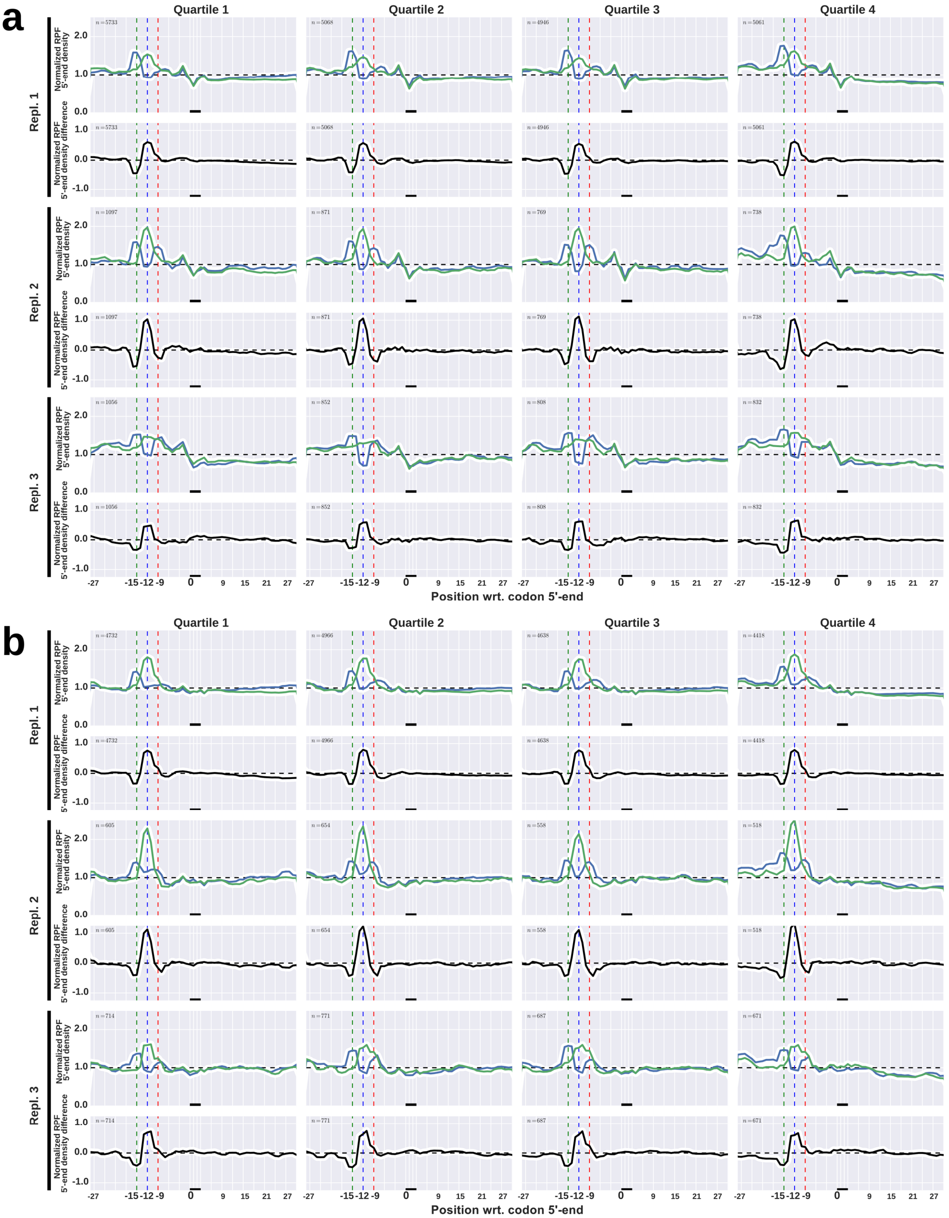
Thoreen, C.C., Chantranupong, L., Keys, H.R., Wang, T., Gray, N.S., and Sabatini, D.M. (2012). A unifying model for mTORC1-mediated regulation of mRNA translation. Nature *485*, 109–113.

Supplemental figures

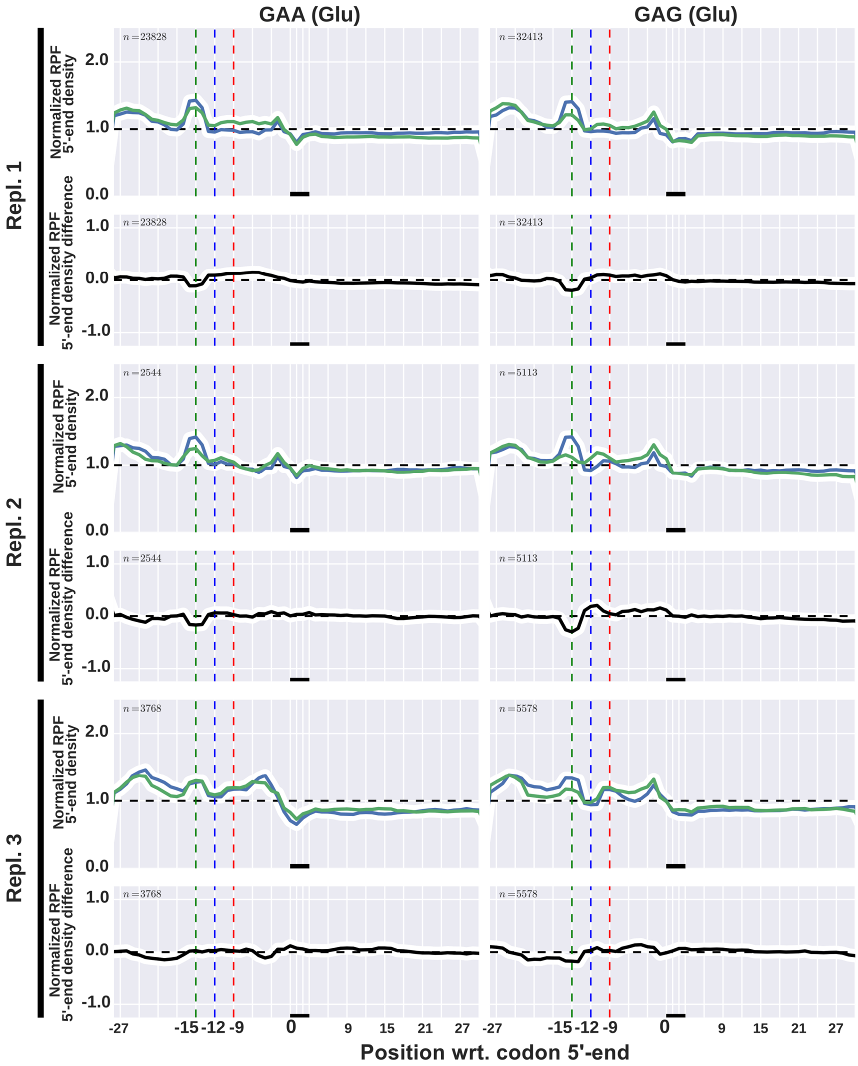




Supplemental Figure S1: Quality and reproducibility of sequencing data.
(a) Correlation between 2 independent biological replicates of RNA sequencing (top panels) and ribosome profiling (bottom panels) in control conditions (left panels) and nocodazole-treated conditions (right panels). **(b)** Correlation between ribosome profiling and RNA sequencing data in two independent biological replicates. **(c)** RPF 5'-end reading frame abundance in control and nocodazole-treated ribosome profiling samples for 3 independent biological replicates. **(d)** RPF 5'-end density profiles for arginine codons as in Figure 1g. **(e)** RPF 5'-end density profiles for proline codons as in Figure 1g.

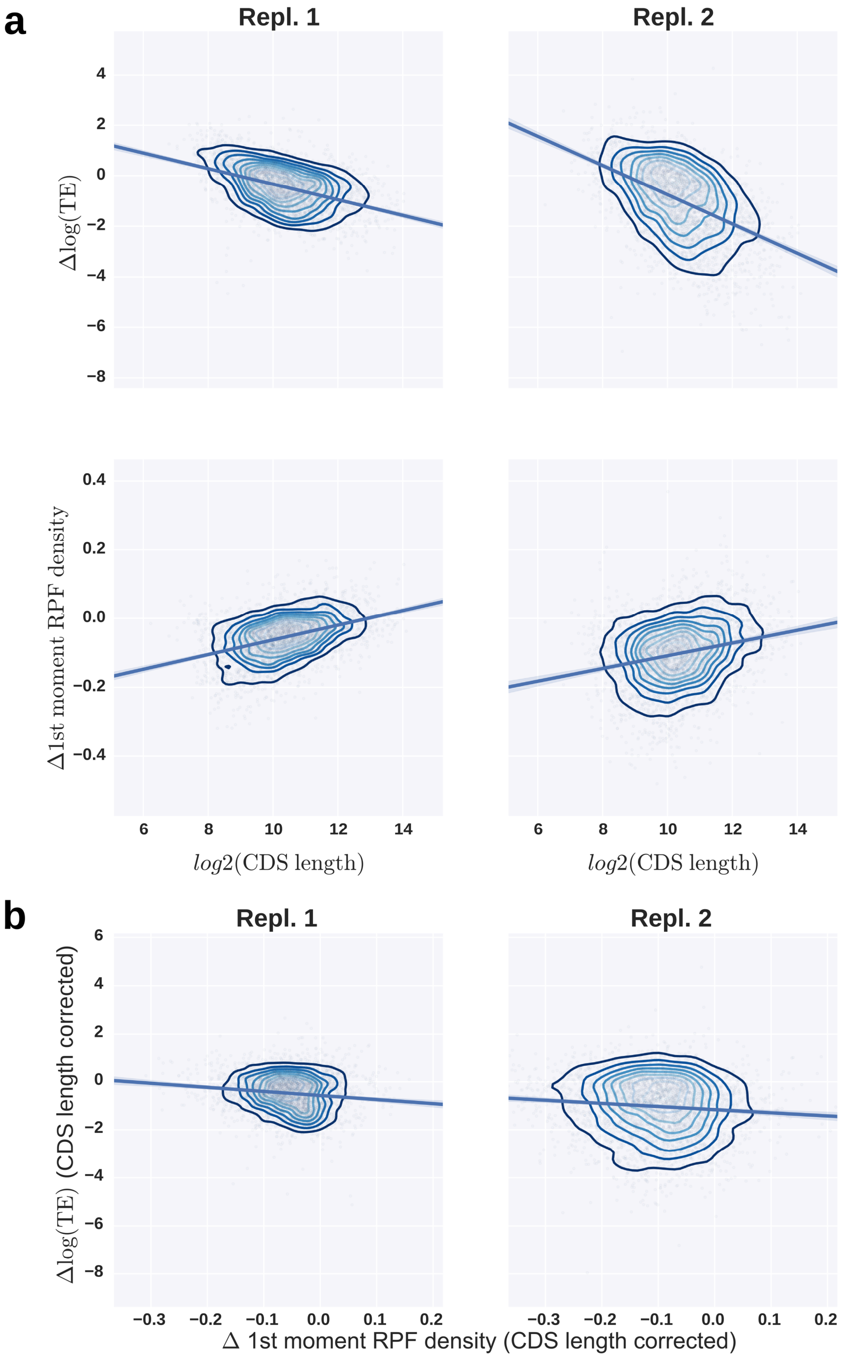


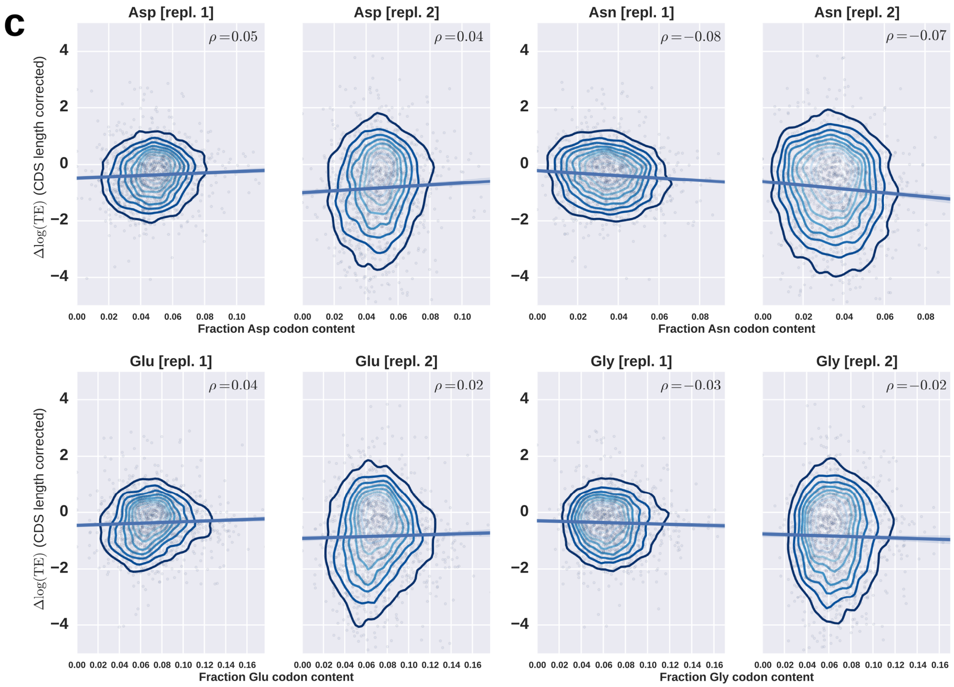
Supplemental Figure S2: RPF density at codons positioned along the CDS.
(a) RPF 5'-end density profiles for aspartic acid codon GAC that are positioned in the four quartiles along the CDSs of transcripts. Blue lines display the density in control samples, green lines display the density in nocodazole-treated samples, black lines display the difference between nocodazole and control samples. **(b)** RPF 5'-end density profiles as in panel (a), but for aspartic acid codon GAT.



Supplemental Figure S3: RPF density at codons coded by tRNAs with altered abundance in mitosis.

RPF 5'-end density profiles at the two glutamic acid (Glu) codons, GAA and GAG, in 3 independent biological replicates. Blue lines display the density in control samples, green lines display the density in nocodazole-treated samples, black lines display the difference between nocodazole and control samples.


a
Repl. 1
Repl. 2
4
2
0
−2
−4
−6
−8
Δlog(TE)
0.4
0.2
0.0
−0.2
−0.4
Δ1st moment RPF density
6
8
10
12
14
$log2$(CDS length)
$log2$(CDS length)
b
Repl. 1
Repl. 2
6
4
2
0
−2
−4
−6
−8
Δlog(TE) (CDS length corrected)
−0.3
−0.2
−0.1
0.0
0.1
0.2
Δ 1st moment RPF density (CDS length corrected)



Supplemental Figure S4: Correlation between translation efficiency changes, changes in RPF center of mass and CDS length.

(a) Correlation between translational efficiency changes upon nocodazole treatment and CDS length (top panels) and center of mass changes and CDS length (bottom panels) **(b)** Correlation between translational efficiency changes and RPF center of mass changes after adjusting both metrics for CDS length. **(c)** Correlations between translational efficiency changes and amino acid/codon content of genes for amino acids aspartic acid, asparagine, glutamic acid and glycine, in two biological replicates. Pearson correlation coefficients are indicated by ρ. In all panels, dark blue straight lines indicate linear regressions and blue shading around the linear regressions indicate a 99% confidence interval of the linear regression.

CHAPTER 5

Discussion of Part A – Protein translation studied by ribosome profiling

The chapters of this part of the thesis describe an exploration of translational processes using ribosome profiling. We tailored the protocol and analyses to investigate direct or indirect changes in availability of aminoacylated tRNAs, the essential component of the translational machinery that delivers amino acids. We expect that our work will provide tools that can identify novel molecular targets and biomarkers as well as uncover factors that regulate translation.

Translation by mitochondrial ribosomes

In chapter 2 we uncovered that the conventional ribosome profiling protocol was unable to faithfully capture translation by mitochondrial ribosomes. We adapted the ribosome profiling protocol and were able to capture translation by mitochondrial ribosomes to the same extent as cytosolic ribosomes. Our adapted protocol provided a useful new tool for mitochondrial research, as it allowed us to investigate the effects of a genetic disorder, caused by a mutation in the mitochondrial tRNA-Trp gene, on translation. This work also highlights the importance of distinguishing between true phenomena and technical artefacts, in this case the lower apparent translational efficiencies of mitochondrial genes. While it seems fully plausible that the set of mitochondrial genes yield mRNA transcripts that have a distinct translational efficiency, or that mitochondrial ribosomes have a lower overall affinity for mRNA than cytosolic ribosomes, the low apparent translational efficiency measured with the conventional ribosome protocol was merely due to technical artefacts. We showed that stringent size selection of ribosome protected fragments was the culprit in capturing mitochondrial ribosomes, which are distinctly larger than their cytoplasmic counterparts, and this in turn lowered the apparent translational efficiencies of mitochondrial mRNAs. Relieving the stringent size selection allowed us to capture ribosome protected fragments from both cytosolic and mitochondrial ribosomes in parallel. The data we generated using this adapted protocol indicated that mitochondrial ribosomes have affinity for mitochondrial mRNAs to a similar extent as the cytosolic ribosomes for nuclear-encoded mRNAs. The established protocol included the stringent size selection in order to reduce ribosomal RNA (rRNA) contamination (Ingolia et al., 2009, 2012). We initially feared that adaptation of the size selection step would lead to an increased amount of contaminating reads. However, the fraction of rRNA contamination was virtually unaffected by widening of the size selection window (Chapter 2, Figure 2d).

rRNA contamination

The ribosome profiling protocol involves RNA digestion in order to identify the ribosome protected mRNA fragments. However, the ribosomes themselves are complexes of protein and ribosomal RNA. The latter is inherently prone to digestion during the ribosome profiling protocol, and as a result, many fragments of ribosomal RNA end up contaminating the library which ideally contains only ribosome protected mRNA fragments. This contamination can severely hamper sensitivity of the obtained data, as the fraction of contaminating reads exceed 95% of the raw data, thereby reducing the amount of mRNA fragments in the data. In order to reduce rRNA contamination from the data, we made use of the observation that specific fragments of the ribosomal RNAs are far more prevalent than others, a consequence of their size, propensity to be digested and location wrt. the surface of the ribosome. Using hybridization to biotinylated oligonucleotides we were able to remove highly abundant target sequences through streptavidin affinity. Using this approach we were able to reduce the fraction of contaminating reads to 60% in most data sets.

Bimodality of size of mitochondrial-ribosome protected fragments

Our adapted ribosome profiling protocol increased the window of RPF size selection, and allowed for improved detection of mitochondrial RPFs, that had previously fallen outside of this size selection window. Surprisingly we noticed that the size of mitochondrial RPFs was not unimodal like the size of cytosolic RPFs, which peaked at 29nt., but instead showed a bimodal size distribution with peaks at 27 and 33nt. We could not identify any functional relevance or consequence of the size of the RPFs, and their distributions along the mitochondrial mRNAs was similar. Concurrently, work from colleagues investigating translation in bacteria showed a bimodal size distribution of RPFs, which was dependent on the location of a signal sequence (the Shine-Dalgarno sequence) with respect to the RPF (O’Connor et al., 2013). We investigated whether the bimodal size distribution of mitochondrial RPFs in our data could be explained by signal sequences but we were unable to identify such an element. Important to notice is that analysis of overrepresented elements was hampered by the small size of the mitochondrial protein-coding transcriptome (i.e. the cumulative length of mitochondrial mRNAs), reducing the statistical power to find significantly overrepresented sequence motifs. Thus, the existence of sequence motifs that underlie the bimodal size distribution of mitochondrial RPFs cannot be excluded.

Investigating the consequences of mitochondrial tRNA mutations in genetic disorders

We used the adapted ribosome profiling protocol to provide a view of translation of mitochondrial mRNAs in cells harboring mitochondrial tRNA mutations that occur in patients with known genetic disorders. We used a *cybrid cell system*, where mitochondria harboring the mutations were fused with cells lacking mitochondria, allowing us to investigate the effects of the mutations free of variability in the nuclear and cytosolic content. We found that a mitochondrial tRNA-Trp$^{5556G>A}$ mutation causes ribosome stalling on tryptophan codons, compared to wild-type mitochondria. We show that the mutation leads to lower production rates of mitochondrial protein (chapter 2) and this is in line with the phenotype of patients with this genetic disorder (Smits, 2010). Since the cells harboring this mutation are viable, and ribosomes are found throughout the CDSs of mitochondrial mRNAs, it is likely that protein production occurs albeit at a significantly lower rate. This implies that the tRNA-Trp$^{5556G>A}$ mutation does not fully impair the function of the tRNA. Instead, its availability for translation and uptake by ribosomes is reduced. This might be caused by less efficient folding of the tRNA into its three-dimensional structure, or reduced efficiency for aminoacylation or chemical modifications that are part of tRNA maturation. We also investigated the consequences of several other genetic disorders where mutations in mitochondrial tRNAs occur, that lead to a similar disease manifestation and similar reduction in mitochondrial protein production rates to what is observed with the tRNA-Trp$^{5556G>A}$ mutation. However, with other mutations that we profiled (in the mitochondrial tRNA genes for tRNA-His and tRNA-Arg) we found that ribosome protected fragments were distributed indistinguishable from the ribosomes in the control cells and no stalling of ribosomes when encountering the complementary codon was apparent (see Figure 1, Rooijers, Loayza-Puch and Agami, unpublished data). In these mutations, which we validated by deep-sequencing, a reduced availability of the tRNA for translation does not seem to be causal for the disease phenotype. Instead, the disease might be a consequence of toxic effects of the tRNA mutations. Alternatively, the tRNA mutations that elicit no effect on ribosome progression are passenger mutations, and the causal mutations have been missed in analysis of the patient material.

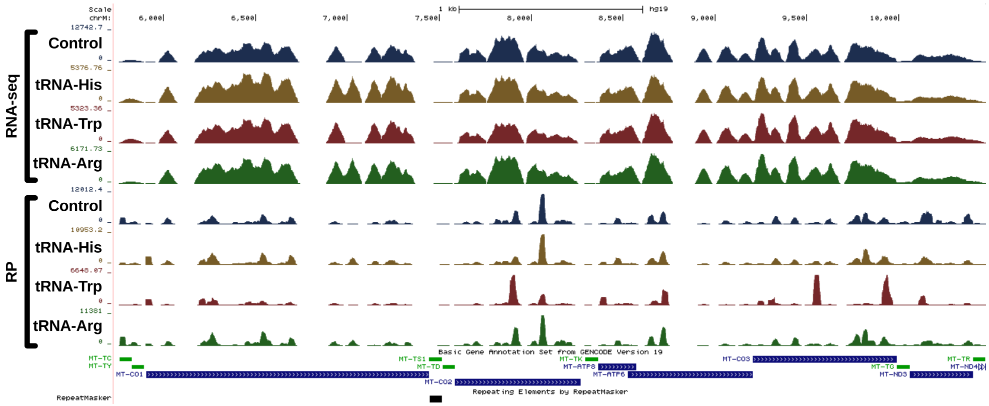


Figure 1: RNA sequencing and ribosome profiling performed on control and three mitochondrial tRNA mutant cybrid celllines.
UCSC data session showing RNAseq on one control (blue) and three mitochondrial tRNA mutant cybrid celllines (brown: tRNA-His, red: tRNA-Trp and green: tRNA-Arg) in the first four tracks and ribosome profiling in the next four tracks (identical coloring scheme). While transcript coverage measured by RNAseq is generally comparable between control and mutant celllines, the ribosome profiling data shows that the tRNA-Trp mutant exhibits a strongly deviating pattern, indicative of ribosomal stalling.

Ribosome profiling as sensor for differential amino acid availability

In chapter 3 the use of ribosome profiling as tool to expose differential amino acid availabilities is explored. Our work with mitochondrial tRNA mutants underscores that ribosome profiling is able to expose differential availability of tRNA, as a mutation in the mitochondrial tRNA-Trp gene leads to ribosomes stalling on complementary (Trp) codons. We reasoned that tRNA availability could not only be modulated by mutations that possibly alter stability or folding, but also by differential availability of amino acids, or differential activity of aminoacyl tRNA synthetases, which catalyze the esterification of amino acids to their appropriate tRNA (Figure 2).

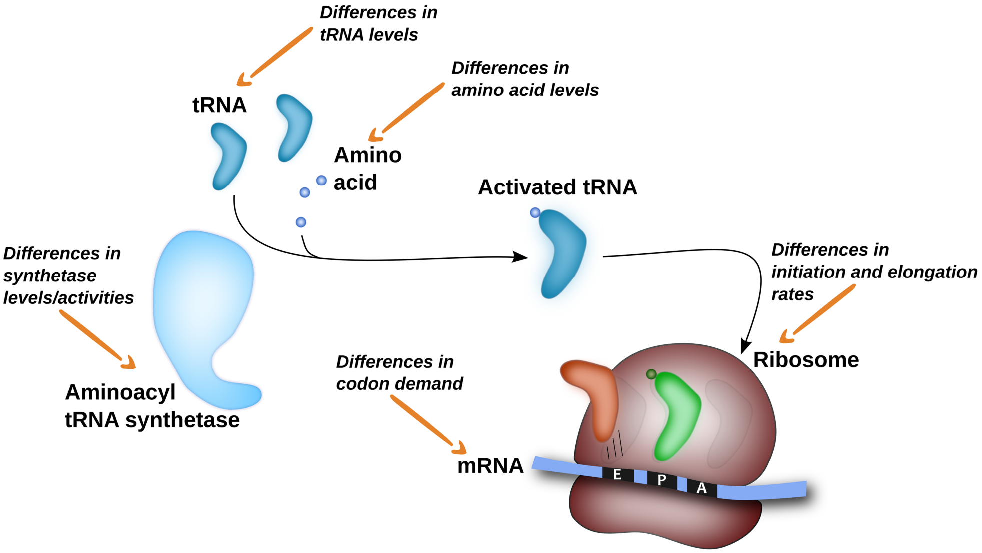

Differences in tRNA levels
tRNA
Differences in amino acid levels
Amino acid
Activated tRNA
Differences in synthetase levels/activities
Differences in initiation and elongation rates
Ribosome
Aminoacyl tRNA synthetase
Differences in codon demand
mRNA
E
P
A

Figure 2: Upstream factors involved in differential ribosomal codon occupancies. Factors that can alter the availability of activated tRNAs for translation and thereby alter the codon occupancies of ribosomes are shown. Note that this scheme focuses factors discussed in this thesis, and that other factors (such as mRNA folding or factors acting upon translation elongation) are omitted.

A tool that offers insight in amino acid availability would be valuable because it could pinpoint specific vulnerabilities of cancers. Furthermore, using ribosome profiling as a sensor for amino acid availability would be more relevant than measurements of amino acid levels on cell lysate, because ribosome profiling would indicate any differential availability that would result from metabolic flux rewiring, independent of the amino acid concentration. Measurements on the translational level bypass the need to investigate the regulatory levels that appear between amino acid and incorporation into protein, as it can directly indicate the incorporation (or lack thereof) of amino acids.

Translation initiation is tightly controlled upon restriction of nutrient availability

To validate ribosome profiling as a tool capable of detecting differential amino acid availability we starved cells of specific amino acids, and measured a response of ribosome stalling using ribosome profiling. For instance, Figure 1d in Chapter 3 displays the response of ribosomes upon treatment with L-asparaginase, which reduces the amount of available asparagine. Another experimental setup, where a panel of breast cancer celllines was starved of glutamine, taught us that not all cells respond in equal ways. We show that luminal celllines exhibit ribosome stalling specifically on glutamine codons, while the basal type celllines show either no or a marginal response to glutamine starvation. These cells instead exhibit a reduced occupancy of ribosomes on the initiation (START) codon, indicative of a global reduction of translation initiation. Analysis of phosphorylation of eEF2, a key regulator of this process, confirmed that the luminal celllines have a response distinct from the basal celllines. Altogether these results reveal a limitation in the approach: regulation of translation initiation (and possibly elongation) can mask amino acid-specific signals in the ribosome profiling data. Conversely, it shows that ribosome profiling can detect global regulation of translation, from which a cell's commitment to proliferate can be inferred.

Revealing reduced leucine availability upon TGF-beta treatment, and reduced proline availability in kidney tumor

When treated with TGF-beta, a signaling molecule capable of triggering a wide range of responses, MCF10A cells undergo epithelial-to-mesenchymal transition. Using ribosome profiling and our analysis protocol, we show that this is accompanied with stalling on all leucine codons, indicative of reduced leucine availability. We found a leucine transporter gene (SLC3A2) to be transcriptionally attenuated and showed its importance in the uptake of leucine. Furthermore we show that attenuation of this gene is at least partially responsible for the stalling of ribosomes on leucine codons, and that supplementing the growth medium with an esterified leucine, which is cell permeable, reduces the response.

To take our approach to a clinically relevant setting, we performed ribosome profiling on tissue samples taken from dissected kidney tumors. One tumor showed elevated transcript levels of PYCR1, the prime gene responsible for *de novo* synthesis of the amino acid proline. However, our ribosome profiling data indicated that contrary to what might be inferred from the transcriptional activation of PYCR1, these cells showed a ribosomal stalling on proline codons. This highlights the power of the ribosome profiling approach to amino acid availability, as it shows the net effect of amino acid concentration, metabolic fluxes and translational regulation.

Ribosome profiling reveals a distinct redistribution of ribosomes in mitosis

The final chapter of this part focuses on translational regulation and ribosome redistribution during mitosis. Mitosis is an intriguing part of the cell cycle in which the cell undergoes drastic morphological changes, usually followed by cytokinesis during which the cells' content is split amongst two daughter cells. Cap-dependent translation initiation has been reported to be reduced during mitosis (Fan and Penman, 1970; Pyronnet et al., 2001; Sivan et al., 2007) however recent reports suggest that cap-dependent translation is stimulated, through alternative, mTOR-independent pathways (Shuda et al., 2015). The regulation of translation during this phase has been of interest, and it was hypothesized that a stalling of ribosomes protects mRNAs from degradation (Sivan and Elroy-Stein, 2008), as well as the idea that several mitosis-specific factors would escape the translational repression. Our data shows a strong and reproducible global effect on ribosome positioning wrt. the CDS. A strong global shift of RPFs towards the translation START site indicates a reduced elongation activity. Interestingly, we see no evidence of reduced translation initiation. A remarkable observation is that ribosomes protect different codons preferentially in mitosis, and this differential codon occupancy seems independent of the shift of RPFs towards the translation START site. This suggests a differential tRNA availability, either directly through differential levels of the tRNAs themselves or other upstream factors such as amino acid availability or tRNA synthetase activity. We addressed the possibility that differential tRNA levels are causal to the differential codon occupancy by deep-sequencing of tRNA in control and mitotic cell samples. While we observed differences in the tRNA levels upon mitosis, they were unable to explain the differential codon occupancies.

Future perspectives

Our work has demonstrated the ability of ribosome profiling to identify several mechanisms of translational regulation. Global mechanisms such as repression of elongation, for instance during mitosis, or repression of initiation, which occurs in certain celllines upon nutrient starvation are readily detected using ribosome profiling and appropriate analyses. Also regulation of specific components such as tRNAs and amino acids can be detected by ribosome profiling. Finally, we show that ribosome profiling is also able to detect regulation of specific mRNAs, as demonstrated by the activation of translation of ribosomal protein mRNAs during mitosis. We have found that ribosome profiling can be used to identify differential regulation of translation, and aid in the identification of factors that mediate this differential regulation. Furthermore we show that ribosome profiling on tumor sections is feasible and that high-quality data can be obtained, and from this data regulation at the level of translation can be inferred. We expect that this in particular will allow the identification of metabolic vulnerabilities of tumors and the identification of novel molecular targets. In a follow-up study we are currently investigating how nutrient requirements of tumor cells differ between growth *in vitro* and growth *in vivo*, and amongst different tumor cells, in order to exploit these differences and to be able to put a halt to growth of the tumor cells specifically.

References

Fan, H., and Penman, S. (1970). Regulation of protein synthesis in mammalian cells. II. Inhibition of protein synthesis at the level of initiation during mitosis. J. Mol. Biol. *50*, 655–670.

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

Ingolia, N.T., Brar, G.A., Rouskin, S., McGeachy, A.M., and Weissman, J.S. (2012). The ribosome profiling strategy for monitoring translation in vivo by deep sequencing of ribosome-protected mRNA fragments. Nature Protocols *7*, 1534–1550.

O'Connor, P.B.F., Li, G.-W., Weissman, J.S., Atkins, J.F., and Baranov, P.V. (2013). rRNA:mRNA pairing alters the length and the symmetry of mRNA-protected fragments in ribosome profiling experiments. Bioinformatics *29*, 1488–1491.

Pyronnet, S., Dostie, J., and Sonenberg, N. (2001). Suppression of cap-dependent translation in mitosis. Genes Dev. *15*, 2083–2093.

Shuda, M., Velásquez, C., Cheng, E., Cordek, D.G., Kwun, H.J., Chang, Y., and Moore, P.S. (2015). CDK1 substitutes for mTOR kinase to activate mitotic cap-dependent protein translation. PNAS *112*, 5875–5882.

Sivan, G., and Elroy-Stein, O. (2008). Regulation of mRNA Translation during cellular division. Cell Cycle *7*, 741–744.

Sivan, G., Kedersha, N., and Elroy-Stein, O. (2007). Ribosomal Slowdown Mediates Translational Arrest during Cellular Division. Mol. Cell. Biol. *27*, 6639–6646.

Smits, P. (2010). Functional consequences of mitochondrial tRNATrp and tRNAArg mutations causing combined OXPHOS defects. Eur. J. Hum. Genet. *18*, 324–329.

Regulation of Coding and Non-coding Genes;
New insights obtained through analysis of high-throughput sequencing data

Koos Rooijers

PART B

Cover (Part B): “Just do a UCSC screenshot”, Koos Rooijers, 2015

The UCSC genome browser has been a ubiquitous tool in my work for the past 5 years. At first I was somewhat repulsed: the '90s look and feel (and page loading times[*]) did not align with my expectations of a tool used in cutting-edge science. However, over time I came to appreciate the UCSC browser and learned to deal with its quirks. The available data and its currency are unparalleled and make it an incredibly powerful tool. And I guess, over time, I even came to terms with using screenshots in publications...

---

[*] Did you know that for every view you take in the browser, a new figure is generated based on the position and tracks you are currently looking at, at some remote server, which is then transmitted to you? A modern solution would send your browser portions of track data and let your browser do the graphing.

Contents

Part B – Transcription of the non-coding genome

Chapter 6 Introduction to Part B B 5
Chapter 7 Genome-wide profiling of p53-regulated enhancer RNAs uncovers a subset of enhancers controlled by a lncRNA B 9
Chapter 8 Discussion of Part B B 47

Summary B 49
Samenvatting B 51
Curriculum Vitae B 53
List of publications B 54

Part A – Protein translation studied by ribosome profiling

Chapter 1 Introduction to Part A A 9
Chapter 2 Ribosome profiling reveals features of normal and disease-associated mitochondrial translation A 21
Chapter 3 Sensing amino acid deficiencies through differential ribosome codon occupancies A 43
Chapter 4 Specific changes of ribosome distribution over messenger-RNAs during mitotic arrest A 77
Chapter 5 Discussion of Part A A 103

CHAPTER 6

Introduction to Part B – The non-coding genome

Enhancers as functional component in the non-coding genome

Part A of this thesis centers around the translation of messenger RNA into protein. It hails proteins as important components, the major workhorses of the cell. The importance of proteins for the functioning of a cell has been long recognized, and hereditary content of a cell (its DNA) is often studied in the context of the impact it has on the expressed proteins that result from the genetic code. One example is the study of SNPs, variations in the genetic code that exist in a population. The impact of SNPs that occur in parts of the genome that encode for proteins is more readily discerned, and often considered more significant. It is paradoxical that in the human genome only 2% of the genomic code actually encodes for protein sequences, and suggestive of functional elements other than protein-coding genes that are kept in the genome. Since the energetic costs that are associated with replication and maintenance of the genome are significant, dysfunctional or "useless" parts of the genome would be lost during evolution. However, within the non-coding parts of the genome, many conserved regions can be identified. Indeed, other functional non-coding elements can be identified in the genome, for instance promoters, which regulate the expression of genes (DNA that is transcribed into protein-coding or non-protein-coding RNA). Another important type of functional non-coding element that regulates the expression of genes are enhancers. Their existence and function has been long known (Banerji et al., 1981; Benoist and Chambon, 1981; Khoury and Gruss, 1983).

Akin to promoter elements, they form docking sites for transcription factors and transcription regulators, which bind to the enhancer through sequence elements or chromatin environment, such as histone modifications. In contrast to promoters, they are positioned not necessarily close to the target gene whose expression they modulate (see Figure 1). Instead, they come in contact with the target gene through looping of the chromatin, and because of this they may be located millions of basepairs away from their target gene, or even be located at a different chromosome. This poses a difficulty in identification of the target enhancers of a particular gene (and vice versa, the target gene(s) of an enhancer). Many years of research, and more recently advances in high-throughput and deep-sequencing technologies, have allowed for the identification and functional impact of these enhancer elements within the genome (Shlyueva et al., 2014). This has led to estimates exceeding 300.000 functional enhancers in the human genome (Pennacchio et al., 2007), a number far exceeding the number of protein-coding genes in the human genome (approx. 22.000), resulting in part from the large degree of tissue-specificity of enhancers. Indeed, in different tissues and celltypes, even expression of ubiquitous genes

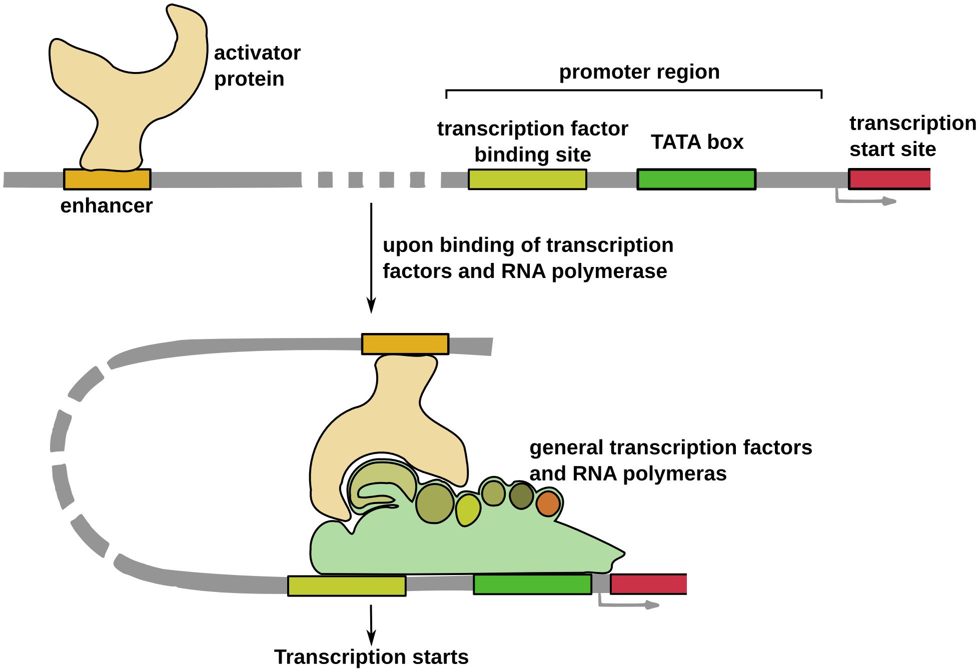

activator
protein
promoter region
transcription factor
binding site
TATA box
transcription
start site
enhancer
upon binding of transcription
factors and RNA polymerase
general transcription factors
and RNA polymeras
Transcription starts

Figure 1: Regulatory capacity of enhancers at distant promoters
A transcription start site carries a region upstream known as the promoter, commonly 0.5-2kbp. in length. In the promoter region the molecular machinery (with RNA polymerase as major component) that carries out transcription is recruited. This recruitment is often facilitated by other factors (transcription factors, not shown in the figure) that are in turn recruited by auxiliary sequences in the DNA itself (transcription factor binding sites). Enhancers are sequences that also promote the recruitment of transcription machinery (or activate the start of transcription by the machinery (Jonkers and Lis, 2015)) but in contrast to promoter-localized transcription factor binding sites, they can be located far away from the transcription start site DNA sequence (>>1000kbp.) and instead become physically located close to the transcription start site through DNA looping.

such as housekeeping genes may be driven by different enhancers (de Laat and Duboule, 2013).

Long non-coding RNAs as functional gene products

Promoter and enhancer elements act as docking platform for transcription factors and chromatin modifiers, which in turn promote transcription of target genes. Genes can be further subdivided into two groups: those genes that encode for protein and thus require translation and which have long been considered the main component of the genetic information, and genes that are transcribed but do not encode for protein, and thus the transcribed RNA is the end-product (non-coding RNAs). Some classes of non-coding RNAs have been long known: transfer-RNAs (tRNAs) and ribosomal RNAs (rRNAs) are important components of the translational machinery. Another class of non-coding RNAs are microRNAs (miRNAs), short transcripts which act to decrease expression of certain target transcripts through sequence complementarity. The discovery of miRNAs has in turn led to the discovery of an important level of regulation of gene expression. Another class of non-coding genes are those encoding for long non-coding RNAs (lncRNAs), a class commonly defined through their transcript size (> 200 nt.) which sets them apart from most other non-coding RNAs discovered thus far. As a loosely defined class, they can have a wide range of functions. For example lncRNA XIST functions as structural RNA to cover the inactive X-chromosome in female cells, and is an important regulator in the inactivation of the inactive X-chromosome (Brown et al., 1992). More specifically, lncRNAs may function by repression of transcription factors by acting as decoy (Kino et al., 2010), and through interaction with other RNAs, chromatin (DNA) and other proteins such as transcription factors, they can act at virtually any level of regulation that encompasses the cell (Vance and Ponting, 2014). The expression patterns of lncRNAs show high tissue-specificity, like the use of enhancers (Derrien et al., 2012; Tsoi et al., 2015). Furthermore, their expression is often lower than protein-coding genes, hampering their identification. As with enhancers, in the area of lncRNA research advances in technologies such as deep-sequencing have thus also allowed more sensitive detection of the functional and expressed elements, catalyzing discoveries in this area and discoveries of new functions of lncRNAs.

In this part of the thesis the non-coding targets of p53, a master regulator of cellular processes and cell fate, such as cell cycle progression and apoptosis are explored. As the tumor suppressor gene that is found most commonly mutated in cancers, it has been studied extensively. Functionally, the protein P53 acts as a transcription factor to regulate downstream genes. Through techniques such as ChIP the loci in the genome that are bound by P53 have been characterized, although advances in these techniques still improve the sensitivity which results in identification of more and more P53 binding sites. Also the impact of P53 on transcriptional regulation has been characterized in comprehensive fashion, by array transcriptome profiling and in recent years deep-sequencing of RNA. However, the impact on transcriptional regulation has mostly been studied with focus on the protein-coding genes and transcripts. We hypothesized that P53 also activates many non-coding genes that have remained undiscovered because (i) many of the P53 binding sites identified by ChIP can not be attributed to activation of protein-coding genes, (ii) a transcription factor cannot make the distinction between a protein-coding gene and a lncRNA gene and (iii) only recent advances have made detection of lowly abundant and novel transcripts such as lncRNAs and enhancer RNAs (eRNAs, products of transcription at enhancers and indicators of enhancer activity) possible, thus creating a potential for undiscovered P53 targets. We start off by using start-of-the-art technologies and available data to identify P53-regulated enhancers and P53-regulated lncRNAs. We then assessed whether the P53-regulated lncRNAs could play a role in the activation of the P53-activated enhancers that were found to be P53-unbound and identified lncRNA 'LED'. We show that LED is a potential tumor suppressor in the p53 pathway, as it can be inactivated by hypermethylation.

References

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.

Benoist, C., and Chambon, P. (1981). In vivo sequence requirements of the SV40 early promoter region. Nature *290*, 304–310.

Brown, C.J., Hendrich, B.D., Rupert, J.L., Lafrenière, R.G., Xing, Y., Lawrence, J., and Willard, H.F. (1992). The human XIST gene: analysis of a 17 kb inactive X-specific RNA that contains conserved repeats and is highly localized within the nucleus. Cell *71*, 527–542.

Derrien, T., Johnson, R., Bussotti, G., Tanzer, A., Djebali, S., Tilgner, H., Guernec, G., Martin, D., Merkel, A., Knowles, D.G., et al. (2012). The GENCODE v7 catalog of human long noncoding RNAs: analysis of their gene structure, evolution, and expression. Genome Res. *22*, 1775–1789.

Jonkers, I, and Lis, J (2015). Getting up to speed with transcription elongation by RNA polymerase II. Nat Rev Mol Cell Biol *16*, 167–177

Khoury, G., and Gruss, P. (1983). Enhancer elements. Cell *33*, 313–314.

Kino, T., Hurt, D.E., Ichijo, T., Nader, N., and Chrousos, G.P. (2010). Noncoding RNA gas5 is a growth arrest- and starvation-associated repressor of the glucocorticoid receptor. Sci Signal *3*, ra8.

de Laat, W., and Duboule, D. (2013). Topology of mammalian developmental enhancers and their regulatory landscapes. Nature *502*, 499–506.

Pennacchio, L.A., Loots, G.G., Nobrega, M.A., and Ovcharenko, I. (2007). Predicting tissue-specific enhancers in the human genome. Genome Res. *17*, 201–211.

Shlyueva, D., Stampfel, G., and Stark, A. (2014). Transcriptional enhancers: from properties to genome-wide predictions. Nat Rev Genet *15*, 272–286.

Tsoi, L.C., Iyer, M.K., Stuart, P.E., Swindell, W.R., Gudjonsson, J.E., Tejasvi, T., Sarkar, M.K., Li, B., Ding, J., Voorhees, J.J., et al. (2015). Analysis of long non-coding RNAs highlights tissue-specific expression patterns and epigenetic profiles in normal and psoriatic skin. Genome Biology *16*, 24.

Vance, K.W., and Ponting, C.P. (2014). Transcriptional regulatory functions of nuclear long noncoding RNAs. Trends in Genetics *30*, 348–355.

CHAPTER 7

Genome-wide profiling of p53-regulated enhancer RNAs uncovers a subset of enhancers controlled by a lncRNA

Nicolas Léveillé[1,*], Carlos A. Melo[1,2,*], Koos Rooijers[1,*], Angel Díaz-Lagares[3], Sonia A. Melo[4], Gozde Korkmaz[1], Rui Lopes[1], Farhad Akbari Moqadam[5], Ana R. Maia[6], Patrick J. Wijchers[7], Geert Geeven[7], Monique L. den Boer[5], Raghu Kalluri[4], Wouter de Laat[7], Manel Esteller[3,8,9] and Reuven Agami[1,10]

1 Division of Biological Stress Response, The Netherlands Cancer Institute
Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands
2 Doctoral Programme in Biomedicine and Experimental Biology,
Centre for Neuroscience and Cell Biology Coimbra University, 3004-504 Coimbra, Portugal
3 Division of Cancer Epigenetics, Cancer Epigenetics and Biology Program (PEBC)
Bellvitge Biomedical Research Institute (IDIBELL), L'Hospitalet de Llobregat
Barcelona, Catalonia 08908, Spain
4 Department of Cancer Biology
University of Texas MD Anderson Cancer Center, Houston, Texas 77054, USA
5 Department of Pediatric Oncology and Hematology
Erasmus University Medical Center, Erasmus MC, 3015 CE Rotterdam, The Netherlands
6 Department of Cell Biology and Cancer Genomics Center, The Netherlands Cancer Institute
Plesmanlaan 121, 1066 CX Amsterdam, The Netherlands
7 Hubrecht Institute-KNAW, University Medical Centre Utrecht
Uppsalalaan 8, 3584CT Utrecht, The Netherlands
8 Department of Physiological Sciences II, School of Medicine
University of Barcelona, Barcelona, Catalonia 08036, Spain
9 Institució Catalana de Recerca i Estudis Avançats (ICREA)
Barcelona, Catalonia 08010, Spain
10 Erasmus MC, Rotterdam University
3000 CA Rotterdam, The Netherlands
* These authors contributed equally to this work

Published in Nature Communications 6, article number: 6520, doi:10.1038/ncomms7520

Introduction

For several decades the foundations of molecular biology leaned against the dogma that genetic information is stored in protein-coding genes (Crick et al., 1961). Although this concept was, and is still, largely true in prokaryotes, where genomes are mainly composed of protein-coding genes, it does not hold true for higher eukaryotes, where protein-coding sequences occupy less than 3% of the genome. Once considered transcriptionally inactive or simply referred to as ‘junk DNA’, the predominant fraction of the genome is in fact pervasively transcribed into thousands of different noncoding RNAs (ncRNAs), which can further be divided into two groups: small ncRNAs and long noncoding RNAs (lncRNAs). In addition, lncRNA genes have been classified based on the epigenetic state of their chromatin. For instance, the long intergenic noncoding RNAs (lincRNAs) are known for the presence of high histone 3 lysine 4 trimethylation (H3K4me3) at their promoters and high H3K36me3 along their transcribed regions, also referred as the K4K36 signature (Guttman, 2009). Alternatively, enhancer RNAs (eRNAs) are produced from transcriptionally active enhancer regions, which are epigenetically defined by high level of H3K4me1, low level of H3K4me3 (Heintzman, 2007; Visel, 2009) and high level of histone 3 lysine 27 acetylation (H3K27Ac) and H3K9Ac (Creyghton, 2010).

Importantly, lincRNAs have recently emerged as potent regulators of gene expression. Recent publications have shown that lincRNAs are able to form complexes with various chromatin modifiers and to specifically direct them to different genomic regions. For example, the lincRNA-p21 was shown to interact with and guide the heterogeneous nuclear ribonucleoprotein K to repress a subset of p53 target genes (Huarte, 2010). However, although lincRNAs can mediate their effect in cis and in trans, eRNAs have been so far mainly characterized for their function in cis. Classically expressed as bidirectional transcripts from enhancer regions, eRNAs can alter the expression of their neighbouring genes through the formation of DNA loops, which help to bridge the interaction between enhancers and nearby promoters. Several transcription factors (TFs) were found to be important coordinators of eRNA expression (Lam, 2013; Li, 2013; Melo, 2013). An interesting case revealed that the tumour-suppressor p53 directly regulates the expression of eRNAs upon cellular stresses (Melo, 2013).

P53 function is frequently compromised in tumours, in part as a consequence of somatic mutations, which occur in more than 50% of all human cancers (Beckerman and Prives, 2010). Moreover, it was also shown that p53 is inactivated in various cancers by dysregulation of its regulatory pathway, such as the amplification and over-expression of its negative regulators MDM2 and MDM4 (Gembarska, 2012; Momand et al., 1998). Upon cellular stresses, p53 is activated and acts primarily as a TF to mediate and coordinate a complex transcriptional response that regulates hundreds of target genes. Until recently, the p53 network was mainly characterized by its impact on protein-coding target genes (Wei, 2006). However, we now begin to discover and appreciate the great potential of ncRNAs in the intricate regulatory network of p53. The recent discoveries that p53 can mediate its function in collaboration with diverse lncRNAs, suggest a potential role for this novel regulatory layer in disease such as cancer, and therefore urge the importance of an in-depth reassessment of the p53 transcriptional response.

Here, by using Global Run-On sequencing (GRO-seq), we mapped p53-responsive enhancers bound by p53. Surprisingly, we also found a large group of p53-activated enhancers that were not associated with p53. Although motif-search analysis identified the p53 signature in the enhancers bound by p53, no single TF was found to govern the majority of p53-unbound enhancer groups. However, further analysis revealed that nutlin-3a-induced Signal transducer and activator of transcription 3 (STAT3), B-cell lymphoma 3-encoded protein (BCL3), FBJ murine osteosarcoma viral oncogene homolog (FOS) might largely contribute to the transcriptional regulation of indirect p53 target genes. Next, we assessed whether p53-responsive lncRNAs could play a role in p53-mediated enhancer activation. Our data revealed that a prominent p53-induced lncRNA termed LED (LncRNA activator of Enhancer Domains) is required for p53-induced cell cycle arrest and is involved in the activation of a subset of p53-bound and unbound enhancers by inducing an epigenetic change. Strikingly, promoter-associated hypermethylation of LED was uncovered in several cancer cell lines and human tumours with preference to p53 wild-type (WT) status, suggesting its implication in tumorigenesis. Altogether, we propose that LED is an important regulator and a potential tumour suppressor of the p53 pathway.

Results

Genome-wide identification of p53-regulated eRNAs

To detect active enhancers, we relied on the observation that eRNA production marks enhancer activity. Using GRO-seq of MCF-7 cells treated with nutlin-3a, a specific activator of p53, we obtained a genome-wide quantitative snapshot of transcriptional activity. As expected, the activation of CDKN1A/p21, and many other known target genes of p53, was readily apparent (Fig. 1a). Moreover, also several previously described p53-induced eRNAs could be confirmed (Fig. 1a) (Melo, 2013). We proceeded to generate a global view of the putative p53-regulated enhancers in MCF-7 cells. By taking the union of the enhancer domains defined by the Broad chromatin segmentation (Ernst, 2011), we selected only regions showing RNA polymerase II (RNAPII) and p300 binding in MCF-7 cells (using publicly available chromatin immunoprecipitation sequencing (ChIP-seq) data (Nikulenkov, 2012)) and excluded those having annotated transcripts on both strands, as well as those having transcription start sites. The remaining regions were extended by 1 kb for the purpose of read counting and used in conjunction with the aforementioned GRO-seq data (Fig. 1b). This analysis resulted in the detection of 50,502 putative enhancers of which 6,270 were regulated (at least one direction) by nutlin-3a treatment, and referred here as p53-regulated enhancer regions or p53RERs.

Since p53 mainly functions as an activator of transcription (Wang et al., 2009), the vast majority (72%) of the differentially expressed eRNAs showed induction upon nutlin-3a treatment, and activated enhancers were more often bound by p53 than the repressed enhancers (Fig. 1c). Using ENCODE ChIP-seq data obtained from MCF-7 cells, the presence of enhancer-specific histone modifications was confirmed (Fig. 1d). Moreover, on average, nutlin-3a-induced enhancers are positioned closer to p53-regulated canonical genes (median 41 kb), than to non-regulated genes (median 161 kb; Fig. 1e). This observation is in agreement with the notion that eRNAs are potent regulators of neighbouring target genes (Lam, 2013; Li, 2013; Melo, 2013). In support of a role for nutlin-3a-regulated enhancer regions within the p53 pathway, the gene ontology analysis on neighbouring genes revealed enrichment for genes involved in DNA damage response/signal transduction by p53 (GO term 0030330, P=1.6e−3). Moreover, a de novo motif analysis using HOMER (Heinz, 2010) confirmed the presence of a p53 response element at p53-bound enhancer regions (p53BERs; Fig. 1f).

Next we reanalysed published p53 ChIP-seq data (Nikulenkov, 2012) to identify which of the p53RERs were direct targets of p53 (hereafter referred to as p53BERs). The enhancers in the remaining subset of p53RERs were considered p53-free enhancer regions, or p53FERs, as no enrichment for p53 or any known TF signature was found. Intriguingly, activation of p53BERs and p53FERs was different, as we observed a faster transcription

Figure 1 (next page): Identification of p53-regulated enhancer RNAs (p53RERs).
(a) GRO-Seq snapshot showing the induction of CDKN1A/p21 transcription upon 12 h nutlin-3a treatment (upper scheme). Display of nutlin-3a-induced bidirectional transcription at p53-bound and unbound enhancers (lower schemes). Binding of p300, presence of H3K27 acetylation and chromatin states in nine cell lines are also presented. **(b)** Diagram showing the outline of the algorithm to identify enhancers using chromatin segmentation data, and ChIP-Seq and GRO-Seq data. **(c)** Venn diagram showing the number of retrieved regions at the steps of the enhancer identification algorithm. **(d)** Boxplot showing the abundance of several enhancer marks at different regions (grey: 5,000 random non-repeat regions; red: promoters of the 5,000 most abundant genes as identified by GRO-seq in the nutlin-3a-treated condition; orange: all putative enhancer regions showing bidirectional transcription; blue: subset of the putative enhancers showing significant induction upon nutlin-3a treatment (induced p53RERs); dark blue: subset of the induced p53RERs having a p53 peak within 1 kb (p53BERs). **(e)** Boxplot showing the distances between enhancer region and the nearest annotated gene induced upon nutlin-3a treatment, for induced p53RERs (UP Enhancers) and nutlin-3a unresponsive enhancers (nonUP enhancers). **(f)** Motif identified in induced p53BERs using HOMER. **(g)** Normalized density of transcription downstream of the point of bidirectional transcription for p53BERs (red), p53FERs (dark grey) and uninduced enhancers (light grey). The lines indicate the median across all regions. The boxplot in the inset shows the distance from the point of bidirectional transcription to the 75% quantile of read density.

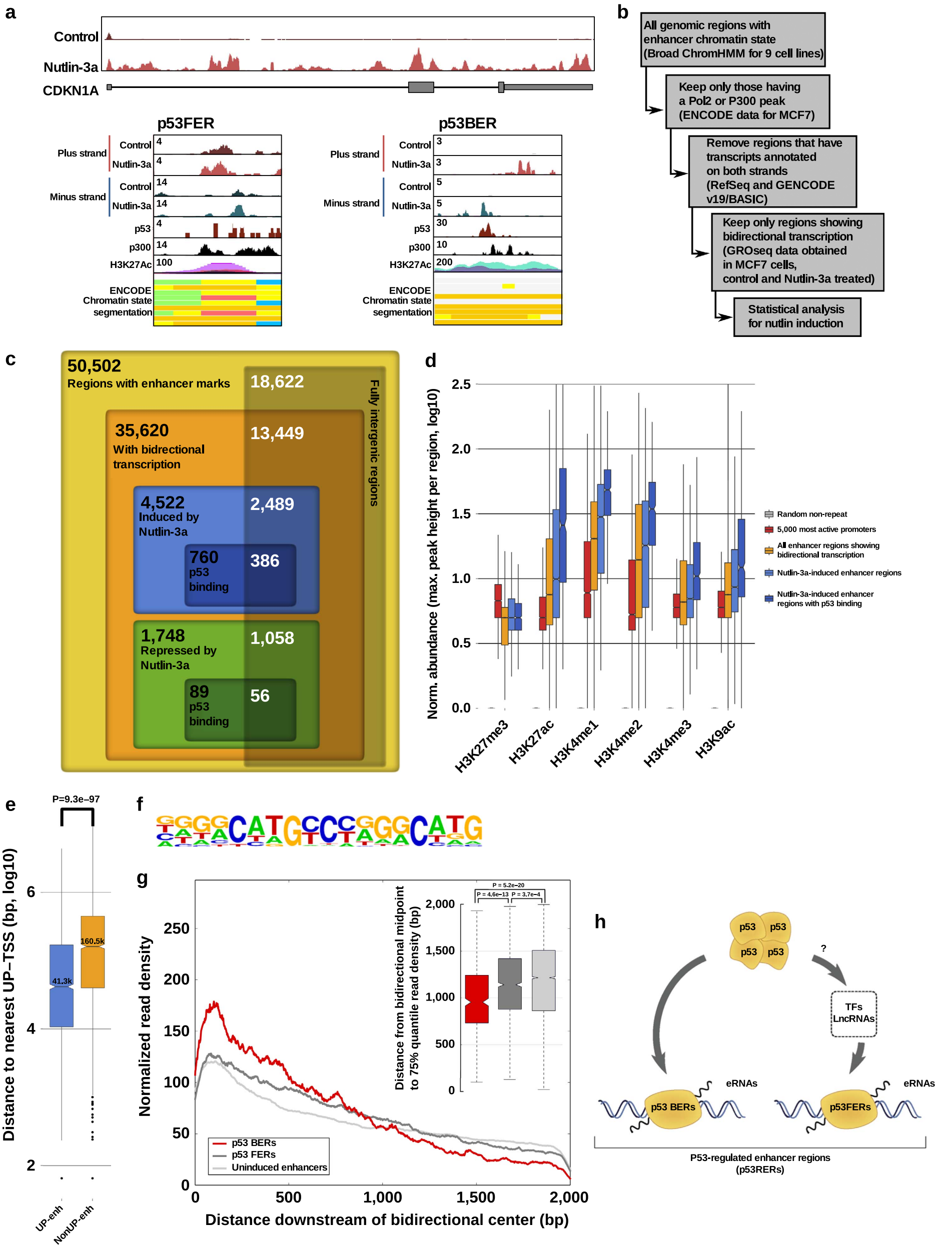


Figure 1 (cont.)

(h) Schematic representation of p53-regulated enhancers (p53RERs). P53 can directly bind to enhancers (p53BERs) or regulated intermediate factors (for example, lncRNAs or TFs) to indirectly influence another subset of enhancers (p53FERs).

drop-off for the first group compared with the second (Fig. 1g and Supplementary Fig. 1a). Although we did not further investigate this difference, we suggest that it reflects a secondary (p53-indirect) and, thus, differentially regulated wave of transcription (Supplementary Fig. 1b). Therefore, we hypothesized that activation of p53FERs could be mediated by a combination of several different factors (Fig. 1h). Indeed, the analysis of TF-binding sites (using the ENCODE Uniform TFBS data) showed that while several TFBSs were enriched in induced p53RERs, with respect to all enhancers, no TFBS was specifically enriched in the p53-free enhancer group (Supplementary Fig. 1c). However, further analysis revealed that among these potential regulators, three (STAT3, BCL3 and FOS) were regulated by nutlin-3a (Supplementary Fig. 1d,e). Intersection between their binding sites and p53FERs revealed that STAT3, BCL3 and FOS may directly regulate 55% of these enhancer regions (Supplementary Fig. 1f). Interestingly, despite their significant contribution, the transcriptional regulation of a large number of p53FERs remains elusive. An additional or complementary possibility is that the transcriptional activation of enhancers is mediated by p53-dependent lncRNAs, as lncRNAs were recently shown to be able to associate with and modulate regulatory elements (Vance, 2014; Yang, 2013).

LED is required for the p53 transcriptional response

We therefore set to identify relevant lncRNAs by profiling the transcriptome of nutlin-3a-treated MCF-7 cells. Using RNA-sequencing (RNA-seq) in combination with an annotation catalogue comprised of Ensembl, Refseq and the Broad Linc Catalog (Cabili, 2011), we identified 194 nutlin-3a-responsive lncRNA genes (Fig. 2a and Supplementary data 1). We then reasoned that the most upregulated transcripts might have a greater biological importance, and consequently selected the top three most activated lncRNAs (that is, RP3-510D11.2, loc643401 and linc00086 (hereinafter referred to as LED)) for further characterization (Fig. 2b). Interestingly, these three lncRNAs were also recently identified,

Figure 2 (next page): Novel p53-regulated lncRNA LED.
(a) Outline of pipeline for identification of p53-regulated lncRNAs. MCF-7 cells were treated with 8 μM nutlin-3a for 12 h and subjected to RNA sequencing (RNA-seq). **(b)** Display of genomic location and RNA-seq data showing the nutlin-3a induction of selected lncRNAs in MCF-7 cells. Values are represented by RPM (reads per million). **(c)** Stress-dependent regulation of selected lncRNAs upon nutlin-3a (8 μM) and ionizing radiation (IR; 10 Gy) treatment in MCF-7 cells measured by quantitative reverse transcription–PCR (qRT–PCR). Values are represented by fold induction (n=3; ***P<0.005, **P<0.01, *P<0.05, two-tailed Student's t-test). **(d)** P53-dependent regulation of validated lncRNAs in MCF-7 cells transfected with a control (Ctrl) or p53 siRNA in the presence or absence of nutlin-3a. Values are represented by fold induction (n=3; **P<0.01, *P<0.05, two-tailed Student's t-test). **(e)** Schematic representation of p53 response element (p53 RE) in LED gene body. Chromatin immunoprecipitation performed in nutlin-3a-treated MCF-7 cells using IgG or p53 antibodies followed by qPCR in the p53 RE region. Values represent the percentage of input (n=3; *P<0.05, two-tailed Student's t-test). **(f)** MCF-7 cells were co-transfected with an empty or LED (exon 2) p53BS pGL3-basic vector and either a Ctrl or p53-targeting siRNA. The relative luciferase activities (Firefly/Renilla) were normalized to the Ctrl reaction (empty pGL3-basic vector; n=3; ***P<0.005, **P<0.01, two-tailed Student's t-test). **(g)** qRT–PCR measuring relative LED RNA levels in MCF-7 cells transfected with a Ctrl or two independent LED siRNAs (LED-kd; n=3; ***P<0.005, *P<0.05, two-tailed Student's t-test).

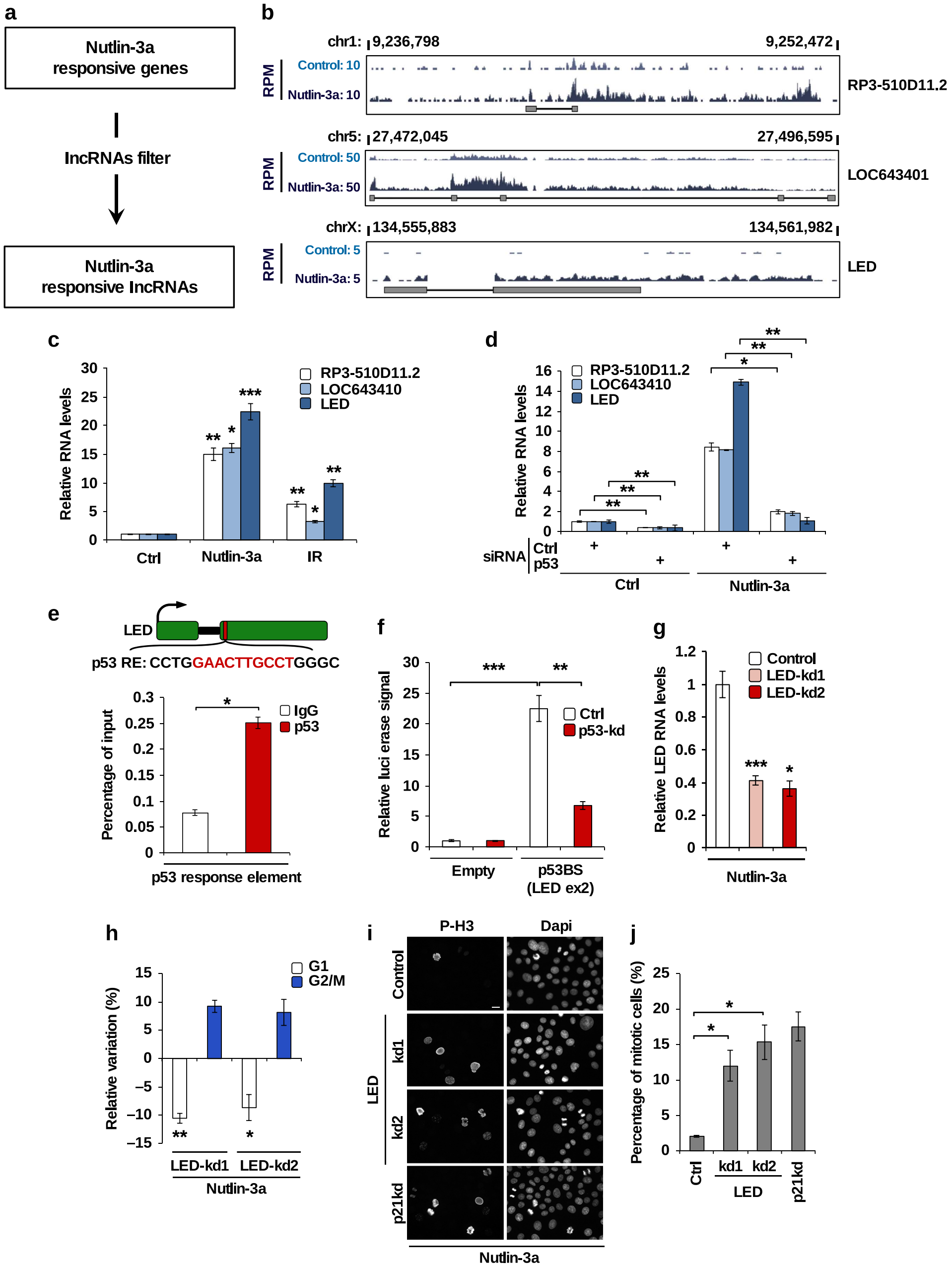


Figure 2 (cont.)

(h) Relative cell cycle variation (LED-kd minus control-kd) of MCF-7 cells transfected with a Ctrl or two independent LED siRNAs, treated with nutlin-3a for 12 h. To capture cycling cells in G2/M, cells were treated with nocodazole for 24 h, before flow cytometric analysis (n=3; **P<0.01, *P<0.05, two-tailed Student's t-test). **(i)** Immunostaining detection of the mitotic marker phospho-histone H3 ser10 (P-H3) and 4,6-diamidino-2-phenylindole (DAPI) staining in MCF-7 cells treated as in h and using a siRNA targeting p21 (p21kd) as a positive Ctrl. Scale bar, 25 μM. **(j)** Quantification of the marker P-H3 from i indicating the percentage of mitotic cells (n=3; *P<0.05, two-tailed Student's t-test).

but not functionally assessed, in two genome-wide studies performed in HCT-116 and Cal-51 cancer cell lines (Allen, 2014; Rashi-Elkeles, 2014). Validation confirmed that the selected lncRNAs were induced in MCF-7 cells upon both nutlin-3a and ionizing radiation treatment (Fig. 2c). Similar results were also obtained in ZR-75-1 and MALM-3M cell lines (Supplementary Fig. 2a). Next, we determined whether these lncRNAs were regulated by p53. As expected, we observed that p53 depletion decreased both the basal and nutlin-3a-induced levels of all tested lncRNAs (Fig. 2d). Moreover, we demonstrated the direct binding of p53 at each lncRNA locus by using publicly available p53 ChIP-seq data and ChIP-quantitative PCR (qPCR; Fig. 2e and Supplementary Fig. 2b). Using a luciferase reporter, we also showed the p53-dependent promoter activity of the p53 response element found in LED exon 2 (Fig. 2f). Altogether, these results demonstrate that our selected lncRNAs are bona fide p53 targets.

Next, we assessed whether the depletion of our selected p53-induced lncRNAs phenotypically influenced the p53 transcriptional response using short interfering RNAs (siRNAs). Among the investigated candidates, only LED significantly influenced the G1 checkpoint arrest following nutlin-3a treatment, as shown by flow cytometry (Fig. 2g,h). To corroborate this finding, we first evaluated cellular entry into mitosis using phospho-H3 (ser10) staining. Cells treated with siRNAs targeting LED showed a significant increase of phospho-H3 (ser10) compared with cells transfected with a non-targeting siRNA (Fig. 2i,j). Furthermore, cell proliferation assays confirmed this observation, as LED-suppressed cells proliferated more following nutlin-3a treatment in comparison with control-transfected cells (Supplementary Fig. 2c,d). Thus, the induction of LED lncRNA is required for efficient sustenance of p53 transcriptional response.

To investigate the mechanism by which LED impacts the p53 transcriptional response, we performed gene expression analysis by RNA-seq following knockdown of LED in MCF-7 cells treated 12 h with nutlin-3a. A total of 1,983 genes were responsive to LED depletion (FDR less than 1%), of which 1,340 were upregulated and 643 downregulated (Fig. 3a). Interestingly, LED knockdowns significantly reduced the levels of the cell-cycle regulator p21 (FDR of 2.7e-8; Fig. 3b), without influencing p53 levels (FDR of 0.12). We further validated this observation by showing the LED-dependent regulation of p21 at both the mRNA (Fig. 3c) and protein levels (Fig. 3d). Similar results were also obtained in ZR-75-1 and MALM-3M cell lines (Supplementary Fig. 3). Altogether, our results indicate that LED is required for an efficient p53-dependent checkpoint by maintaining high levels of p21.

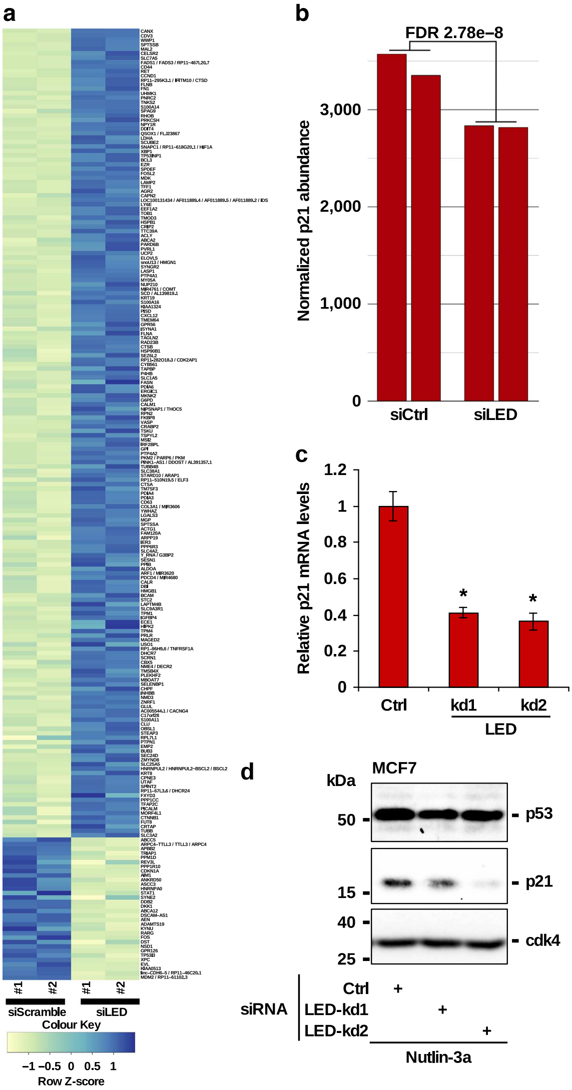


Figure 3: LED is required for the proper p53 transcriptional response. **(a)** RNA sequencing heatmap showing a subset of genes differentially expressed upon LED knockdown in MCF-7 cells treated 12 h with nutlin-3a (only the subset with absolute fold-change>30% and FDR<5% is shown). **(b)** Barplot derived from the RNA-sequencing showing the normalized p21 mRNA levels in control and LED knockdown conditions. **(c)** Relative mRNA levels of p21 upon transfection of a control or two independent LED siRNAs (LED-kd), measured by qRT–PCR in MCF-7 cells treated with Nutlin-3a for 12h (n=3; *P<0.05, two-tailed Student's t-test). **(d)** Western blot showing p53 and p21 protein levels in MCF-7 cells transfected with a control or two independent LED siRNAs (LED-kd) and treated with nutlin-3a for 12 h. Detection of CDK4 protein levels is also displayed.

LED associates with and regulates enhancer domains

Next, to investigate whether LED, a bona fide lncRNA of ~5 kb, exerts its function in the nucleus or in the cytoplasm, we examined its subcellular localization (Fig. 4a,b and Supplementary Fig. 4a,b). As LED is partially located in the nucleus, we set out to assess its putative interaction with chromatin. We performed chromatin isolation by RNA purification technique (ChIRP, (Chu et al., 2011) using anti-sense oligos to LED (odd and even) or the bacterial β-galactosidase (lacZ) and confirmed the specific enrichment for LED, but not glyceraldehyde 3-phosphate dehydrogenase RNA (Supplementary Fig. 4c). Then, we sequenced the DNA fragments co-purified in the two pools, aligned reads to the genome and processed them using peak calling software in a pipeline developed for ChIRP-seq data. Overlap of the peaks from the odd and even purifications indicated LED binding in 1,698 putative sites (Supplementary data 2). To investigate the nature of the genomic features present at LED-associated domains, we made use of chromatin state annotations previously defined by Ernst and colleagues (Ernst, 2011). Intriguingly, although LED-associated sites were present in all chromatin states, significant enrichment was observed in strong enhancer regions (Fig. 4c and Supplementary Fig. 4d). Moreover, intersection with the GRO-Seq data revealed that a subgroup of LED-bound enhancers was sensitive to nutlin-3a (Fig. 4d). Interestingly, further analysis showed that this subgroup was partially overlapping with p53, STAT3, BCL3 and FOS (Supplementary Fig. 4e). This observation suggests LED as a co-factor in the nutlin-3a-dependent regulation of enhancers.

To assess the regulatory potential of LED on enhancers, we first selected a subset of LED-associated enhancer domains (Supplementary Fig. 4f). Then, we reasserted that bound enhancers harbour hallmarks of active enhancers (Birney, 2007; Heintzman, 2007, 2009) (Supplementary Fig. 4g). Furthermore, we performed ChIP for H3K4me1 and H3K4me3, to confirm the relative deposition of these histone modifications in our cell system (Fig. 4e). Next, we tested a selected group of LED-associated enhancers for eRNA production by quantitative reverse transcription–PCR; after DNase-treatment of RNA isolated from MCF-7 cells incubated with or without nutlin-3a. As previously observed with the GRO-Seq, this analysis confirmed the nutlin-3a-dependent transcriptional induction of eRNAs at all tested LED-associated enhancers (Fig. 4f). This nutlin-3a induction of eRNAs was specific, as the abundance of a control, LED-unbound, FOXC1 enhancer (FOXC1e) RNAs remained unaffected. Strikingly, RNA interference-mediated LED knockdown reduced the level of activation of these putative eRNAs (Fig. 4g), indicating direct regulation of eRNA production by LED. Intriguingly, we noticed among the LED-associated enhancers a prominent peak located within the first intron of p21. We further validated the association of LED to p21 enhancer (p21e) domain using ChIRP-qPCR (Supplementary Fig. 4 h). To evaluate the enhancing potential of p21e, we cloned a 1.2-kb fragment into a pGL3-promoter luciferase reporter vector. As expected from an enhancer domain, p21e activated the luciferase gene in an orientation-independent manner (Fig. 4h). A detailed analysis of p21e 1.2 kb fragment revealed the presence of a p53 response element overlapping LED-binding site (Supplementary Fig. 4i). Thus, we suggested that both LED and p53 may participate in regulating p21e enhancing activity. Indeed, we demonstrated that LED or p53 knockdown decreases the enhancing activity of both sense and antisense p21e luciferase reporters (Fig. 4i). Using northern blotting and RNAPII ChIP experiments, we further supported the presence of an antisense eRNA at p21e locus and its regulation by nutlin-3a and LED (Fig. 4j,k and Supplementary Fig. 4j). Last, we examined whether p21e could interact with distant promoters by DNA looping, using circular chromosome conformation capture (4C) experiments. This analysis failed to reveal long-distance enhancer–promoter interactions, suggesting that p21e acts within its functional domain on the p21 promoter (Supplementary Fig. 4k). Collectively, these results demonstrate that LED associates with chromatin regions marked as enhancers and regulates the production of eRNAs.

To further delineate the mode of action by which LED regulates enhancers, we hypothesized that LED controls enhancer activity by remodelling the epigenetic state of enhancer domains. To investigate this possibility, we assessed whether LED influences the deposition of active enhancer histone marks, such as H3K27ac and H3K9ac. ChIP analyses revealed that the levels of H3K9ac, but not H3K27ac, were decreased at p21 enhancer domain upon LED knockdown (Fig. 4l and Supplementary Fig. 4l). Similar results were also obtained with another LED-associated enhancer (Supplementary Fig. 4m). Interestingly, in concomitance with H3K9ac reduction, we also noticed a lower p53-binding affinity at distal and proximal enhancers located upstream of p21 transcription start site (Supplementary Fig. 4n,o). These results indicate that LED may influence the production of eRNAs, by influencing the deposition of H3K9ac at specific enhancer loci. In addition or as a consequence of its influence on H3K9ac, LED may influence the binding of TFs at or in the vicinity of enhancer domains.

LED is inactivated by promoter hypermethylation in cancer

Gene expression comparison analysis suggests not only that LED is activated by p53, but also that its function is intimately linked to the transcriptional response of p53. We therefore examined whether LED is inactivated in cancer. Inspection of the LED promoter sequence identified a large CpG island region (Fig. 5a). As CpG islands are often subject to hypermethylation and silencing, we asked whether LED promoter hypermethylation leads to a reduced LED expression in cancers. We first measured the methylation status of LED CpG islands in 135 cancer cell lines covering a wide range of cancers. Notably, we find LED promoter methylation in ~44% (59/135) of all tested cell lines, with a large proportion in leukaemia (Supplementary Fig. 5a–c). Moreover, we observed a strong preference for methylation in p53 WT cell lines (60%; 29/48) as compared with p53 mutants (34%; 30/87, $P=0.004$ ($\chi 2$); Fig. 5a, Table 1). Most importantly, we then assessed and validated the transcriptional silencing of LED by its promoter-associated hypermethylation on several cancer cell lines. As expected, there was a significant anti-correlation between LED expression and its methylation status (Fig. 5b and Supplementary Fig. 5d,e). Also, treatment of LED-promoter-hypermethylated cell lines with the DNA-demethylating agent 5-Azacytidine resulted in LED re-expression (Fig. 5c). Moreover, we observed that methylation-dependent inactivation of LED may delay or reduce the induction of p21 mRNA, as compared with unmethylated cell lines (Supplementary Fig. 5f). Finally, we

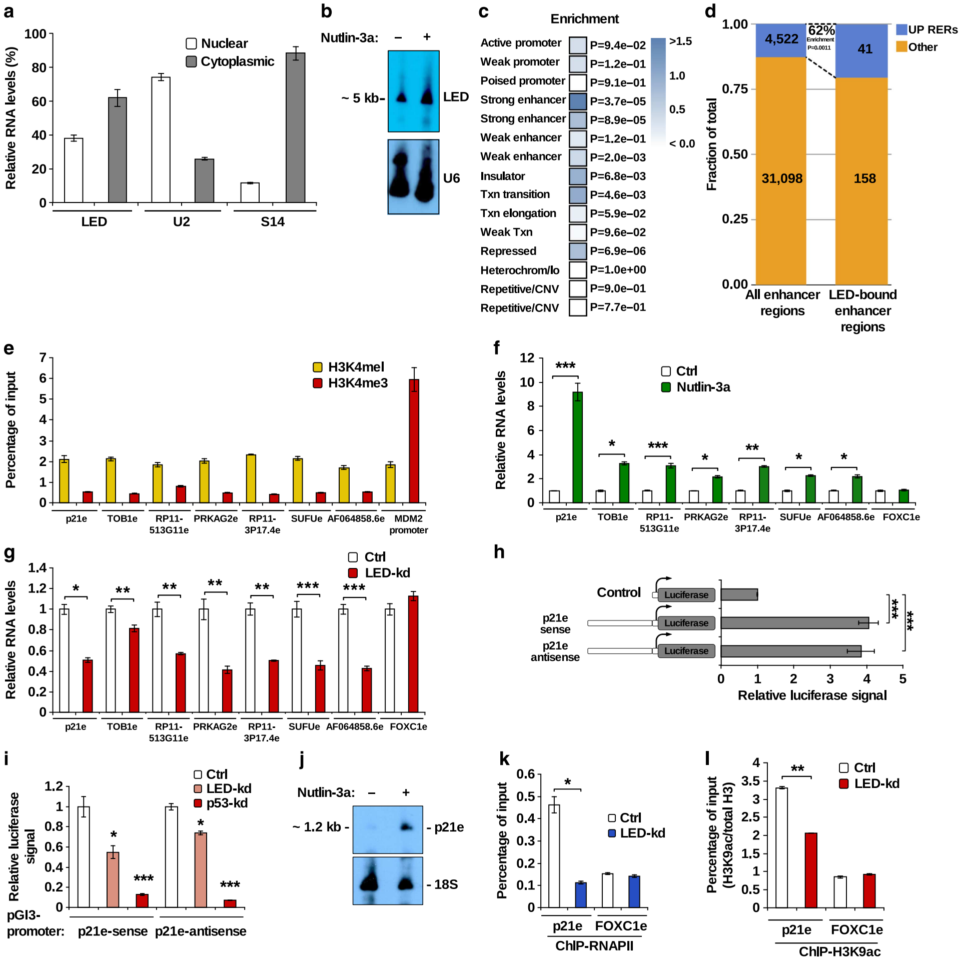


Figure 4: LED binds preferentially to enhancers and regulates enhancer RNA production from p53RERs.

(a) LED subcellular localization in MCF-7 cells treated with nutlin-3a. U2 and S14 genes were used as controls (Ctrl) for nuclear and cytoplasmic fraction, respectively. **(b)** Northern blot analysis showing LED transcript in MCF-7 cells incubated or not with nutlin-3a. U6 was used as a loading Ctrl. **(c)** Enrichment of LED ChIRP peaks in genomic features defined by ENCODE. **(d)** Bar graph showing the fraction of induced p53RERs of all found putative enhancer regions (left) and of all LED-bound enhancer regions (right). The enrichment of induced p53RERs in the LED-bound fraction is significant with P=0.0011 (hypergeometric distribution). **(e)** Quantification of H3K4me1 and H3K4me3 at LED-associated p53RERs by ChIP-qPCR in MCF-7 cells. Values were corrected to total H3 and MDM2 promoter was used as a Ctrl. Mean±s.d. are shown. **(f)** Nutlin-3a regulation of LED-associated p53RERs expression in MCF-7 cells. FOXC1e was used as a negative Ctrl (n=3; ***P<0.005, **P<0.01, *P<0.05, two-tailed Student's t-test). **(g)** LED-dependent regulation of p53RERs upon nutlin-3a treatment in MCF-7 cells transfected with a Ctrl or LED siRNA. FOXC1e was used as a negative Ctrl (n=3; ***P<0.005, **P<0.01, *P<0.05, two-tailed Student's t-test). **(h)** MCF-7 cells were transfected with an empty, p21e-sense or p21e-antisense reporter construct. The firefly/renilla luciferase activities were normalized to the Ctrl reaction (n=3; ***P<0.005, two-tailed Student's t-test).

evaluated the prevalence of LED-promoter hypermethylation in various human tumours. Using methylation-specific PCR, we observed LED-promoter hypermethylation in various tumour types, most prominently reaching 22% of all samples in acute lymphocytic leukaemia (ALL; Fig. 5d and Table 2).

Table 1: Association of LED CpG island hypermethylation with TP53 mutational status in human cancer cell lines.

Methylation	TP53 WT	TP53 mut
Methylated	29 (60%)	30 (34%)
Unmethylated	19 (40%)	57 (66%)
Total (135)	48	87

Table 2: Frequency of LED promoter associated-hypermethylation (M) in cancer patients.

Tumour type	Methylation	Total	%
Acute lymphocytic leukaemia	21	95	22
Cutaneous lymphomas	1	6	16
Folicular lymphomas	1	10	10
Melanomas	4	46	9
Acute myeloid leukaemia	1	13	8
Chronic lymphocytic leukaemia	1	33	3

Figure 4 (cont.)
(i) MCF-7 cells were co-transfected with an empty, p21e-sense or p21e-antisense pGL3-promoter vector and either a Ctrl, LED siRNA or p53 siRNA. The relative luciferase activities were normalized to the Ctrl reaction (empty vector) and subsequently to the Ctrl siRNA (n=3; ***P<0.005, *P<0.05, two-tailed Student's t-test). **(j)** Northern blot analysis showing p21e antisense transcript in MCF-7 cells treated or not with nutlin-3a. 18S was used as a loading Ctrl. **(k)** Quantification of RNAPII binding at p21e and FOXC1e regions by ChIP-qPCR. MCF-7 cells were transfected with a Ctrl or LED siRNA and treated with nutlin-3a (n=3; *P<0.05, two-tailed Student's t-test). **(l)** Quantification of H3K9Ac at p21e and FOXC1e regions by ChIP-qPCR. MCF-7 cells were transfected with a Ctrl or LED siRNA and treated with nutlin-3a. Values were normalized to total H3 (n=3; **P<0.01, two-tailed Student's t-test).

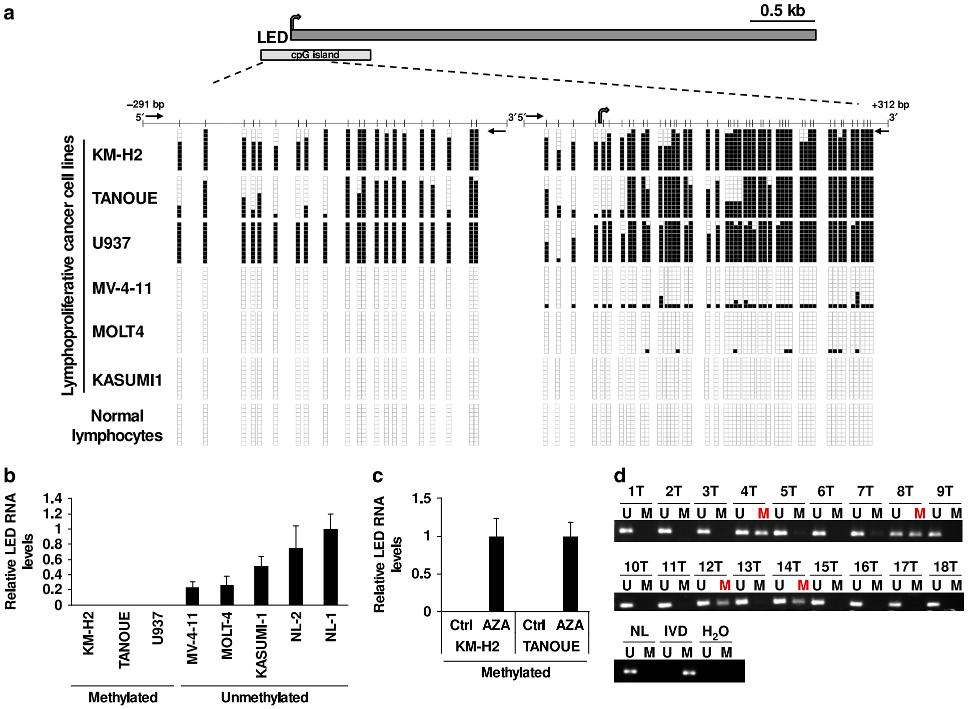


Figure 5: DNA methylation-associated silencing of LED in lymphoproliferative tumours. **(a)** Schematic representation of LED genomic loci and CpG island. Bisulfite genomic sequencing analysis of LED CpG island in human lymphoproliferative cancer cell lines and normal lymphocytes as tissue control (Ctrl). Location of bisulfite genomic sequencing PCR primers (black arrows), CpG dinucleotides (vertical lines) and the transcriptional start site (grey arrow) are shown. Ten single clones are represented for each sample. Presence of unmethylated or methylated CpGs is indicated by white or black squares, respectively. **(b)** LED expression levels in methylated or unmethylated human lymphoproliferative cell lines and in normal lymphocytes (NLs) as Ctrl. Values were determined by quantitative reverse transcription–PCR (qRT–PCR) in triplicates and are expressed as mean±s.e.m. (n=2–4). **(c)** Restored LED expression after treatment with DNA demethylating agent 5-aza-2′-deoxycytidine (AZA) in LED CpG island methylated cell lines. Values were determined in triplicates by qRT–PCR and are expressed as mean±s.e.m (n=3). **(d)** Methylation-specific PCR analyses for LED methylation in primary leukaemias. The presence of a band under the U lane indicates unmethylated alleles, whereas the presence of a band under the M lane indicates methylated alleles. Normal lymphocytes and in vitro methylated DNA (IVD) are shown as negative and positive Ctrls for methylated alleles, respectively.

Discussion

Coordination of gene expression components within response programmes is a delicate task crucial for the maintenance of cellular homeostasis. One key player for such coordination is the tumour-suppressor p53, which organizes the implementation of an appropriate cellular response to stress cues such as DNA damage and emerging oncogenes. With the discovery that the genome is pervasively transcribed (Core et al., 2008), it is likely that novel p53-sensitive transcripts and regulatory networks will be uncovered. Although many target promoters of p53 are well-established, little is known about the role of this master tumour suppressor as enhancer factor.

eRNAs were recently suggested as transcriptional regulators (Hah et al., 2013; Melo, 2013). Moreover, eRNA level emerges as robust readout for determining enhancer activity, as it correlates with the expression levels of neighbouring target genes. The GRO-seq is a very powerful technique that can be used to globally measure newly synthesized eRNAs and to infer enhancer activity in a genome-wide manner. Here we used GRO-seq to map and quantify eRNAs induced by the p53 inducer nutlin-3a, and identified hundreds of regulated enhancer domains. Although many enhancers are direct targets of p53, most nutlin-3a-regulated enhancer domains were not bound by this TF. Thus, it is likely that those enhancers are bound and influenced by factors regulated by p53. In this respect, bioinformatics analyses revealed three TFs (STAT3, BCL3 and FOS) with potential regulatory impact on p53-free enhancer regions. However, despite the potential combined influence of these three TFs on approximately 55% of p53FERs, the regulation of a large fraction remains unexplained. In search of novel p53FER regulators, we discovered LED, a lncRNA induced by p53, and subsequently demonstrated its involvement in the regulation of p53-sensitive enhancers, including both p53BERs and p53FERs. In support of our finding, two recent studies reported that not only TFs but also trans-acting lncRNAs are present at transcriptional regulatory regions (Vance, 2014; Yang, 2013). For instance, the lncRNA Paupar was found to interact with the TF PAX6 at enhancer domains in order to modulate the expression of genes involved in neural stem cell fate. Thus, for the first time, we demonstrate the contribution of a p53-induced lncRNA, termed LED, in the regulation of enhancer-derived transcripts.

LED is a direct transcriptional target of p53. Suppression of LED expression attenuated the activation of target enhancer domains, as demonstrated by reduced eRNA production and by a lower H3K9 acetylation. We found that LED was associated with different genomic loci and especially enriched at enhancer domains producing eRNAs. Notably, a subgroup of these enhancers is regulated by p53. Moreover, some, but not all, LED-bound enhancers were concomitantly bound by p53. Despite this observation, all tested LED-bound p53-induced eRNAs responded to siRNA-mediated LED depletion. This suggests that LED is a p53-induced factor that contributes to both the direct and indirect p53 transcriptional response.

How exactly does LED trigger enhancer activation? Modulation of the chromatin epigenetic state plays an important role in the regulation of gene expression. Thus far, several studies have put forward the idea that lncRNAs are important epigenetic regulators.

For example, the lncRNA HOTAIR represses gene expression by interacting with and guiding the polycomb repressive complex 2 to target promoters, where it contributes to chromatin compaction by catalysing the methylation of histone H3 at lysine 27 (Gupta, 2010). Alternatively, ribonucleoprotein complexes such as HOTTIP:MLL/WDR5 activate gene expression by promoting the deposition of an active mark (H3K4me3) on promoters (Wang, 2011b). Here we complement these observations by showing that LED is essential for the acetylation of H3K9 at bound enhancers, a modification associated with active gene transcription. Moreover, the p21 locus analysis also revealed the potential implication of LED in the epigenetic regulation of nearby contacted enhancers. This finding is consistent with the fact that LED is required for proper p53 binding at p21 upstream enhancers, as well as for RNAPII loading and eRNA transcription at bound enhancers. Moreover, genome-wide deposition of H3K9 acetylation was previously reported to be enriched at regulatory elements such as promoters, enhancers and repetitive sequences (Ernst, 2011). Consequently, active transcriptional programmes may primarily be epigenetically governed by the action of a subset of activating lncRNAs. However, whether LED influences the epigenetic features of regulatory elements before the TF-binding dysregulation, remains to be elucidated.

P53 is one of the most commonly inactivated gene in human cancer, with somatic mutations occurring in approximately half of all human cancers (Hollstein et al., 1991). In addition, alterations in the p53 pathway often represent an alternative route to attenuate the function of WT p53 in tumour (Leveille, 2011; Voorhoeve, 2006). Here we demonstrate that LED lncRNA is largely silenced in p53 WT primary human ALL. Although our DNA methylation analysis mainly focused on ALL tumours, it is likely that LED inactivation also occurs in other p53 WT tumours, such as on breast, liver and prostate. Nevertheless, this important observation pinpoints the dysregulation of lncRNAs as a potent mechanism in tumorigenesis. In support of this concept, other lncRNAs have been linked with cancer. For example, the oncogenic lncRNA HOTAIR is highly expressed in breast tumours and promotes cancer metastasis by guiding polycomb repressive complex 2 to specific genomic loci (Gupta, 2010). The lncRNA ANRIL and SChLAP1 are overexpressed in prostate cancers and antagonize the tumour-suppressive activity of INK4a/b and SWI/SNF complex, respectively (Prensner, 2013; Yap, 2010). Finally, tumour-suppressive lncRNAs such as GAS5 have been shown to be downregulated in cancer (Mourtada-Maarabouni et al., 2009).

Collectively, our results highlight a novel tumour suppressive mechanism involving a p53-induced lncRNA acting on enhancers (Fig. 6). The existence of a crosstalk between different lncRNA species uncovers an emerging regulatory network with potential considerable impacts in cancer development.

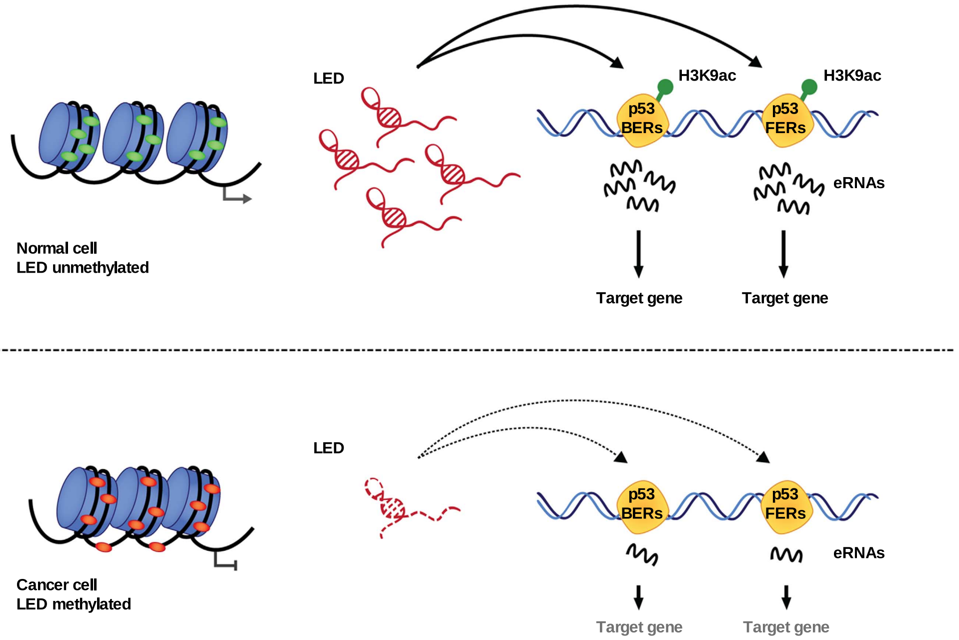

LED
H3K9ac
H3K9ac
p53
BERs
p53
FERs
eRNAs
Normal cell
LED unmethylated
Target gene
Target gene
LED
p53
BERs
p53
FERs
eRNAs
Cancer cell
LED methylated
Target gene
Target gene

Figure 6: Schematic representation of LED function in normal and cancer cells.
P53-bound enhancer regions (p53BERs), p53-free enhancer regions (p53FERs) and H3K9 acetylation are displayed.

Methods

Analysis of GRO-seq data and determination of enhancer regions

GRO-seq protocol was performed as previously described (Wang, 2011a). Briefly, MCF-7 cells were incubated with or without 8 μM nutlin-3a for 12 h and 5 million nuclei were isolated for each condition. rRNA reads were removed from the data by alignment to a rRNA index compiled from Ensembl annotations ('rRNA', 'rRNA_pseudogene' and 'Mt_rRNA') using bowtie2 (v.2.0.6, parameters '--seed 42 --end-to-end -N1 -L20 -i C,1 -D5 -R5') and keeping the unmapped reads. GRO-seq data were aligned to hg19 (including unassembled contigs) using bowtie2 (v. 2.0.6) with parameters '--seed 42 --sensitive'. Alignments with mapping quality lower than 10 and non-primary alignments were not considered in further analyses. Broad ChromHMM data for nine cell lines (Ernst, 2011) were used to screen putative enhancer regions. Along each chromosome, positions that were marked as enhancer regions (feature IDs 4, 5, 6 and 7) in at least one cell line were merged into regions. Transcription start sites, as annotated by RefSeq (obtained from the UCSC database server, 9 August 2012) and GENCODE v19/BASIC were extended by 1,000 bases and used to blacklist positions (that is, those positions were excluded as putative enhancer regions). Each merged region was tested for the presence of p300 and Pol2 as determined by ENCODE in MCF-7 cells (accessions GSM822295 and GSM1010800 (Johnson et al., 2007; Lee, 2012)) and each region without p300 and Pol2 peak was removed. The remaining regions were considered putative enhancer regions with enhancer marks. GRO-seq counts were obtained for each region, after extending each region by 1 kb. Regions having detectable transcription on both strands were considered putative enhancer regions with bidirectional transcription and used for downstream analyses. edgeR (Robinson et al., 2010) was used to determine statistical significance of differential expression of the enhancer regions (separately for each strand).

Generation of omnibus annotation

Ensembl annotations (v37.65), RefSeq gene annotations (obtained from UCSC database server 9 August 2012) and the Broad Linc RNA catalogue (Cabili, 2011) were merged in a single GTF (annotation file) using the gffread utility supplied with the Cufflinks package (v. 1.3.0), using parameters '-M -K -F -G'. This essentially collapses overlapping exons/transcripts so as to create an omnibus with low degree of redundancy yet high coverage of known and novel transcripts.

Analysis of RNA-seq data

RNA-seq samples were processed according to Illumina's protocol. Raw RNA-seq data were aligned using TopHat2 (v. 2.0.3) (Trapnell et al., 2009), using parameters '-m1 -p4 -F0.0 --segment-length 21 --segment-mismatches 1' and an exon annotation GTF file that was generated as described before. Reads with mapping quality less than 10 and non-primary alignments were discarded. Remaining reads were counted using HTSeq-count (http://www-huber.embl.de/users/anders/HTSeq/doc/overview.html), per gene ID. Statistical analysis of the differential expression of genes was performed using edgeR (Robinson et al., 2010). Genes with False Discovery Rate (FDR) for differential expression lower than 0.01 were considered significant.

Analysis of p53 ChIP data

p53 ChIP-seq data obtained from MCF-7 cells upon untreated and p53-stimulated conditions were obtained from SRA project SRP007261. Alignment was done using bowtie2 (v. 2.0.6) with parameters '--seed 42 --sensitive' to hg19 (including unassembled contigs). Only primary alignments with quality of at least ten were kept. Peaks were called by MACS (v. 2.0.10.20130501) using default parameters. Peaks with a fold-change (w.r.t. input) lower than 3.0 or a -log10(q-value) lower than 2.0 were discarded.

Sequence motif enrichment analysis

For enhancer regions the midpoint of bidirectional transcription was established, after pooling the GRO-seq data of the nutlin-3a-treated replicates. Two hundred bases of DNA around the midpoint of bidirectional transcription were extracted. For the analysis of sequence enrichment of BERs and FERs, the backgrounds were enhancers with bidirectional transcription that had an FDR for differential expression of 0.75 or greater (for both strands). In addition, for FERs the background contained only enhancers that had no p53 peak within 1 kb. HOMER (Heinz, 2010) was used to search for overrepresented sequence motifs, using parameters '-nogo -nlen 4 -len 18 -S 5 -mis 2'.

Constructs

p21e domain sense and antisense were PCR amplified using gDNA derived from MCF-7 cell lines and subsequently cloned (Nhe1/Xho1) into pGL3-promoter luciferase reporter vector. All primers used are listed in Supplementary table 1.

Cell culture and transfection

MCF-7 cells were cultured in DMEM containing 10% FBS, penicillin and streptomycin at 37 °C and 5% CO2. Identification and validation of lncRNA regulated by nutlin-3a were carried out by treating MCF-7 cells with 2–8 μM of nutlin-3a for a period of 4–12 h. To induce a p53 stress-response, cells were also treated with 10 Gy of ionizing radiation for 12 h. RNA interference experiments were performed using Dharmafect transfection reagent-1 and between 20 and 60 nM of siRNA. For epigenetic study, cells were treated with 2 μM 5-aza-2′-deoxycytidine (A3656, Sigma) for 72 h.

Protein analysis

Whole-cell lysates were prepared as previously described (Leveille, 2011). Protein detection was performed using primary antibodies detecting p53 (DO1, Santa Cruz, 1:1,000), p21 (Sc-397, Santa Cruz, 1:1,000), CDK4 (Sc-260, Santa Cruz, 1:1,000), phospho-histone H3 (ser 10) (9701, Cell Signaling, 1:100). Proteins were visualized using adequate secondary antibody (Dako) and ECL reagents (GE Healthcare).

Quantitative real-time PCR

RNA isolation and cDNA preparation were carried out as previously described (Melo, 2013). Real-time qPCR analysis was performed using the LightCycler 480 SYBR Green I Master mix (Roche). The glyceraldehyde 3-phosphate dehydrogenase was used as an internal control.

Flow cytometry

Control or nutlin-3a-treated cells were arrested in mitosis using 250 ng ml−1 of nocodazole for 24–36 h. Cells were then trypsinized, washed and resuspended in PBS containing 0.6% NP-40, 50 mg ml−1 RNaseA and 50 mg ml−1 propidium iodide for 10 min. Cell cycle profiles were captured using FACSCalibur (BD Biosciences) and analysed with the Flowjo software.

Immunofluorescence

Cells were first plated at a density of 3 × 10^5 cells per well and concomitantly reverse transfected with a control siRNA or siRNAs against LED or p21. After 24 h, cells were trypsinized and seeded on microscope coverslips coated with polylysine. Next, cells were fixed with 3% formaldehyde and subsequently permeabilized with PBS-Triton X-100 (0.3%) solution. After blocking 1 h with 2% PBS-Milk, cells were successively incubated with the primary antibody phospho-histone H3 (ser 10) and the Alexa Fluor 488 Dye-conjugated secondary antibody. Images were captured using an AxioCam MRc CCD camera (Carl Zeiss Microimaging).

RNA fluorescence in situ hybridization

MCF-7 cells treated with nutlin-3a (8 µM) and non-treated controls were grown on coverslips in six-well plates overnight. The media were aspirated and cells washed 3 × in cold PBS. Fixation solution (5 ml of 10 × PBS, 5 ml of 37% formaldehyde (100% formalin) and 40 ml of Diethylpyrocarbonate (DEPC) treated H_2O) was added and cells were incubated for 20 min at 4 °C. Cells were washed 3 × in cold PBS and 70% cold ethanol was used to permeabilize cells at 4 °C for 24 h. Cells were washed with cold PBS and left in hybridization buffer (1 g of dextran sulfate, 7 ml of DEPC water, 1 ml of formamide and 1 ml of 20 × SSC buffer) for 1 h. A measure of 50 ng of stellaris probe were used in hybridization buffer and cells were kept incubating for 48 h at 37 °C. After washing with wash SSC buffer, coverslips were covered with Draq5 during 20 min for nuclear staining, washed with cold PBS and mounted using antifade buffer (850 µl of DEPC H_2O, 100 µl of 20 × SSC, 40 µl of 10% glucose, 10 µl of Tris). Images were captured in a Zeiss confocal microscope.

ChIRP

ChIRP was performed as previously described (Chu et al., 2012). ChIRP probes (48 × 40-mer) targeting LED and lacZ were designed at http://www.singlemoleculefish.com /designer.html. Probes antisense to LED were divided into two sets (odd and even). The input and odd and even probe samples were sequenced individually. After clipping of adapters from the obtained reads, data were aligned to hg19 using bowtie2 (v. 2.0.6) using parameters '--seed 42 -N 1 -p 2'. Reads with mapping quality less than 10 and non-primary alignments were excluded from further analysis. Peak detection was run using MACS (v. 2.0.10.20130501 (Zhang, 2008)) using parameters '-g hs -B -p 0.1'. Peaks with a −log10(q-value) ≥5 and an enrichment ≥4 with respect to the input were kept, and peaks found in the odd and even samples were intersected. Overlapping peaks in both samples that had a position-wise Pearson correlation of abundance ≥0.2 and at least 25 reads in both samples were merged. From the resulting set of peaks, plasmid contaminants were discarded (Supplementary Table 2).

ChIP

MCF-7 (5×10^6) cells were first transfected with a control or specific siRNAs. Next, cells were fixed with 1% formaldehyde for 8 min at room temperature and subsequently quenched with 125 mM glycine for 5 min on ice. The cells were pelleted 10 min at 470 g and re-suspended in 300 µl of cold lysis buffer (50 mM Tris-HCl, pH 8.0, 10 mM EDTA and 1% SDS) supplemented with protease inhibitor cocktail (Roche). The suspension was sonicated 20 min (30 s on/off at maximum power) and further diluted with 800 µl of dilution buffer (10 mM Tris-HCl, pH 7.5, 140 mM NaCl, 1 mM EDTA, 0,5 mM EGTA and 1% Triton X-100) supplemented with protease inhibitor cocktail. The lysate was centrifuged for 10 min at maximum speed and the soluble fraction (chromatin) was transferred to a new tube. For each ChIP reaction, 100 µl of chromatin preparation was diluted with 300 µl of dilution buffer and incubated on an end-to-end rotator with 2–10 µg of antibody at 4 °C overnight. Then, 30 µl of protein A/G beads, previously blocked 1 h with PBS/BSA (0.1%) solution, was add to each ChIP reaction and incubated 2–3 h at 4 °C. The immune-purified chromatin was washed 2 × 5 min with the dilution buffer and 1 × with TE (50 mM Tris-HCl pH 8.0 and 10 mM EDTA) and finally eluted in 300 µl elution buffer (20 mM Tris-HCl pH 7.5, 5 mM EDTA, 50 mM NaCl and 1% SDS) at 65 °C overnight. Eluted chromatin was purified using QIAquick PCR purification kit (Qiagen) and subjected to real-time PCR analysis. Antibodies and amounts used in this study were as follows: pol2 (8 µg, CTD4H8, Upstate), H3K9ac (3 µl, ab4441, Abcam), H3K27ac (3 µl, ab4729, Abcam), H3K4me1 (6 µl, ab8895, Abcam), H3K4me3 (3 µl, MC315, Upstate), histone H3 (5 µl, 2650, Cell Signaling).

Chromosome conformation capture on chip (4C)

4C templates preparation and analysis were performed as previously described (Splinter, 2011). Briefly, 10^7 of MCF-7 cells were harvested and crosslinked with formaldehyde 2% for 10 min at room temperature, and neutralized with 125 M glycine. After washing with PBS, cells were lysed in 150 mM NaCl, 50 mM Tris-HCl (ph 7.5), 5 mM EDTA, 0.5% NP40, 1% Triton X-100 and nuclei were recovered by spinning 8 min at 600 g. Nuclei were digested overnight with 400 U DpnII (NEB) and Csp6I (NEB) and re-ligated in 7 ml with 100 U of T4 DNA ligase (Roche) overnight at 16 °C. Purified DNA circles were digested with 50 U of DpnII (Csp6I circles) and Csp6I (DpnII circles) overnight at 37 °C, followed by heat inactivation and ligation in 14 ml with 200 U of T4 DNA ligase. The 4C template was then purified and used for PCR amplification.

DNA methylation analysis

The Methyl Primer Express v1.0 software was used to identify the CpG islands and design-specific primers for the methylation analysis (Supplementary table 1). DNA methylation status was established by bisulfite genomic sequencing of multiple clones or methylation-specific PCRs in DNA samples previously treated with sodium bisulfite (EZ DNA methylation Gold kit, Zymo Research). The Illumina 450 K methylation array was used to anlyse the methylation status in multiple human cancer cell lines. For epigenetic drug treatments, cells were treated with 1 μM 5-aza-2′-deoxycytidine (Sigma) for 72 h.

References

Allen, M.A. (2014). Global analysis of p53-regulated transcription identifies its direct targets and unexpected regulatory mechanisms. eLife *3*, e02200.

Beckerman, R., and Prives, C. (2010). Transcriptional regulation by p53. Cold Spring Harb. Perspect. Biol. *2*, a000935.

Birney, E. (2007). Identification and analysis of functional elements in 1% of the human genome by the ENCODE pilot project. Nature *447*, 799–816.

Cabili, M.N. (2011). Integrative annotation of human large intergenic noncoding RNAs reveals global properties and specific subclasses. Genes Dev. *25*, 1915–1927.

Chu, C., Qu, K., Zhong, F.L., Artandi, S.E., and Chang, H.Y. (2011). Genomic maps of long noncoding RNA occupancy reveal principles of RNA-chromatin interactions. Mol. Cell *44*, 667–678.

Chu, C., Quinn, J., and Chang, H.Y. (2012). Chromatin isolation by RNA purification (ChIRP). J. Vis. Exp.

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

Creyghton, M.P. (2010). Histone H3K27ac separates active from poised enhancers and predicts developmental state. Proc. Natl Acad. Sci. USA *107*, 21931–21936.

Crick, F.H., Barnett, L., Brenner, S., and Watts-Tobin, R.J. (1961). General nature of the genetic code for proteins. Nature *192*, 1227–1232.

Ernst, J. (2011). Mapping and analysis of chromatin state dynamics in nine human cell types. Nature *473*, 43–49.

Gembarska, A. (2012). MDM4 is a key therapeutic target in cutaneous melanoma. Nature Med. *18*, 1239–1247.

Gupta, R.A. (2010). Long non-coding RNA HOTAIR reprograms chromatin state to promote cancer metastasis. Nature *464*, 1071–1076.

Guttman, M. (2009). Chromatin signature reveals over a thousand highly conserved large non-coding RNAs in mammals. Nature *458*, 223–227.

Hah, N., Murakami, S., Nagari, A., Danko, C.G., and Kraus, W.L. (2013). Enhancer transcripts mark active estrogen receptor binding sites. Genome Res. *23*, 1210–1223.

Heintzman, N.D. (2007). Distinct and predictive chromatin signatures of transcriptional promoters and enhancers in the human genome. Nature Genet. *39*, 311–318.

Heintzman, N.D. (2009). Histone modifications at human enhancers reflect global cell-type-specific gene expression. Nature *459*, 108–112.

Heinz, S. (2010). Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. Mol. Cell *38*, 576–589.

Hollstein, M., Sidransky, D., Vogelstein, B., and Harris, C.C. (1991). p53 mutations in human cancers. Science *253*, 49–53.

Huarte, M. (2010). A large intergenic noncoding RNA induced by p53 mediates global gene repression in the p53 response. Cell *142*, 409–419.

Johnson, D.S., Mortazavi, A., Myers, R.M., and Wold, B. (2007). Genome-wide mapping of in vivo protein-DNA interactions. Science *316*, 1497–1502.

Lam, M.T. (2013). Rev-Erbs repress macrophage gene expression by inhibiting enhancer-directed transcription. Nature *498*, 511–515.

Lee, B.K. (2012). Cell-type specific and combinatorial usage of diverse transcription factors revealed by genome-wide binding studies in multiple human cells. Genome Res. *22*, 9–24.

Leveille, N. (2011). Selective inhibition of microRNA accessibility by RBM38 is required for p53 activity. Nat. Commun. *2*, 513.

Li, W. (2013). Functional roles of enhancer RNAs for oestrogen-dependent transcriptional activation. Nature *498*, 516–520.

Melo, C.A. (2013). eRNAs are required for p53-dependent enhancer activity and gene transcription. Mol. Cell *49*, 524–535.

Momand, J., Jung, D., Wilczynski, S., and Niland, J. (1998). The MDM2 gene amplification database. Nuclei. Acids Res. *26*, 3453–3459.

Mourtada-Maarabouni, M., Pickard, M.R., Hedge, V.L., Farzaneh, F., and Williams, G.T. (2009). GAS5, a non-protein-coding RNA, controls apoptosis and is downregulated in breast cancer. Oncogene *28*, 195–208.

Nikulenkov, F. (2012). Insights into p53 transcriptional function via genome-wide chromatin occupancy and gene expression analysis. Cell Death Differ. *19*, 1992–2002.

Prensner, J.R. (2013). The long noncoding RNA SChLAP1 promotes aggressive prostate cancer and antagonizes the SWI/SNF complex. Nature Genet. *45*, 1392–1398.

Rashi-Elkeles, S. (2014). Parallel profiling of the transcriptome, cistrome, and epigenome in the cellular response to ionizing radiation. Sci. Signal. *7*, rs3.

Robinson, M.D., McCarthy, D.J., and Smyth, G.K. (2010). edgeR: a Bioconductor package for differential expression analysis of digital gene expression data. Bioinformatics *26*, 139–140.

Splinter, E. (2011). The inactive X chromosome adopts a unique three-dimensional conformation that is dependent on Xist RNA. Genes Dev. *25*, 1371–1383.

Trapnell, C., Pachter, L., and Salzberg, S.L. (2009). TopHat: discovering splice junctions with RNA-Seq. Bioinformatics *25*, 1105–1111.

Vance, K.W. (2014). The long non-coding RNA Paupar regulates the expression of both local and distal genes. EMBO J. *33*, 296–311.

Visel, A. (2009). ChIP-seq accurately predicts tissue-specific activity of enhancers. Nature *457*, 854–858.

Voorhoeve, P.M. (2006). A genetic screen implicates miRNA-372 and miRNA-373 as oncogenes in testicular germ cell tumors. Cell *124*, 1169–1181.

Wang, D. (2011a). Reprogramming transcription by distinct classes of enhancers functionally defined by eRNA. Nature *474*, 390–394.

Wang, K.C. (2011b). A long noncoding RNA maintains active chromatin to coordinate homeotic gene expression. Nature *472*, 120–124.

Wang, B., Xiao, Z., and Ren, E.C. (2009). Redefining the p53 response element. Proc. Natl Acad. Sci. USA *106*, 14373–14378.

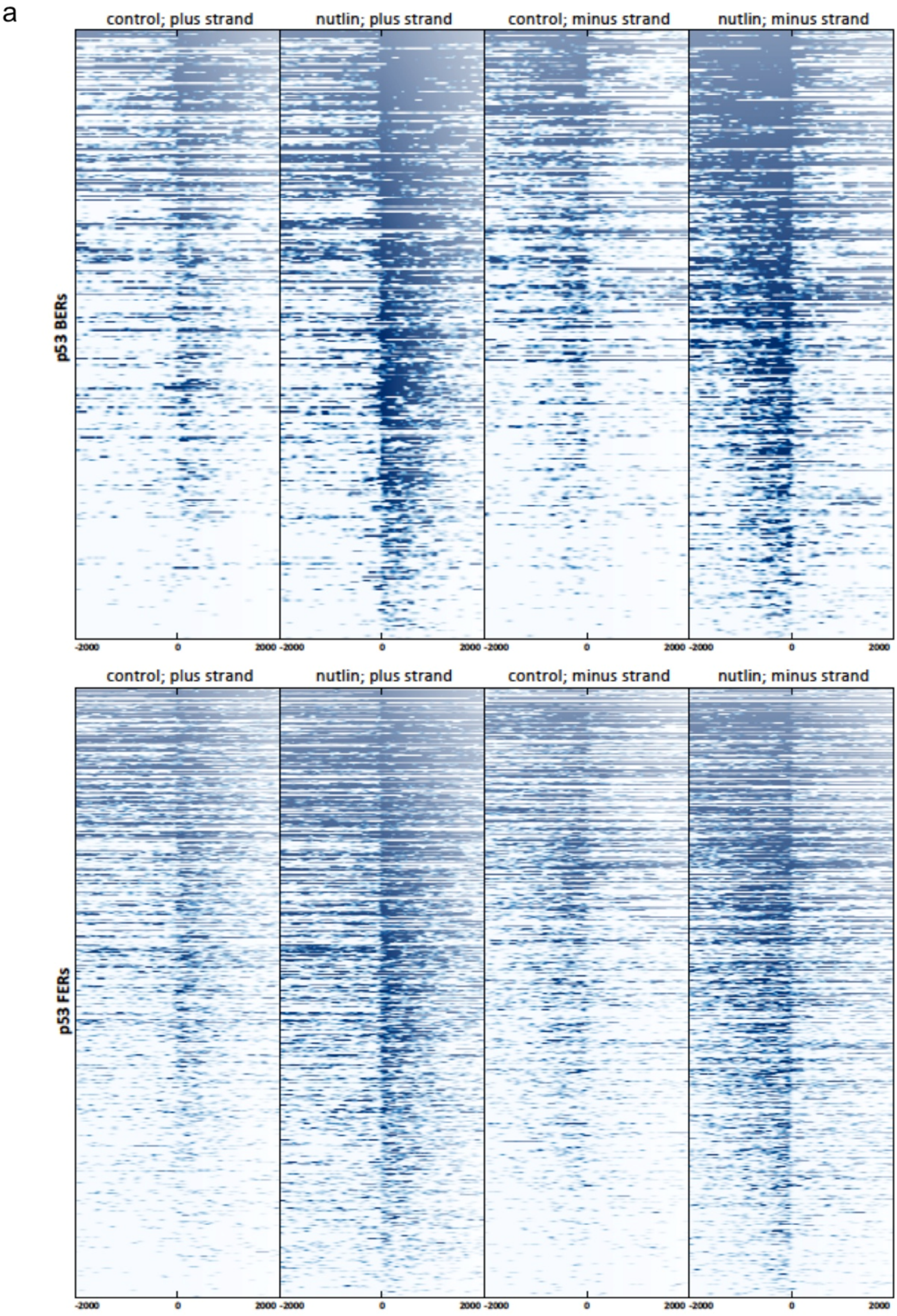
Wei, C.L. (2006). A global map of p53 transcription-factor binding sites in the human genome. Cell *124*, 207–219.

Yang, L. (2013). lncRNA-dependent mechanisms of androgen-receptor-regulated gene activation programs. Nature *500*, 598–602.

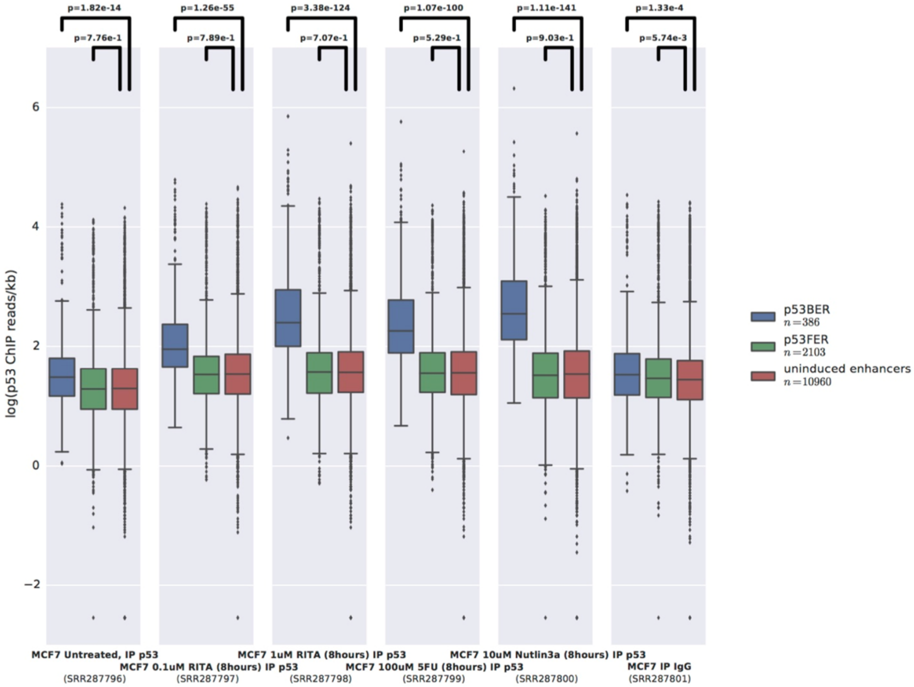
Yap, K.L. (2010). Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. Mol. Cell *38*, 662–674.

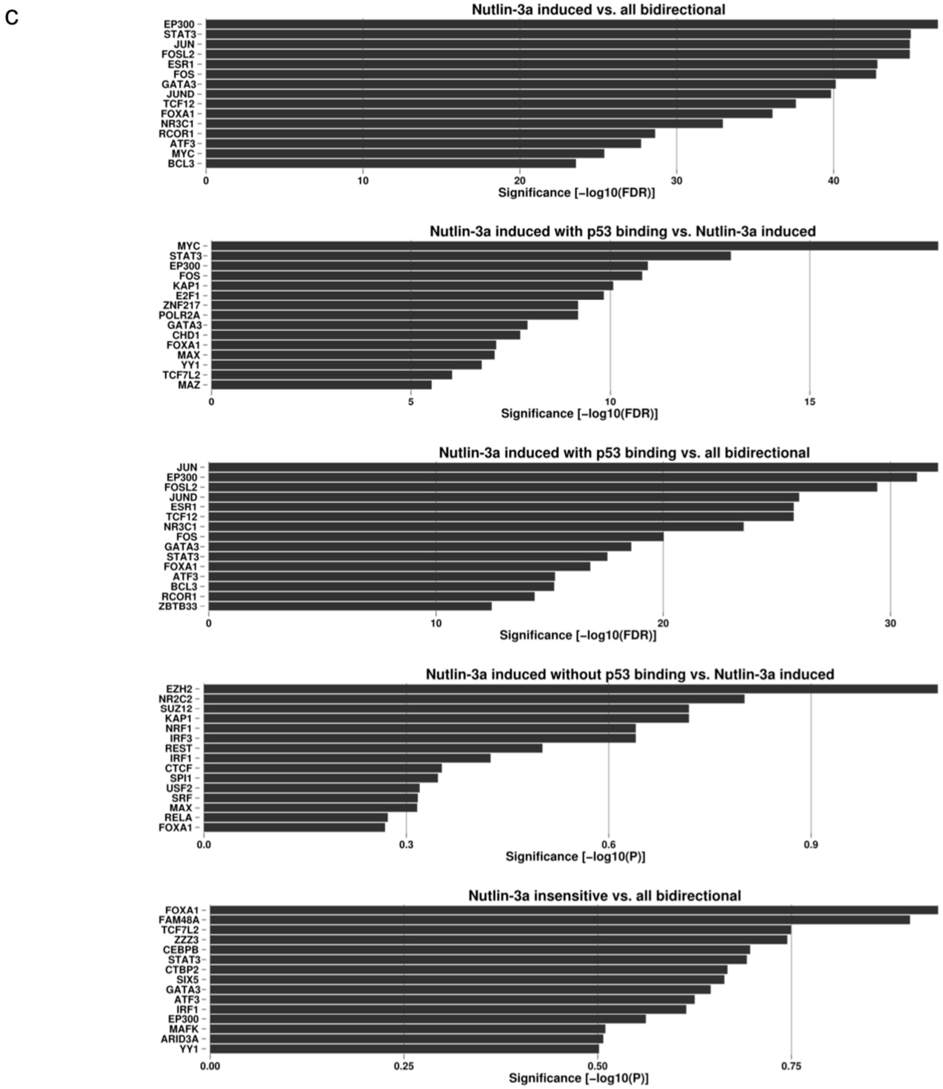
Zhang, Y. (2008). Model-based analysis of ChIP-Seq (MACS). Genome Biol. *9*, R137.

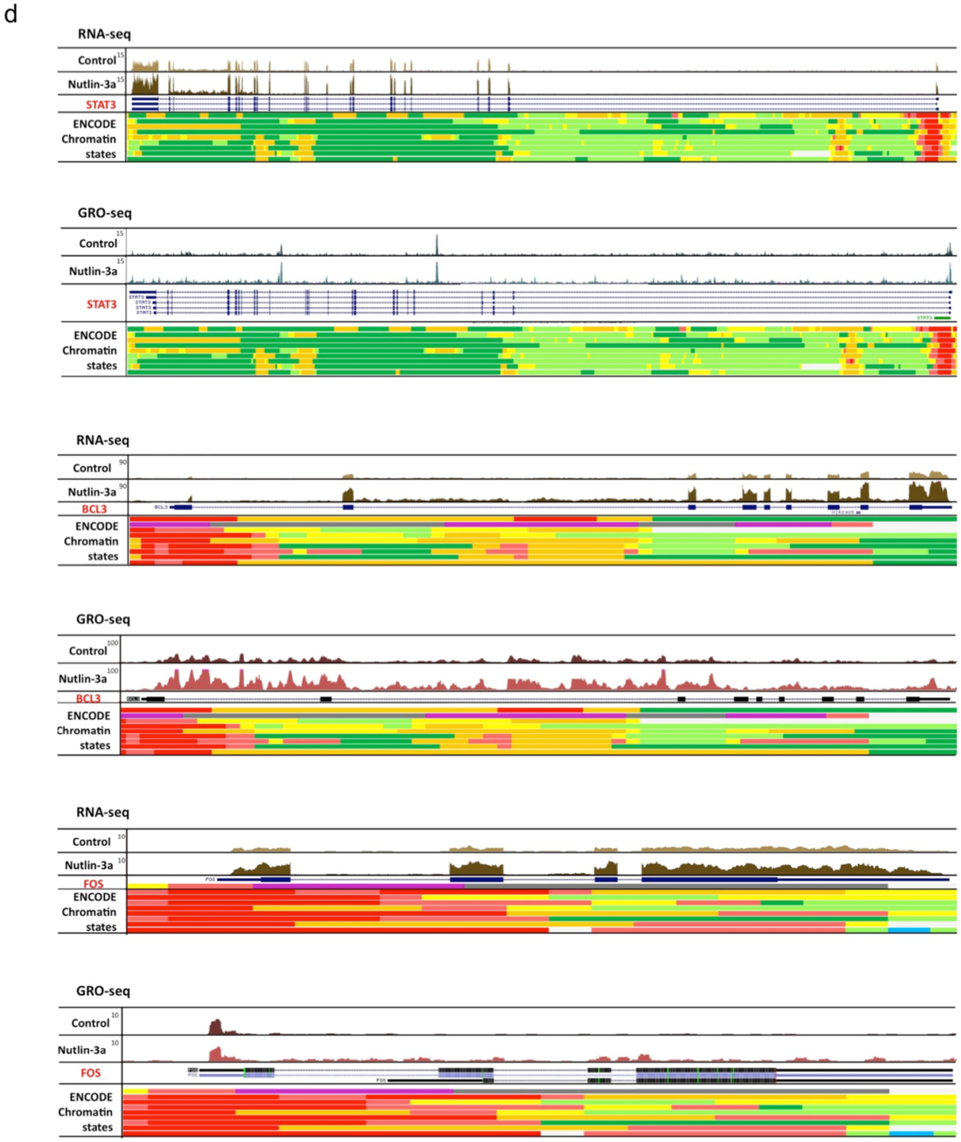
Supplementary figures

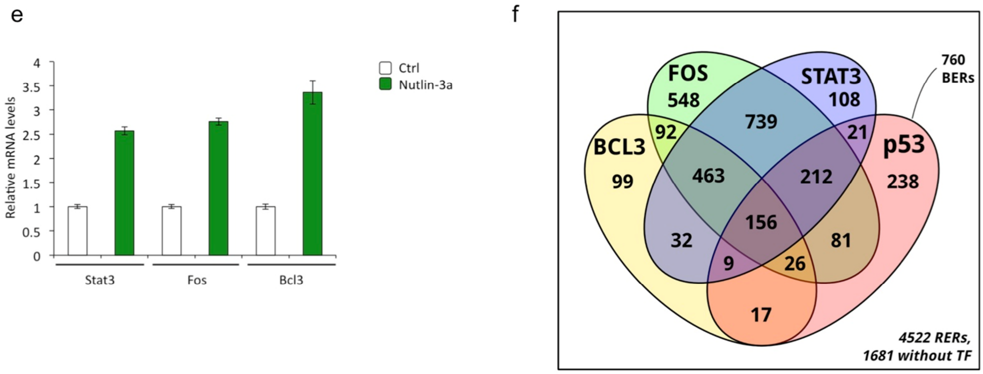


b

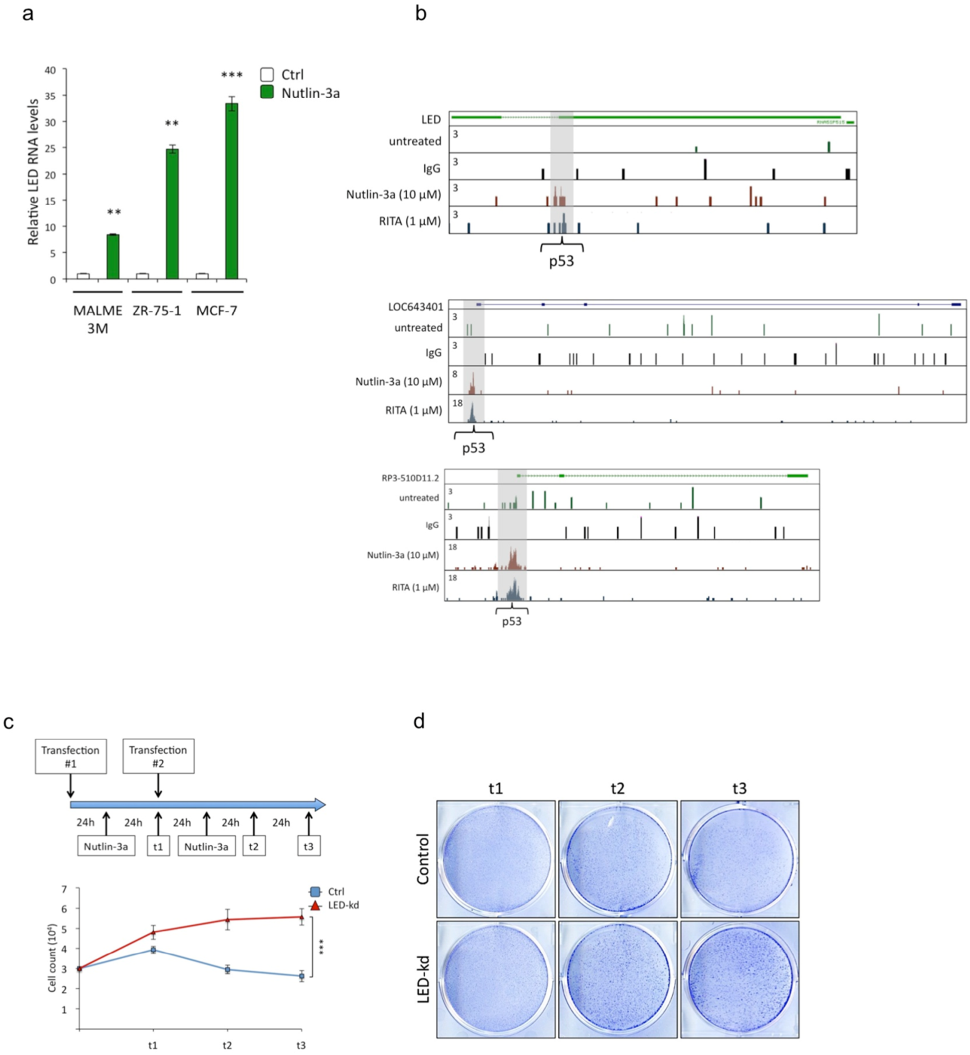

p=1.82e-14
p=7.76e-1
p=1.26e-55
p=7.89e-1
p=3.38e-124
p=7.07e-1
p=1.07e-100
p=5.29e-1
p=1.11e-141
p=9.03e-1
p=1.33e-4
p=5.74e-3
log(p53 ChIP reads/kb)
6
4
2
0
−2
p53BER $n=386$
p53FER $n=2103$
uninduced enhancers $n=10960$
MCF7 Untreated, IP p53 (SRR287796)
MCF7 0.1uM RITA (8hours) IP p53 (SRR287797)
MCF7 1uM RITA (8hours) IP p53 (SRR287798)
MCF7 100uM 5FU (8hours) IP p53 (SRR287799)
MCF7 10uM Nutlin3a (8hours) IP p53 (SRR287800)
MCF7 IP IgG (SRR287801)


c
Nutlin-3a induced vs. all bidirectional
EP300
STAT3
JUN
FOSL2
ESR1
FOS
GATA3
JUND
TCF12
FOXA1
NR3C1
RCOR1
ATF3
MYC
BCL3
0
10
20
30
40
Significance [−log10(FDR)]
Nutlin-3a induced with p53 binding vs. Nutlin-3a induced
MYC
STAT3
EP300
FOS
KAP1
E2F1
ZNF217
POLR2A
GATA3
CHD1
FOXA1
MAX
YY1
TCF7L2
MAZ
0
5
10
15
Significance [−log10(FDR)]
Nutlin-3a induced with p53 binding vs. all bidirectional
JUN
EP300
FOSL2
JUND
ESR1
TCF12
NR3C1
FOS
GATA3
STAT3
FOXA1
ATF3
BCL3
RCOR1
ZBTB33
0
10
20
30
Significance [−log10(FDR)]
Nutlin-3a induced without p53 binding vs. Nutlin-3a induced
EZH2
NR2C2
SUZ12
KAP1
NRF1
IRF3
REST
IRF1
CTCF
SPI1
USF2
SRF
MAX
RELA
FOXA1
0.0
0.3
0.6
0.9
Significance [−log10(P)]
Nutlin-3a insensitive vs. all bidirectional
FOXA1
FAM48A
TCF7L2
ZZZ3
CEBPB
STAT3
CTBP2
SIX5
GATA3
ATF3
IRF1
EP300
MAFK
ARID3A
YY1
0.00
0.25
0.50
0.75
Significance [−log10(P)]


d
RNA-seq
Control 15
Nutlin-3a 15
STAT3
ENCODE Chromatin states
GRO-seq
Control 15
Nutlin-3a 15
STAT3
ENCODE Chromatin states
RNA-seq
Control 90
Nutlin-3a 90
BCL3
ENCODE Chromatin states
GRO-seq
Control 100
Nutlin-3a 100
BCL3
ENCODE Chromatin states
RNA-seq
Control 10
Nutlin-3a 10
FOS
ENCODE Chromatin states
GRO-seq
Control 10
Nutlin-3a 10
FOS
ENCODE Chromatin states

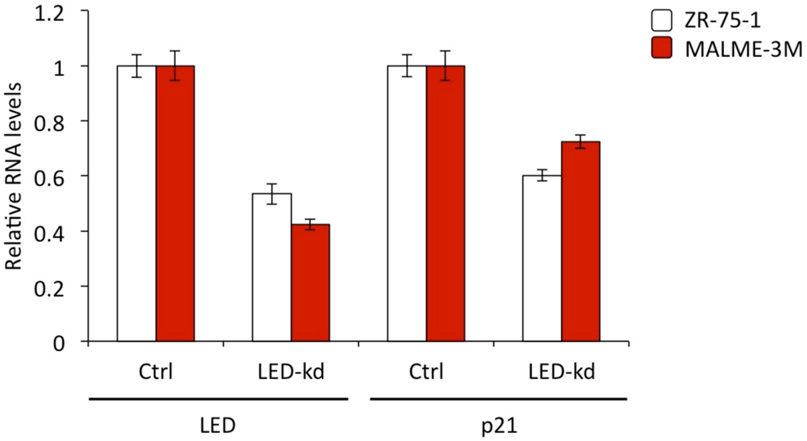


Supplementary Figure 1. Identification of p53-regulated enhancer RNAs (p53RERs).
(a) GRO-seq read densities centered at the point of bidirectional transcription for p53BERs (upper panel) and p53FERs (lower panel). **(b)** Boxplot showing publically available p53 ChIP-seq coverage for p53BERs, p53FERs and uninduced enhancers regions. P-values are for the 2-sided Mann-Whitney-U test. **(c)** Top 15 transcription factor enrichments. Top to bottom: p53RERs vs all bidirectional enhancers, p53BERs vs. p53RERs, p53BERs vs. all bidirectional, p53FERs vs. p53RERs, nutlin-3a-insensitive vs. all bidirectional (as negative control). Note that in the last two panels the significance is indicated by uncorrected p-value, while in the other panels the FDR is used. **(d)** Genome browser representation of STAT3, FOS and BCL3 activation upon nutlin-3a treatment in MCF-7 cells, shown by both RNA-seq and GRO-seq. **(e)** Relative mRNA levels of STAT3, FOS and BCL3 upon nutlin-3a (8 µM) treatment in MCF-7 cells. Values were determined by qRT-PCR (Mean ± SD are shown). **(f)** Venn diagram showing the number of p53-responsive enhancers (from the total of 4522, see Fig. 1c in the main text) that have binding of transcription factors BCL3, FOS, STAT3 and p53. The binding positions of BCL3, FOS and STAT3, as determined by ChIP-seq, were obtained from ENCODE data (wgEncodeReg TFBS Clustered V3).

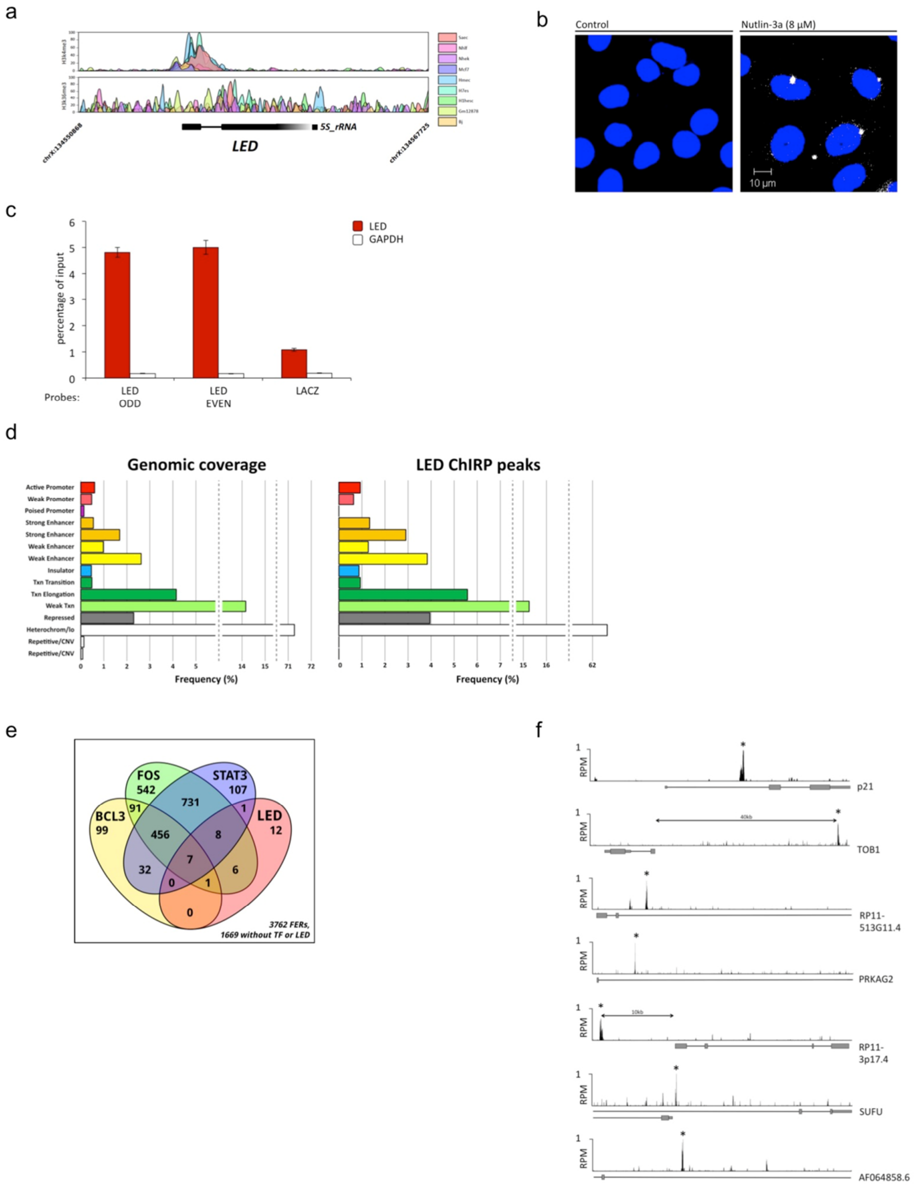


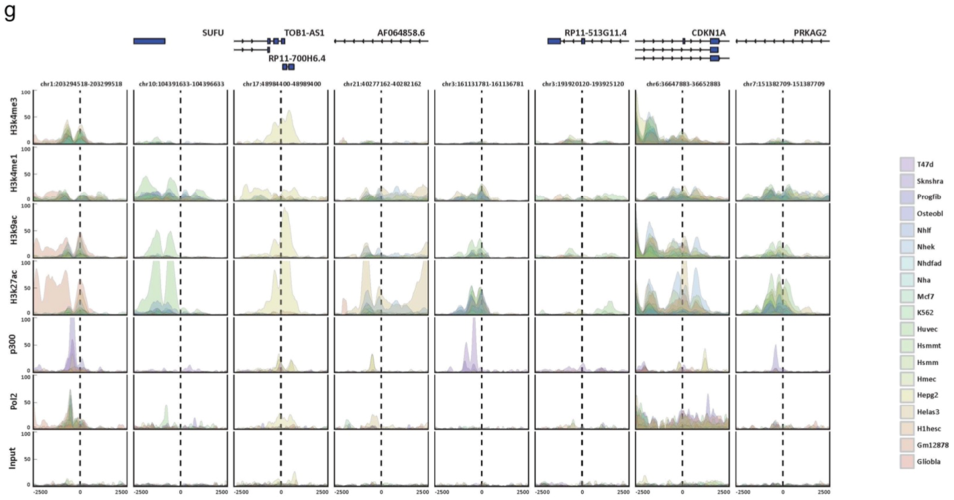
Supplementary Figure 2. Novel stress-regulated lncRNA LED.

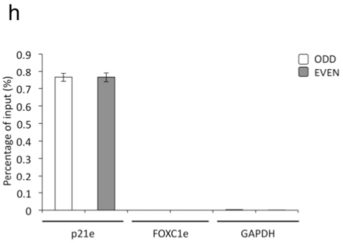
(a) LED expression levels in MALME-3M, ZR-75-1 and MCF-7 cells upon 12h nutlin-3a treatment (8 µM). Values were determined by qRT-PCR (n=3; p-values were calculated with a two-tailed student's t-test. **p<0.01, ***p<0.005). **(b)** P53 or IgG ChIP in MCF-7 cells treated with or without nutlin-3a (8 µM) and RITA (1µM) treatments. The binding of p53 within LED gene body is displayed. **(c)** Schematic representation of cell proliferation assay in MCF-7 treated with 2 pulses of nutlin-3a and transfected with a control or LED siRNA. (n=3; p-values were calculated with a two-tailed student's t-test. ***p<0.005). **(d)** Colony formation assay using the same condition as in (c).

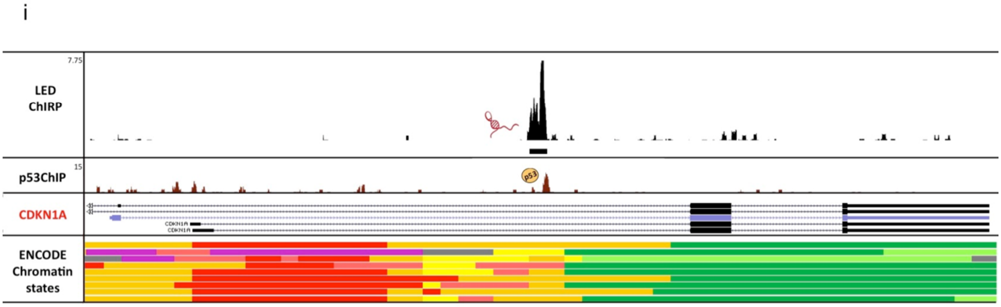


Supplementary Figure 3. The lncRNA LED influences p21 levels.
LED and p21 expression levels in ZR-75-1 and MALME-3M cell lines upon transfection of a control (Ctrl) or LED siRNA (LED-kd). The cells were treated 12h with nutlin-3a (8 μM) and expression levels were measured by qRT-PCR (Mean ± SD are shown).

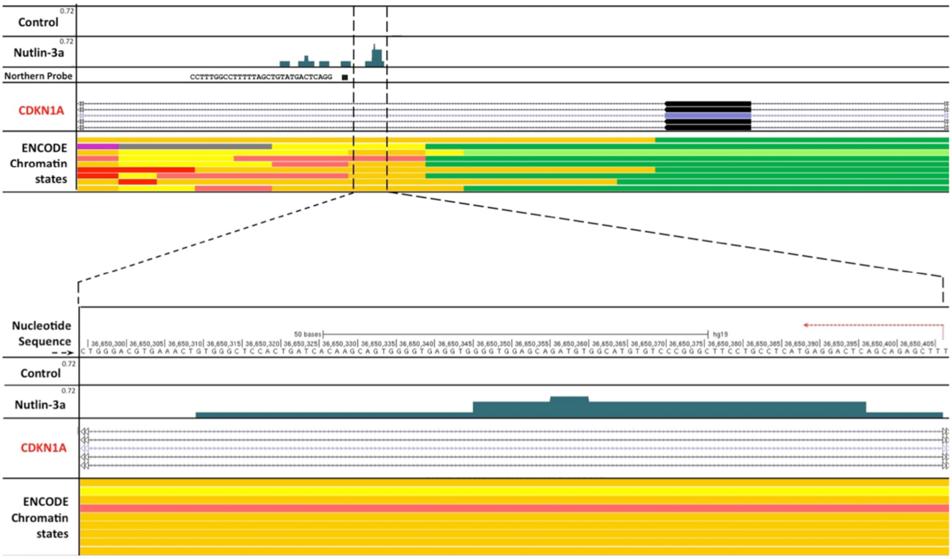

a
H3k4me3
H3k36me3
Saec
Nhlf
Nhek
Mcf7
Hmec
H7es
H1hesc
Gm12878
Bj
chrX:134550868
chrX:134567725
5S_rRNA
LED
b
Control
Nutlin-3a (8 μM)
10 μm
c
percentage of input
LED
GAPDH
Probes:
LED ODD
LED EVEN
LACZ
d
Genomic coverage
LED ChIRP peaks
Active Promoter
Weak Promoter
Poised Promoter
Strong Enhancer
Strong Enhancer
Weak Enhancer
Weak Enhancer
Insulator
Txn Transition
Txn Elongation
Weak Txn
Repressed
Heterochrom/lo
Repetitive/CNV
Repetitive/CNV
Frequency (%)
Frequency (%)
e
FOS 542
STAT3 107
BCL3 99
LED 12
731
91
1
456
8
7
32
0
1
6
0
3762 FERs, 1669 without TF or LED
f
RPM
p21
40kb
TOB1
RP11-513G11.4
PRKAG2
10kb
RP11-3p17.4
SUFU
AF064858.6


g
SUFU
TOB1-AS1
RP11-700H6.4
AF064858.6
RP11-513G11.4
CDKN1A
PRKAG2
chr1:203294518-203299518
chr10:104391633-104396633
chr17:48984400-48989400
chr21:40277162-40282162
chr3:161131781-161136781
chr3:193920120-193925120
chr6:36647883-36652883
chr7:151382709-151387709
H3k4me3
H3k4me1
H3k9ac
H3k27ac
p300
Pol2
Input
100
50
0
-2500
0
2500
T47d
Sknshra
Progfib
Osteobl
Nhlf
Nhek
Nhdfad
Nha
Mcf7
K562
Huvec
Hsmmt
Hsmm
Hmec
Hepg2
Helas3
H1hesc
Gm12878
Gliobla

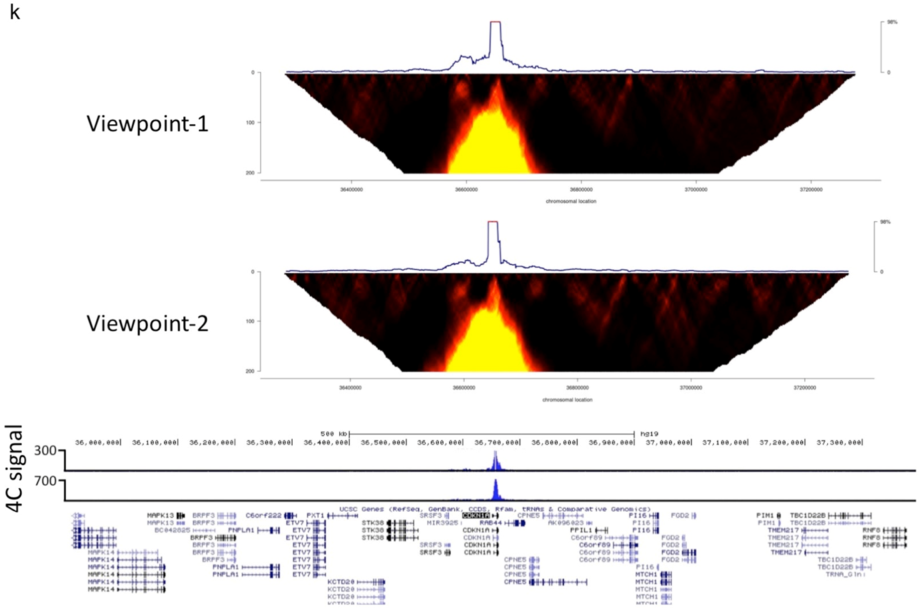

h
ODD
EVEN
Percentage of input (%)
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0
p21e
FOXC1e
GAPDH

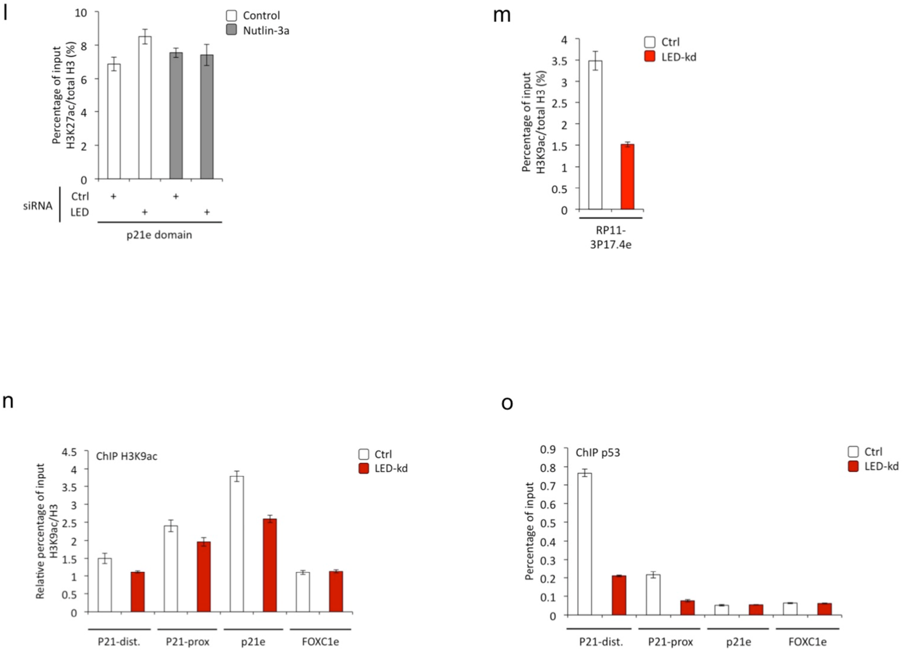

i
LED ChIRP
7.75
p53ChIP
15
p53
CDKN1A
ENCODE Chromatin states

j


Control
0.72
Nutlin-3a
0.72
Northern Probe
CCTTTGGCCTTTTTAGCTGTATGACTCAGG
CDKN1A
ENCODE Chromatin states
Nucleotide Sequence
50 bases
hg19
36,650,300 36,650,305 36,650,310 36,650,315 36,650,320 36,650,325 36,650,330 36,650,335 36,650,340 36,650,345 36,650,350 36,650,355 36,650,360 36,650,365 36,650,370 36,650,375 36,650,380 36,650,385 36,650,390 36,650,395 36,650,400 36,650,405
CTGGGACGTGAAACTGTGGGCTCCACTGATCACAAGCAGTGGGGTGAGGTGGGGTGGAGCAGATGTGGCATGTGTCCCGGGCTTCCTGCCTCATGAGGACTCAGCAGAGCTTT
Control
0.72
Nutlin-3a
0.72
CDKN1A
ENCODE Chromatin states

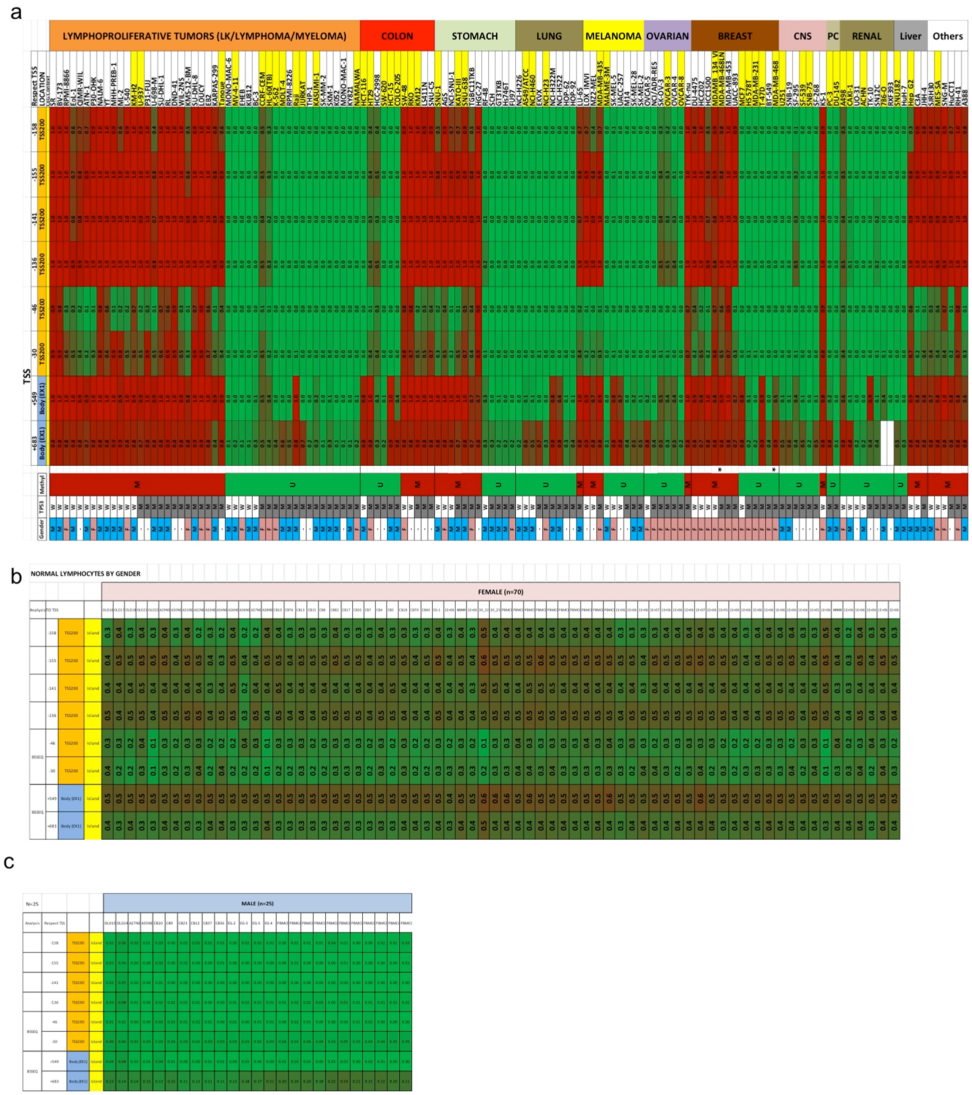
k

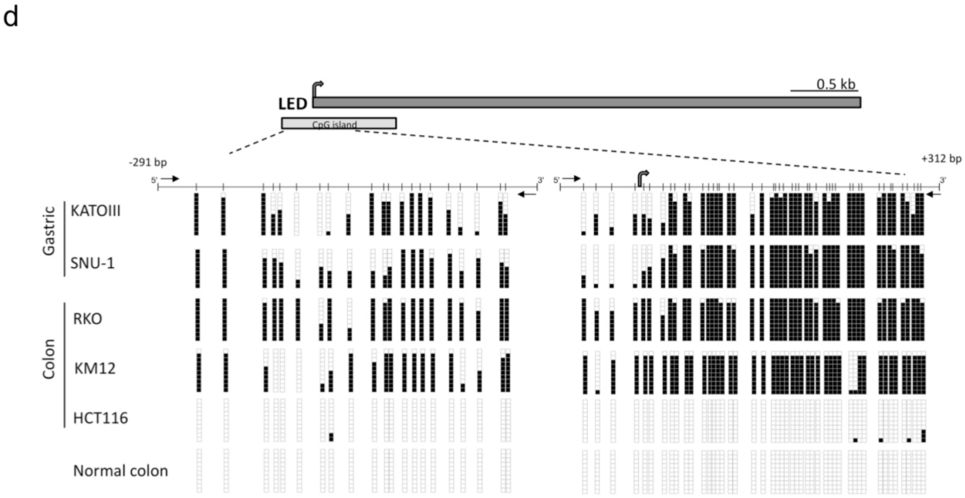

Viewpoint-1
Viewpoint-2
0
100
200
98%
0
36400000
36600000
36800000
37000000
37200000
chromosomal location
4C signal
300
700
500 kb
hg19
36,000,000 36,100,000 36,200,000 36,300,000 36,400,000 36,500,000 36,600,000 36,700,000 36,800,000 36,900,000 37,000,000 37,100,000 37,200,000 37,300,000
UCSC Genes (RefSeq, GenBank, CCDS, Rfam, tRNAs & Comparative Genomics)
MAPK13
MAPK14
BRPF3
BC042825
C6orf222
PNPLA1
PXT1
ETV7
KCTD20
STK38
SRSF3
MIR3925
CDKN1A
RAB44
CPNE5
AK096023
PPIL1
C6orf89
PI16
MTCH1
FGD2
PIM1
TBC1D22B
TMEM217
RNF8
TRNA_Gln

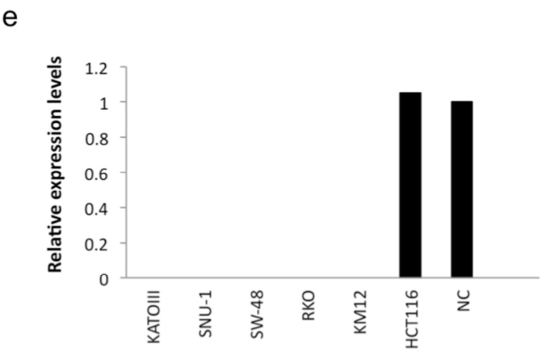


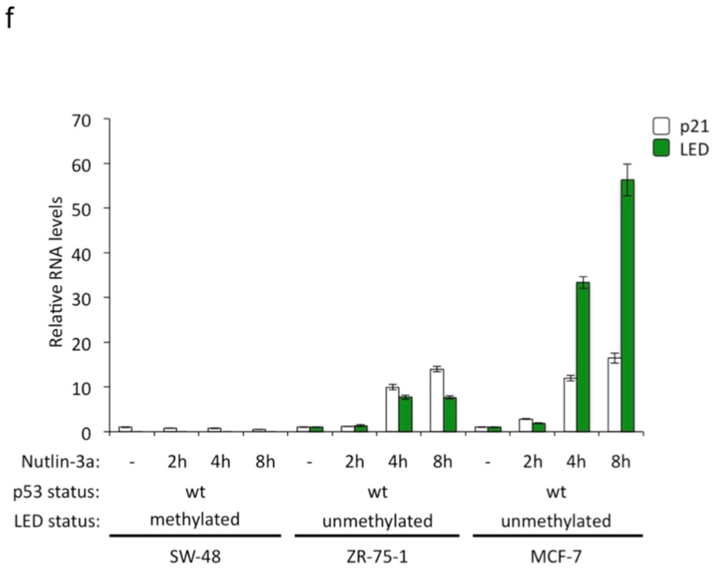
Supplementary Figure 4. LED binds preferentially to enhancers and regulates enhancer RNA production from p53RERs.

(a) H3K4 trimethylation and H3K36 trimethylation (K4K36 signature) at LED gene locus for several cell lines, as determined by the ENCODE consortium. **(b)** RNA Fluorescence in situ hybridization (FISH), with probes against LED in MCF-7 cells incubated with or without nutlin-3a. **(c)** ChIRP enrichment for LED RNA using LED-ODD, LED-EVEN and LACZ probes. GAPDH was used as a control (mean ± SD). **(d)** Distribution of the ENCODE features over the genome (left panel) and LED ChIRP peaks in the different ENCODE features (right panel). **(e)** Intersection of LED-bound p53FERs with BCL3, FOS and STAT3 transcription factors. **(f)** ChIRP sequencing data, showing LED association to a subset of enhancer domains. Values are represented by RPM (reads per million). Star (*) indicates the significant binding site. **(g)** Levels of several histone marks (H3K27Ac, H3K4me1, H3K4me3 and H3K9Ac), p300 and RNAPII in different cell lines in the loci (+/- 5kb) of selected LED-induced enhancers, as determined by the ENCODE consortium. **(h)** DNA ChIRP-qPCR enrichment for p21e, but not for FOXC1e or GAPDH, using ODD and EVEN LED-tiling probes in MCF-7 cells. Values are represented as percentage of input. Mean ± SD are shown. **(i)** Genome browser depicting the overlapping LED-binding site (ChIRP) with a p53BS (ChIP) at p21 intronic enhancer (p21e). **(j)** Genome browser session showing p21 enhancer domain (p21e) and the putative p21 enhancer RNA transcription start site (red arrow). Control and nutlin-3a-treated MCF-7 cells GRO-seq tracks are shown as well as the "+" strand nucleotide sequence. **(k)** Domainograms (de Wit et al., 2008) visualizing significance of interactions at different window sizes (viewpoints) for p21 enhancer domain with surrounding chromosomal regions in MCF-7 cells. Color ranges (see scale bar) reflect different levels of significance, from black (low significance, $P=1x10^{-2}$) to yellow (high significance, $P=1x10^{-8}$). To account for the fact that the majority of the data are very close to the viewpoint, we set the data range of the vertical axis to the 98% quantile value for the analyzed region. Values are represented by intensity of 4C signal. Schematic representation of p21 gene with the location of the bait used for the 4C experiments is also shown. **(l)** ChIP-qPCR for total H3 and H3K27Ac at p21 enhancer domain (p21e) in control and nutlin-3a-treated MCF-7 cells transfected with a control or LED siRNAs (LED-kd). Values represent the percentage of input and were normalized to total H3. Mean ± SD are shown. **(m)** ChIP-qPCR for total H3 and H3K9Ac at RP11-3P17.4e in nutlin-3a-treated MCF-7 cells transfected with a control or LED siRNA (LED-kd). Values represent the percentage of input and were normalized to total H3. Mean ± SD are shown. **(n)** ChIP-qPCR for total H3 and H3K9Ac at p21 distal (dist.), proximal (prox.) and intronic (p21e) enhancers in nutlin-3a-treated MCF-7 cells transfected with a control or LED siRNA (LED-kd). Values represent the percentage of input and were normalized to total H3. Mean ± SD are shown. **(o)** ChIP-qPCR for p53 at p21 distal (dist.), proximal (prox.) and intronic (p21e) enhancers in nutlin-3atreated MCF-7 cells transfected with a control or LED siRNA (LED-kd). Values represent the percentage of input. Mean ± SD are shown.




d
LED
0.5 kb
CpG island
-291 bp
+312 bp
5'
3'
Gastric
KATOIII
SNU-1
Colon
RKO
KM12
HCT116
Normal colon


e
Relative expression levels
1.2
1
0.8
0.6
0.4
0.2
0
KATOIII
SNU-1
SW-48
RKO
KM12
HCT116
NC


f
Relative RNA levels
70
60
50
40
30
20
10
0
p21
LED
Nutlin-3a:
-
2h
4h
8h
p53 status:
wt
LED status:
methylated
unmethylated
unmethylated
SW-48
ZR-75-1
MCF-7

Supplementary Figure 5. DNA methylation-associated silencing of LED in colon and gastric cancers.

(a) DNA methylation-associated silencing of LED in several cancer cell lines. Methylation was measured in different LED regions represented by TSS (transcription start site: -158, -155, -141, -136, -46, -30, +549, +683). Presence of unmethylated or methylated CpGs is indicated by green or red, respectively. Methylation is considered when both CpGs at (-155, -141, -136) and at (+549,+683) are methylated or when all the CpGs are methylated. Indication of TP53 status (W-wild type ; M-mutant) and gender (M-male ; F- female) are represented. **(b, c)** LED methylation in normal lymphocytes by gender- female (b) and male (c) and analysed as in (a). **(d)** Schematic representation of LED genomic loci and CpG island. Bisulfite genomic sequencing analysis of LED CpG island in human solid cancer cell lines and normal colon as tissue control. Location of bisulfite genomic sequencing PCR primers (black arrows), CpG dinucleotides (vertical lines) and the transcriptional start site (grey arrow) are shown. Ten single clones are represented for each sample. Presence of unmethylated or methylated CpGs is indicated by white or black squares, respectively. **(e)** LED expression levels in methylated or unmethylated human solid cancer cell lines and in normal colon (NC) as control. Values were determined by qRT-PCR in triplicates. **(f)** LED and p21 expression levels in SW-48, ZR-75-1 and MCF-7 cells upon 2h, 4h and 8h of nutlin- 3a treatment (8 μM). Values were determined by qRT-PCR and normalized to the untreated amples. Mean ± SD are shown.

NB: Supplementary tables have been omitted for brevity and can be found online, using identifier doi:10.1038/ncomms7520

CHAPTER 8

Discussion of Part B – The non-coding genome

The discovery that the human is pervasively transcribed (Djebali et al., 2012), with over 60% of the human genome being expressed as RNA, while only 2% of the genome is protein-coding poses a paradox. It breaks the dogma of the protein-coding genes being the primary product of the genome and its transcription. P53 has been characterized as important transcription factor, directly inducing expression of hundreds of target protein-coding genes. On the other hand, little has been known about its role in the induction of non-coding genes and enhancers. Our work identifies several hundred enhancers directly activated by P53 by combining ChIP-seq and GROseq data. Furthermore, we found thousands of enhancers activated that were not found to be P53-bound. These may constitute secondary targets. Furthermore, using transcriptome profiling by RNAseq we identified about 200 lncRNAs induced by activation of P53. We focused our attention on LED, a lncRNA strongly induced upon P53 activation and important for the full activation of cell-cycle arrest normally induced by activation of P53, as indicated by FACS profiles upon the knockdown of LED.

We found LED to be localized predominantly in the nucleus, and knockdown of LED and subsequent transcriptome profiling by RNAseq showed a pattern of differentially expressed genes further indicating a role of LED in control of cell cycle. A prominent gene that was down-regulated upon LED knockdown was p21, a master cell cycle regulator and important direct target of P53. Interestingly, using ChIRP to identify the localization of LED in the chromatin we identified a binding of LED in the gene of p21. Specifically, LED binds an enhancer region within the p21 gene. We showed that this enhancer was a bona fide enhancer, able to activate transcription in an orientation-independent manner. Furthermore, the enhancers' activity was sensitive to knockdown of LED.

Our findings provide an interesting example of how the different levels of regulation and non-coding and coding genes are integrated: a transcription factor activating a lncRNA, which in turn activates an enhancer which in turn is important for the full activation of a protein-coding gene. As technological advances are made, we get tools to investigate these levels of regulation and transcription more sensitively. It seems therefore likely that more examples where proteins, mRNAs, lncRNAs and enhancers integrate will be found.

Our use of state-of-the-art techniques such as GROseq, ChIRP, RNAseq and ChIPseq allowed us to get a high-quality view with unprecedented sensitivity at the transcription induced by P53. But this does not mean the picture of P53's transcriptional landscape is now complete. Instead, improvements to technologies, experimental and computational procedures will continue to incrementally improve our understanding of the targets of P53. An exciting new technique is CRISPR-Cas9, which allows high-throughput genome editing (Sander and Joung, 2014). While current high-throughput techniques mostly focus on detection of transcription (such as GROseq, RNAseq), CRISPR-Cas9 allows functional interrogation of genes or loci by modulation of the transcription or gene product. Combining a system to systematically modulate transcription with identified transcripts that are differentially expressed upon P53 activation or with loci that are found to be bound by P53 allows the identification of novel functional P53-regulated elements, in a manner that is independent of the transcriptional activity of these elements.

References

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.

Sander, J.D., and Joung, J.K. (2014). CRISPR-Cas systems for editing, regulating and targeting genomes. Nat Biotech *32*, 347–355.

Summary

The genetic code of a cell is kept in its DNA. However, a vast number of functions of a cell are carried out by proteins. Through gene expression the genetic code can be expressed and give rise to proteins. The expression of genes into proteins follows two steps: transcription of DNA into RNA and translation of (messenger-)RNA into proteins. This thesis concerns these two steps in gene expression, which are very different from a molecular point-of-view.

Part A – Protein translation studied by ribosome profiling

Protein coding genes give rise to proteins via an intermediate messenger molecule: messenger-RNA (mRNA). The mRNA is translated into proteins by ribosomes. Recent advances in technology have made it possible to take a snapshot of the positions of ribosomes on mRNA molecules, by a protocol called "ribosome profiling". It is comprehensive and high-throughput: a single snapshot reveals the positions of millions of ribosomes on the mRNA in a sample. Ribosome profiling has paved the way to investigate regulation at the level of mRNA translation in a global manner (i.e. measuring a large part of the expressed genes simultaneously). In this work, we extend the purpose of ribosome profiling by making use of the positional data (i.e. the location of ribosomes within the mRNAs) that is obtained. First we show the impact of a mitochondrial transfer-RNA (tRNA) mutation on ribosome progression. tRNAs form essential components of the translational machinery by delivering amino acids (the protein building blocks for proteins) to the ribosomes. While ribosomes seem to be able to use the tRNA in question for translation, as indicated by the presence of full-length protein, they spend more time waiting for the tRNA, suggestive of stalling. We also investigate the impact of other factors that are required for activation of tRNAs, such as the levels of amino acids. We show that under certain conditions ribosome profiling can be used to identify amino acid levels restrictive for growth of cancer cells. Furthermore, using tumor samples we identify a ccRCC kidney tumor with ribosomal stalling patterns suggestive of proline deficiency. From this tumor data, we were able to untangle some of the metabolic rewiring upon proline deficiencies, and identify a molecular target (the gene PYCR1). We further used ribosomal profiling to investigate the regulation of translation during mitosis, the part of the cell cycle in which the (by then duplicated) DNA content is separated into two equal parts. Interestingly, ribosome profiling data shows that ribosomes undergo strong major changes in their mRNA occupancies. First, ribosomes tend to occupy mRNAs more on the 5'-end. Second, ribosomes show a differential tendency to occupy certain codons in the mRNAs.

Because ribosome profiling has been pivotal in the work in this part of the thesis, we have made improvements to the experimental part of the protocol, and increased data quality. For example in the reduction of ribosomal RNA (rRNA) contamination, an unfortunately

pervasive and abundant contaminant in the data, we have made substantial improvements. Using subtractive hybridization we were able to strongly reduce the amount of contamination in the raw data, and this proved essential when working with e.g. live tissue samples.

Part B – Transcription of the non-coding genome

Proteins are an important product of gene expression. Yet, most of the genetic code does not encode for proteins. As protein-coding genes come to expression as mRNA to be further translated into protein (the focus of part A), it is puzzling that a large fraction of the genome (approx. 60%) comes to expression as RNA but does not serve as template to be translated into protein. The function of this expression is largely unclear. Part B of this thesis investigates these “non-coding” genes and RNAs in the context of cancer. Transcription factor P53 is of paramount importance in the stress-induced arrest of the cell cycle. Using ChIP-seq we investigate which regions of the genome are bound by P53, and using RNA-seq and GRO-seq we identify which genes are differentially expressed upon activation of P53. Using these techniques we unravel novel parts of the transcriptional program of P53 that lead to expression of non-coding genes. We show that the major cell cycle regulator p21 which is induced by P53 requires the activation of one such non-coding gene.

Samenvatting

De genetische code van een cel ligt vast in het DNA. Toch worden een groot aantal functies van een cel uitgevoerd door eiwitten. De genetische code wordt via gen-expressie omgezet tot eiwitten. Die expressie volgt twee stappen: transcriptie van DNA in RNA, en translatie van (messenger-)RNA in eiwit. Deze dissertatie betreft deze twee stappen van gen-expressie, al zijn ze zeer verschillend vanuit moleculair oogpunt.

Deel A – Translatie van mRNA onderzocht met behulp van ribosome profiling
Eiwit-coderende genen komen tot expressie als eiwit via een tussenstap: transcriptie van een gen in messenger-RNA (mRNA). Het mRNA wordt in eiwit vertaald door ribosomen. Recent werk en technologische vooruitgang hebben het mogelijk gemaakt om een momentopname te maken van de posities van ribosomen op het mRNA, met een techniek die “ribosome profiling” wordt genoemd. Het is een veelomvattende techniek: in een momentopname van een monster is het mogelijk de posities van miljoenen ribosomen op mRNAs tegelijk te registreren. Ribosome profiling heeft het mogelijk gemaakt om translatie van mRNA tot eiwit op een globale manier waar te nemen (d.w.z. de translatie van een groot deel van de mRNAs kan in één keer in kaart worden gebracht). In dit deel van de dissertatie vergroten we de mogelijkheden van ribosome profiling, door analyses uit te voeren van de posities die de ribosomen op de mRNAs innemen. Eerst kijken we naar de impact van een mutatie in een mitochondriaal transfer-RNA (tRNA, een component van de moleculaire machines die mRNA vertalen, en verantwoordelijk is voor de aanvoer van aminozuren, de bouwstenen van eiwitten). We zien dat hoewel de ribosomen het tRNA nog wel kunnen gebruiken voor eiwitproductie, ze meer tijd besteden op het codon wat vertaald wordt door het tRNA. We kijken verder naar de impact van factoren die de hoeveelheid beschikbaar tRNA veranderen, zoals aminozuurniveaus, op translatie door ribosomen en hun positionering op mRNAs. We vonden dat onder bepaalde omstandigheden cellulaire aminozuurtekorten op te sporen zijn met behulp van ribosome profiling en adequate analyses. Door analyse van niertumorsamples vonden we een tumor waar de ribosomale posititiies een patroon lieten zien wat duidde op tekort aan proline. Met de data van deze tumor hebben we een moleculaire schakel kunnen vinden die mogelijk van therapeutisch belang is (het gen PYCR1). We hebben ribosome profiling verder gebruikt om te kijken naar regulatie van eiwittranslatie tijdens de mitose – het deel van de celcyclus waar het DNA (wat op dat moment al vermenigvuldigd is) opgesplitst wordt in twee gelijke delen. Curieus genoeg laat ribosome profiling data zien dat tijdens de mitose ribosomen sterke algemene veranderingen in hun positionering ondergaan. Ten eerste komen ze dichter bij het 5'-einde te zitten (het “begin” van het mRNA molecuul), en ten tweede zijn ribosomen in andere verhoudingen aanwezig op bepaalde codons van het mRNA.

Omdat ribosome profiling zo centraal staat in dit werk, hebben we gaandeweg verbeteringen gemaakt aan het protocol, om zo de kwaliteit van de data te verbeteren. Bijvoorbeeld door de hoeveelheid ribosomaal RNA (rRNA, een onfortuinlijke en zeer sterk aanwezige contaminatie) te reduceren. Met behulp van een hybridisatietechniek zijn we succesvol geweest in het terugbrengen van de fractie van rRNA in onze monsters, hetgeen de gevoeligheid van de analyses heeft bevorderd. Mede hierdoor hebben we bruikbare data uit bijvoorbeeld vers-weefsel monsters kunnen krijgen.

Deel B – Transcriptie van het niet-coderende genoom

Eiwitten vormen een belangrijk product van gen-expressie, maar slechts een klein deel van het genoom bepaalt ook echt de volgorde van aminozuren in een eiwit. Eiwit-coderende genen komen tot expressie als mRNA om verder te worden vertaald in eiwit door ribosomen (de focus van deel A). Maar daarnaast komt een groot deel van het genoom (ca. 60%) tot expressie als RNA, om vervolgens niet verder vertaald te worden tot eiwit. De functie hiervan is voor een groot deel nog onduidelijk. Deel B van deze dissertatie betreft deze niet-eiwitcoderende genen en RNA moleculen in de context van kanker. Transcriptiefactor P53 is van monumentaal belang bij het remmen van de celcyclus in het geval van DNA-schade of cellulaire stress. Met behulp van moderne technieken als ChIP-seq onderzoeken we waar in het DNA de transcriptiefactor P53 bindt. Verder gebruiken we RNA-seq en GRO-seq om te onderzoeken welke genen meer of minder tot expressie komen wanneer P53 wordt geactiveerd. Met behulp van deze technieken kunnen we een beter beeld schetsen van het niet-eiwitcoderende gen-regulatieprogramma van P53. We laten zien dat een van deze niet-eiwitcoderende genen benodigd is voor activatie van p21, een van de belangrijkste genen die door P53 wordt geactiveerd om de celcyclus te remmen.

Curriculum vitae

Koos Rooijers was born October 7th, 1986 in Haarlem, the Netherlands. After leaving the "Eerste Christelijk Lyceum Haarlem" with a high-school diploma in 2005, he followed a bachelor in Life Science and Technology, organized and hosted jointly by Leiden University and Delft University of Technology. After graduating *cum laude* in 2008, he continued with a double master's study at Wageningen University, studying Food Technology as well as Bioinformatics, but discontinued the former. During this time he joined prof. Janusz Bujnicki at the IIMCB in Warsaw for a project on the computational modeling of protein structure. Upon return to Wageningen he wrote a thesis and his first peer-reviewed publication under the supervision of dr. Peter Schaap on the subject of metagenomics and metaproteomics. Shortly after *cum laude* graduation from the master's, Koos continued his academic career at the NKI in the group of prof. dr. Reuven Agami, where the work described in this book was performed.

List of publications

(First author or co-first author)
Rooijers, K., Loayza-Puch, F., Nijtmans, L.G., and Agami, R. (2013). Ribosome profiling reveals features of normal and disease-associated mitochondrial translation. Nat Commun *4*.

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*.

Rooijers, K., Kolmeder, C., Juste, C., Dore, J., de Been, M., Boeren, S., Galan, P., Beauvallet, C., de Vos, W.M., and Schaap, P.J. (2011). An iterative workflow for mining the human intestinal metaproteome. BMC Genomics *12*.

(co-author)
Morris, A.R., Bos, A., Diosdado, B., **Rooijers, K**., Elkon, R., Bolijn, A.S., Carvalho, B., Meijer, G.A., and Agami, R. (2012). Alternative cleavage and polyadenylation during colorectal cancer development. Clin Cancer Res *18*, 5256–5266.

Jenal, M., Elkon, R., Loayza-Puch, F., van Haaften, G., Kuhn, U., Menzies, F.M., Oude Vrielink, J.A.F., Bos, A.J., Drost, J., **Rooijers, K**., et al. (2012). The poly(A)-binding protein nuclear 1 suppresses alternative cleavage and polyadenylation sites. Cell *149*, 538–553.

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

Baglio, S.R., **Rooijers, K**., Koppers-Lalic, D., Verweij, F.J., Perez Lanzon, M., Zini, N., Naaijkens, B., Perut, F., Niessen, H.W.M., Baldini, N., et al. (2015). Human bone marrow- and adipose-mesenchymal stem cells secrete exosomes enriched in distinctive miRNA and tRNA species. Stem Cell Res Ther *6*.

Inacio, P., Zuzarte-Luis, V., Ruivo, M.T., Falkard, B., Nagaraj, N., **Rooijers, K**., Mann, M., Mair, G., Fidock, D.A., and Mota, M.M. (2015). Parasite-induced ER stress response in hepatocytes facilitates Plasmodium liver stage infection. EMBO Rep *16*, 955–964.

Konovalova, S., Hilander, T., Loayza-Puch, F., **Rooijers, K**., Agami, R., and Tyynismaa, H. (2015). Exposure to arginine analog canavanine induces aberrant mitochondrial translation products, mitoribosome stalling, and instability of the mitochondrial proteome. Int J Biochem Cell Biol *65*, 268–274.

Acknowledgements

The work described in this book is mostly the product of fruitful collaborations. It should be clear that I have many people to thank for being next to me on the path I had set out for myself some five years ago.

Reuven, I was honored you drafted me in your little legion and working with you has been a tremendous adventure. I have greatly benefited from your expertise: seeing the opportunities for ribosome profiling to measure mitochondrial translation and detect (differences in) metabolic state turned out to be extremely potent ideas, way beyond my imagination. I joined your group having encountered the borders of (molecular biological) knowledge, and you showed me how to surpass and push out these borders… to boldly go where no one has gone before! I've also appreciated the group of inspiring and varied people you've recruited, and all the unconventional (and at times crazy) experiments you let them do, which yielded wonderful (mostly sequencing) data for me to play with.

Which brings me to the group members. First I'd like to thank my (recent) office mates Nicolas and Carlos! After so many collaborations we finally succeeded to have a fruitful one! Anyway, I learned a lot from the failed ones as well, mostly because of our extensive contact. Hearing each other's hypotheses, planning experiments and discussing analyses, it was a great pleasure and greatly educational. Fabricio, your tireless efforts have been of paramount importance to most of the work described in this book. I've been lucky to have had you as close colleague, with your 'can do' attitude and breath of expertise. Alejandro, you've picked up learning bioinformatics and that would already give you a special place in my heart. But instead you deserve it because you always know the right questions, and are not afraid to ask them. I hope you continue on this path (well,... as long as you think it's fun), because man, you are going places :)

Gozde and Rui, thank you for all the great time we spent together, lunch, parties, … Thank you a million times for letting me moan about work&colleagues (no, not you dear reader, don't worry…). Gozde, thank you for your sugary *pick-me-up*s. Rui, thanks for being such an amazing friend (although your expertise with regards to wine is somewhat intimidating. I mean, what can I bring to your place without offending you? I hope the large amounts of ice-cream we ate were pleasing nevertheless…). Furthermore, I had loads of fun contemplating experiments and opportunities for collaboration, even though it never came to that. Boris, Ruiqi, Li, Vittorio, Pieter, Arieh and all other group members: thanks for bearing with me whenever I gave (seemingly) endless talks during work discussion, and whenever I had (seemingly) endless comments during others' talks. I hope I made the bioinformatics/statistics/data analysis work accessible for everyone… I certainly tried my best.

Also I'd like to thank the people from B4 (van Steensel, van Leeuwen, Dannenberg groups, as well as ex-Agami members). Maybe out of sight, certainly not out of mind. The department meetings we had in the first couple of years during my PhD shaped me and my work, and I am very grateful for the criticisms I have received during

those meetings. Furthermore, I have greatly enjoyed listening to the talks about your projects; I've gained a ton of insight into biology as well as various experimental procedures thanks to your beautiful work.

Bas van Steensel, Jos Jonkers and Lodewyk Wessels: thank you for your yearly feedback as part of my evaluation committee. Your wisdom and suggestions have kept me on course, and helped me finish this book in a reasonable timeframe. Guido Jenster, Marieke von Lindern and Wim Vermeulen, thank you for taking the time to read and judge this book.

I'm grateful to the many collaborators I've worked with over the years: Leo Nijtmans, Lenno Krenning, René Medema, Wilbert Zwart, Metello Innocenti, Svetlana Konovalova, Maria Wesolowska, Robert Lightowlers and Zofia Chrzanowska-Lightowlers. Also, thank you, NKI's own Genomics Core Facility (and Roel Kluin in particular), for letting the data flow freely and plentiful.

Something which I've enjoyed thoroughly at the NKI has been attending the talks from invited speakers. Therefore I'd like to thank the organizers at the NKI, as well as the speakers themselves, for this mind-expanding opportunity. The depth of the talks, the choice of the subjects and the quality of the work presented usually put a smile on my face and made me return to my desk with the feeling I had learned something. Time well spent, methinks.

Thank you Noël Wessels, Lucas van den Berg, Mark Caro and your many talented colleagues for clogging my ears with your awesome beats, giving me the energy to go on working late nights, getting out of bed early and face the grind of the day. And thank you Guus, for introducing me to these artists.

Pap en mam en Guus, ik heb zo veel van mijn huidige succes aan jullie te danken. Er zijn zo veel karaktereigenschappen die ik van jullie heb overgenomen, iets wat ik lang gemist (jeugd) of ontkend/tegengewerkt (puberteit) heb, en waar ik nu zo blij mee ben. Zoals de directe houding van mam. En altijd het onderste steentje boven willen hebben. Guus, al deel jij je inzichten veelal op een directe en ruwe manier, ze wegen voor mij zeer zwaar. Ik heb het grootste ontzag voor hoe jij jouw vrijheden in je leven verworven hebt. Pap, zoals je waarschijnlijk zelf al weet, jouw nieuwsgierige en wetenschappelijke trekjes hebben aanstekelijk op mij ingewerkt. En ik ben ontzettend dankbaar dat jullie dat alle drie hebben gestimuleerd. Daarnaast zijn jullie ook een fantastische ondersteuning geweest en heb ik altijd op jullie kunnen rekenen.

Krystyna, moja kochana. Having you in my life makes me the luckiest person in the world. I fully realize the price for my PhD has been paid by you as well. And not only did you have to accept my limited availability, you actively let me lean on you by helping me practice talks, bounce ideas, and letting me vent. It means the world to me that you are in my life and that you have accepted the path I had set out for myself. I seriously can't imagine having done it without you, but luckily I don't have to. You bring so much happiness in my life that any worries are meaningless in comparison. Thus, I'm looking forward to whatever is next in our lives. I know I can count on us to make it great, whatever and wherever it is.