Istituto Italiano di Tecnologia - Research line in Synthetic and Systems Biology for Biomedicine and Università di Genova - Dipartimento di Neuroscienze e Neurotecnologie

Ph.D. Thesis

Optimal genetic circuit design for mammalian synthetic biology applications

Supervisor:
Dr. Velia Siciliano

External Referees:
Prof. Karen Polizzi
Dr. Grégory Batt

Candidate:
Federica Cella

XXXIII cycle Ph.D. course

April 2021

Contents

List of Abbreviations

RBP: RNA-binding protein; **CS**: cleavage site; **miRNA**: microRNA; **siRNA**: short interfering RNA; **UTR**: untranslated region; **TS**: target site; **GOI**: gene of interest; **TF:** transcription factor; **Pol II**: RNA Polymerase II; **XRN1**: exoribonuclease 1; **DRB**: 5,6-dichloro-1-beta-D-ribobenzimidazole; **RBS**: ribosome binding site; **NT**: non-transfected; **FDR**: false discovery rate; **DEG**: differentially expressed gene; **DEP**: differentially expressed protein; **GO**: gene ontology; **GSEA**: gene set enrichment analysis; **FCS**: functional class scoring; **ORA**: over-representation analysis; **ES**: enrichment score; **NES**: normalized enrichment score.

List of Figures

List of Tables

Abstract

Synthetic biology combines biology and engineering to redesign living organisms by conferring them new properties. Synthetic biologists harness the power of nature to solve problems in medicine, manufacturing and agriculture and use engineering principles to control and perturb cell pathways to gather novel insights into biological functions. Specifically, the programming of new functions into mammalian cells has tremendous application in research and medicine. The rapid advances in the ability to synthesize DNA and RNA have increased our understanding of gene function and regulation and have enabled the expansion of genetic parts available for programming cell biology. Genetic parts can be multiplexed to build complex genetic circuits to *(i)* confer new functions to the host cells, *(ii)* regulate endogenous pathways, or *(iii)* sense inputs of interest to actuate a cellular response. This thesis focuses on the optimization of genetic circuit design for mammalian synthetic biology applications. It covers the need for additional regulatory tools for tight spatiotemporal control of gene expression and cell response and the need for a better understanding of how regulation processes work in cells. Finally, it explores strategies to select biologically relevant inputs for sensor-actuator devices that control cell fate.

Chapter 1

Introduction to Mammalian Synthetic Biology

Synthetic Biology aims at designing genetic networks that can provide insights about mechanisms involved in biological functions, or to achieve new desired functionalities in the targeted cells [1, 2]. The ability to engineer new functions within living organisms has led to a number of applications in industry [3], environmental biotechnology [4] and biomedicine [5]. In particular, advances in mammalian synthetic biology have facilitated the study of diverse biological processes including gene regulation [6], developmental patterns [7], evolution [8], and cancer progression [9]. More recently, its clinical relevance has been improved by the powerful new tools for the engineering of recombinant protein-producing cells [10] and for the creation of novel cell-based therapies for clinical use [11, 12]. In this chapter, I will give an overview of the tremendous potential of synthetic biology and the critical challenges of engineering biology.

1.1 Synthetic biology: engineering living systems

The Nobel Prize winning physicist Richard Feynman once wrote on his blackboard: "What I cannot create I do not understand". He meant that unless you can take a concept or a theory apart, so that you can understand each individual step, you did not truly understand it. Despite the different scientific field, this idea is at the core of synthetic biology [13]. Synthetic biology is hybrid discipline that mix engineering principles with biological questions. The vision of synthetic biology is a systematic, hierarchical design of artificial living systems using standardized biological and synthetic parts [14], with the goal of understanding biological processes or to create new ones [15]. Since the advent of the recombinant DNA technologies [16] and the implementation of Sanger sequencing in the 70's [17], the cost of DNA manipulation and synthesis has consistently dropped, making the creation of synthetic systems cheaper and faster [18].

Initially, researchers focused into hacking bacteria to gain novel insights on cellular functions and metabolic processes. The optimization of DNA delivery technologies encouraged the attempt of engineering bacteria to produce exogenous proteins of industrial interest. Bacteria became machines able to produce bioproducts such as therapeutic proteins [19]. At the same time an increasing interest in engineering computational functions for control of cellular response arose. In 2000, two milestone studies contributed to the birth of modern synthetic biology: *(i)* in the group of Professor Jim Collins at Boston University the first toggle switch was implemented [20]; *(ii)* in the meantime, at Princeton University, Elowitz and Leibler implemented the repressilator, a negative regulatory cascade that exhibits oscillatory behaviour of the output [21]. These early synthetic biology designs showed that regulatory components can be characterized and assembled to bring electronics-inspired behaviors in living systems (for example, memory storage and timekeeping) [22]. Since then, circuits have grown in complexity and the techniques to engineer living systems have become more robust and easier to perform [23, 24].

The increasing advances in our ability to manipulate living systems and finely tune gene expression have opened up the way to biomedical applications and gene therapy. To this end, the need to maximize the safety of the products and to find a realistic model for human tissues, has prompted the scientists to focus on more complex systems such as

eukaryotic cells.

1.1.1 Mammalian synthetic biology

In the past years, an increasing number of research groups, as well as several industries, have started harnessing synthetic biology to produce novel therapeutics and implement diagnostics devices, using mammalian cell systems. Mammalian cells naturally produce molecules homologous to the ones in our organism. The post-translational modifications such as glycosylation are naturally done by the cells without any further requirement of product optimization. Moreover, our ability to engineer smart interfaces that sense intracellular and extracellular changes has boosted the diagnostic capabilities [25].

At the same time, a number of groups are turning to mammalian synthetic biology as a tool to probe basic questions in biology [26]. In fact, mammalian gene networks are highly complex and rich in redundant pathways that overlap with each other to secure a reliable and robust machinery. The regulation patterns at transcriptional, post-transcriptional and translational level are finely modulated by a great number of factors, which cooperate in a synergistic manner. Synthetic biology has the potential to become the key to answer questions on aspects of regulation at systems level that really were unattainable a decade or two ago.

One of the most recent branches of mammalian synthetic biology is synthetic immunology [27]. In 2018, Professor James Patrick Allison and Professor Tasuku Honjo were awarded the Nobel Prize in Physiology or Medicine for their discovery of cancer therapy by inhibition of negative immune regulation exploited by engineering patient-derived CD8+ immune cells engineered with chimeric antigen receptors (CAR-T cells) that potentiate the immune system against patient-specific cancer [28]. CAR-T cells are a landmark in cell therapy and a great example of how mammalian synthetic biology can be applied to medicine.

The tremendous achievements of synthetic biology reached in such a short time have been possible thanks to the rigorous methodology and extensive characterization of the tools. The bottom-up approach of engineering living systems has enabled the sequential characterization of simple modules to complex networks, being able to predict in advance their behavior.

1.2 A bottom-up approach for circuit predictability

Synthetic biology focuses on achieving standardization [29], modularity [30], and mathematically predictable behavior in engineered biological systems [31]. In order to achieve these tasks, it is critical to precisely characterise the steps involved in gene expression, such as the processes of transcription and translation. This would eventually result in the construction of appropriate parts (e.g. a promoter that works at a certain transcription rate, or a protein with a given degradation rate) as well as in a more realistic prediction of how they will behave when delivered in the host organism. Earlier studies were focused on the construction of systems in simple prokaryotes, in which the qualitative notions of transcriptional activation, repression and post-transcriptional regulation could be quantified; these constructs were based on the combinatorial promoter libraries driving the expression of reporter genes [32].

Later on, an increasing interest in post-transcriptional and post-translational regulation triggered the design of synthetic circuits harboring elements from the RNA interference pathway, aptamers, riboswitches, or mutated ribosome binding sites [33–36]. All these basic elements were manipulated so that they could work synergistically toward the desired goal, forming small modules including switches, pulse generators, time-delayed circuits, oscillators and so on [20, 21, 37–40].

The rigorous characterization of the single parts' behaviour to later assemble them in more complex networks is called bottom-up approach [41].

1.2.1 Complex networks and regulatory tools

The rigorous step-by-step implementation of circuits design starting from simple well-characterized modules, in principle allows the construction of scaled-up systems with predictable behavior. Moreover, these systems can be modified in their single parts while retaining the overall behavior [42]. To this end, computational model-aided design based on gene modularity has been largely used [43].

Complex genetic circuits are formed by combining biological parts which are able to sense and respond to defined stimuli in a predictable, tunable manner [44]. The interaction of several modular parts enables logical computation to trigger cellular response upon

the input signal, similar to the processing found in electronic devices. The computation performed by synthetic circuits is based on logic gates. Logic gates are a concept borrowed from electrical engineering and applied to biology to introduce a logic by which molecular and environmental signals are processed. For example, if two molecules A and B are both needed to activate cellular response, this form an AND gate. Similarly, OR, NOT, XOR, NAND, etc gates can be engineered to regulate cellular response [45]. The regulatory interactions and the implementation of logic gates between modules in the systems are possible thanks to a wide toolbox of regulatory devices to control gene expression.

To control exogenous and endogenous gene expression at the DNA level, the number of regulatory tools are countless, starting from promoter libraries [46] to synthetic transcription factors [1, 47] and epigenetic modulators [48]. Post-transcriptional and translational regulation has less available tools for complex computation. For a long time, post-transcriptional regulation relied on ribozymes and aptamers that are only able to directly tune the gene they are encoded with. Recently, the use of RNA interference (RNAi) and RNA-binding proteins has been expanded to generate multilayered regulation [49–52]. Finally, post-translational regulation always suffered the lack of proper algorithms to predict protein structure and interactions. However, recent efforts in Professor Baker's lab at Washington University lead to the *de novo* design of proteins that can interact to generate logic gates [53]. Moreover, the breakthrough advances of DeepMind in determining protein folding with their AlphaFold 2 algorithm will revolutionize this field [54]. As described in Chapter 2, advances in implementing regulatory devices at post-transcriptional and translational level were needed to exploit the potential of all types of circuits delivery, including RNA, which enables enhanced safety as compared to DNA [55, 56]. Post-translational regulation will serve to the same scope.

However, despite the rigorous method provided by the bottom-up approach and despite the regulatory devices that enable tremendous precision in the control of timing and localization of gene expression, living systems are different from non-living machines and, as such, they are intrinsically unpredictable. Random noise in the system, stochastic gene expression fluctuations and the dependence of the host organism on the environment make parts and functional modules often behave unpredictably once interacting in the cellular environment [42].

1.2.2 The unpredictability of living systems: random noise, stochasticity of gene expression and overload

As circuits grow in complexity the role of random noise, heterogeneity and cellular capacity to sustain heterologous gene expression production become determinant. All these parameters introduce variability, and therefore unpredictability, in the system. The idea that the combination of two well characterized modules will return the sum of the two operations has to be revised and adapted to reality.

Natural systems are, almost by definition, heterogeneous [57] and gene expression as well as molecular interactions are stochastic processes [58]. In this landscape, gene regulation is decisive for adaptation and biological signals processing. For years, scientists have tried to get rid of randomness to enhance stability against stochastic system fluctuations [59, 60]. However, in the last years studies in bacteria have demonstrated how stochasticity can lead to systems stabilization [61, 62].

Another critical challenge for predictability and scalability of synthetic circuits is represented by cellular context dependency and resource limitation [42, 63–68]. The first effort to implement a tool to regulate cellular resource allocation for transgene expression was performed by Segall-Shapiro *et al* in 2014, when they engineered the first "resource allocator" to set the maximum transcriptional capacity available to a synthetic system in bacteria [69]. After this first acknowledgement of resource sharing as a limit for synthetic biology, Ceroni *et al* defined the effects of genetic load on the host cells as gene expression burden. An extensive characterization of the problem was done in bacteria [64–66], however in mammalian cells the problem has been tackled only recently [67, 68] (Chapter 3), even if evidences of gene overload-related toxicity had been already reported [70].

To summarize, as circuits grow in complexity, synthetic biologists have to face the increasing role of randomness and adopt a resource-aware design approach. To this end, mathematical models can help to make more realistic predictions of circuit's behaviour.

1.3 Modelling completes biology

Although much has been done to elucidate the biochemistry of signal transduction and gene regulatory pathways, it remains difficult to understand or predict quantitative responses of synthetic circuits, due to systems variability and context dependence. To this end, computational methods to generate models that account for resource limitation and to predict optimal circuits architectures are needed.

The discipline that accounts for modelling living systems is systems biology. Systems biology aims at developing a formal understanding of biological processes through the development of quantitative mathematical models. Systems biology takes advantage of quantitative sciences such as physics, engineering and computer science, and acts in two possible directions to either build predictive models using a bottom-up approach (from parts to systems) or to deconstruct complex networks in their basic components with a top-down approach (from systems to parts) [71].

The first approach is quantitative and is widely used to predict circuit's behaviour which is then validated experimentally *in vivo*. Control theory and computational methods mix together for the modeling, simulation, analysis, inference, and control of biological networks. This approach led to the implementation of regulatory motifs and circuit's architectures inspired by electrical engineering, such as antithetic feed-back loops, incoherent feed-forward loops, etc [59, 60, 72].

The top-down approach uncovers the network of gene regulatory interactions (gene networks) of an almost unknown biological process. It facilitates a predictive understanding for signal-activated transcription of other genes in other pathways or organisms.

Overall, the possibility to model biological systems has dramatically improved the predictability of synthetic circuits and our understanding of regulatory interactions in biology. Even simple implementations, such as the first resource-aware models [66], contributed to improve experimental designs otherwise prone to failures. On the other hand, models aided to explain the behaviour of molecular species have given insights that would have been otherwise difficult to depict experimentally [73].

The awareness that computational methods can be successfully applied to biology has led to the implementation of bioinformatic tools to fully predict the optimal circuit

architecture given the biological context and the purpose of the study [31]. Cell-type classifiers are a great example of this approach [74, 75]. As discussed in Chapter 4, the use of several approaches derived from computer science have led to the successful implementation of algorithms able to predict optimal circuits architecture to exploit cellular recognition upon the desired molecular signature [74–76].

1.4 Aim of the project

The focus of my PhD was on implementing optimal genetic circuit design and innovative regulatory tools for mammalian synthetic biology applications. As described in this thesis, I pursued this goal via three main paths: *(i)* the rational engineering of novel regulatory devices for post-transcriptional and translational gene expression control (Chapter 2); *(ii)* the extensive characterization of gene expression burden in mammalian cells (Chapter 3), to exploit possible solutions for its mitigation (Section 3.2), unveiling the molecular mechanisms responsible for resource re-allocation (Section 3.3) and understanding the global effects of payloads burden on host cells (Section 3.4); *(iii)* the implementation of a cancer cell-type classifier (Chapter 4). For these projects, experimental work was combined with computational methods to contribute to the state-of-the-art in mammalian synthetic biology and to the progress towards efficient cellular engineering.

The contents of Chapter 2 and Chapter 3, Section 3.2 have been published (Appendix E), all the other data are unpublished at the time of writing.

Chapter 2

Protein-based regulatory tools for post-transcriptional and translational control

A critical aspect of synthetic biology-based applications is the capability of genetic devices to *(i)* fine tune and *(ii)* robustly control synthetic and endogenous genes expression. To this end, regulatory devices are crucial to achieve robust control over synthetic circuits and for this reason, scientists have always focused on implementing novel tools for gene regulation. Scientists have long worked on implementing novel tools for transcriptional and post-transcriptional regulation of DNA. Transcriptional regulation has involved design and characterization of synthetic promoters [77, 78] and the engineering of novel transcription factors [79]. This approach has led to great advances in synthetic biology; nevertheless, for some applications (i.e. the ones requiring only temporary expression of the genetic devices, or that require fast regulation dynamics) regulatory devices that act at the post-transcriptional and/or translational level are highly desirable. The possibility to fully exploit RNA-encoded synthetic circuit's potential, as well as to multiplex more regulatory levels in the same circuit design, prompted the development of such regulatory devices. RNA delivery of transgenes has particular significance for biomedical applications, where the unwanted risk of mutagenicity and immunogenicity given by exogenous DNA must be avoided [56, 80]. Moreover, RNA has faster expression dynamics as compared to DNA, retaining the possibility to spatially and temporally control gene expression

[5, 33]. So far, RNA-regulatory devices included ribozymes, aptamers, riboswitches that can modulate the translation of the associated output but cannot be interconnected to create modular and scalable circuits [81–84].

In this chapter, I will give an overview of the state-of-the art in RNA-encoded genetic circuits control (Section 2.1) and I will present a novel set of RNA-binding proteins (RBPs) and proteases that we engineered to address the lack of post-transcriptional and translational regulatory devices (Section 2.2). Finally, I will draw the conclusions that arise from the results of our study (Section 2.3).

2.1 Towards more robust post-transcriptional and translational regulation of synthetic circuits

Synthetic circuits in bacteria and mammalian cells with high biocomputing capabilities are based on smart regulatory interfaces that can sense complex inputs to activate a cellular response [1, 85]. As mammalian cells are particularly attractive for biomedical applications, scientists have recently focused on implementing sensor-actuator devices that rely on protein-based parts that sense intracellular and extracellular signals [25]. These devices are able to detect environmental changes and cellular states, being a promising platform for diagnostic purposes. To this end, timely and localized activation of cellular response is essential. Protein-based interfaces serve to a broad range of applications, as proteins are key regulators of all cell activities and play a major role in signal transmission. Among these interfaces, the ones that exert control at post-transcriptional and translational level maximize the insulation of synthetic systems from the cellular machinery. For this reason and the fact that RNA-only delivery modality may address issues of immunogenicity and unwanted chromosomal integration [49, 52, 86], circuits that operate post-transcriptionally are gaining consideration. In addition, with protein-protein regulation the response time reduces to the order of minutes, making these devices advantageous for the treatment of metabolic disorders like diabetes where dynamic changes in metabolites level are quite fast and require precise feedback control. At the same time, they are compatible with the regulation of synthetic circuits encoded by RNA, which are a promising framework for the reprograming of cellular function. These have potential applications both as research tools and for future *in vivo* therapeutics due to their transient expression and lack of chromosomal integration hazards, overcoming safety issues associated with DNA-based circuit delivery.

Indeed, a key aspect to consider when using RNA-encoded devices for biomedical applications is mRNA stability; chemical and structural modifications of *in vitro* transcribed mRNA have enabled production of significant levels of target proteins for durations longer than a week [87, 88]. Recently, self-replicating RNA viruses such as Sindbis virus have raised significant interest for a wide range of medical applications including vaccine delivery, gene therapy and cellular reprogramming [89, 90]. A main advantage of RNA

replicons is that since they self-replicate, they can generate high levels of gene products even when starting from a relatively low dose, while the risk of undesired chromosomal integration is minimal because they do not reverse transcribe. Sindbis replicons with reduced cytopathicity and longer-term expression have been recently developed, making them an attractive platform for engineering RNA-encoded synthetic devices [91, 92].

2.1.1 RNA-encoded circuits

In 2015, Wroblewska *et al* showed for the first time that complex cellular logic can be encoded exclusively at the post-transcriptional level in mammalian cells, pioneering the multi-layered regulation of RNA-encoded genetic circuits [49]. To this end, they used two RBPs, L7Ae and MS2-cNOT7 responsive to endogenous miRNAs to implement a cell-type classifier, a cascade and a switch.

In a follow-up work, Matsuura *et al* used L7Ae engineered to respond to several cell-specific microRNAs to compute complex Boolean logic gates [50]. They could tune their system to create AND, OR, NAND, NOR and XOR gates for the control of a fluorescent protein expression. Eventually, they showed that the AND gate could be used to selectively induce cellular apoptosis in presence of miR-206 and miR-302a.

Inspired by the convenient modularity and orthogonality of the tools described in these studies, we decided to expand the toolbox of regulatory devices for post-transcriptional and translational control.

2.2 Engineering protein-protein devices to regulate mRNA translation

Cella *et al*, Nature Communications (2018)

Our goal was to expand the existing toolbox of RNA regulatory devices with molecular parts that could be multiplexed to achieve complex regulation of the output. We envisioned to use viral proteases orthogonal to intracellular environment in combination with RBPs, previously characterized as efficient mRNA regulators [12, 49].

Proteolytic devices present a possible platform for multi-element circuit engineering, but today are largely used in synthetic biology to modulate protein half-life, or to trigger selected cellular responses [12, 93–95]. Proteases are broadly expressed in all organisms and function by cleaving short, specific amino acid sequences, are involved in all areas of metabolism, and constitute key effectors of several cellular pathways, modulating protein-protein interaction, protein localization, and ultimately regulating cellular fate [96, 97]. They are classified on the basis of cleavage site [98] (i.e. endopeptidases cleave the target protein internally and exopeptidases remove single amino acids from N or C-terminal end of the protein), evolution, substrate specificity and reaction catalyzed [99]. Further, proteases can be adapted to operate in several environmental conditions and optimized to function in different organisms. Viral proteases such as tobacco etch virus protease (TEVp) are widely used in biotechnology for endoproteolytic removal of affinity tags from recombinant proteins [100], and in synthetic biology to help convert extracellular or intracellular ligand sensing into transcriptional output [12, 94, 95, 101]. They are also used to control protein levels via an N-terminal degron fused to the target protein with a protease-responsive cleavage site, as demonstrated in bacteria [93]. To date there are no off-targets substrates for TEVp reported in the human proteome [102] and in other organisms where they are in fact well tolerated [103–105]. The use of TEVp has enabled innovative studies of biological processes. For example, engineering of a TEVp-responsive Rad21 in a cell type-specific and temporally controlled manner unveiled cohesin function in post-mitotic cells like Drosophila neurons [103]. We thus reasoned that proteases represent an attractive tool for engineering more sophisticated post-transcriptional and translational regulatory networks, and that protease-RBP and protease-protease regula-

tory devices could be exploited for circuits that are solely RNA-encoded and are based on protein-protein interactions.

Thanks to their ability to directly bind RNA, we used RBPs as first regulatory step in all the experimental setups. We then created a multilayered system that add further regulatory layers via protein-RNA and protein-protein interactions using a protein engineering strategy. To engineer protein-protein interaction, we used the TEVp [106], and other proteases of the same family, namely the tobacco vein mottling virus protease (TVMVp) [107], the turnip mosaic virus protease (TUMVp) [108] and the sunflower mild mosaic virus protease (SuMMVp) [100], all recognizing specific amino acid sequences. We envisioned that these unique cognate cleavage sites could be inserted in a desired protein, making it susceptible to protease activity. The availability of crystal structures for a large number of proteins and of software for their visualization (e.g. PyMOL and SWISS-MODEL) allows to identify insertion sites that minimize the risk of detrimental structure and function modifications of the protein of interest. This system is potentially highly modular and multiplexable.

By using viral proteases to regulate RBP activity, we implemented complex regulatory architectures that operate at the post-transcriptional and translational level with tight control over gene expression. Our devices are able to sense intracellular signals and actuate a response. We validated our approach by sensing the hepatitis C marker (HCV) serine protease NS3 (Section 2.2.2). Furthermore, we proved that our protein parts can be linked to compute and decode multi-input information. We proved it by implementing a three-stage cascade and a two-input switch that set the system to different states depending on the siRNA input they receive (Section 2.2.3).

Overall, our work expanded the toolbox of parts available for the post-transcriptional and post-translational control of genetic circuit's behavior. These parts can sense intracellular changes, using proteins as input to trigger a cellular response.

2.2.1 Engineering a TEV protease-responsive L7Ae

We started by focusing on the archaeal RNA binding protein L7Ae. L7Ae binds the RNA C/D box that forms a classical kink-turn (K-turn) motif [49, 109] in the 5'UTR of target mRNAs, inhibiting their translation [110]. As L7Ae had been successfully used by Wroblewska *et al* [49] for the regulation of RNA-encoded genetic circuits, we decided to link its activity to the TEV protease (TEVp).

As a first step, we identified possible insertion points for the TEV cleavage site (TCS) based on the crystal structure of L7Ae bound to its native RNA recognition sequence. Protein structure was determined by Moore *et al* [109] (PDB id: 1RLG). L7Ae-TCS structure visualization was performed with PyMOL (The PyMOL Molecular Graphics System, Version 2.0 Schrödinger, LLC, see https://pymol.org/2/). We had two main objectives in choosing the sites: *(i)* TCS insertions should minimally affect L7Ae structure and/or RNA binding, and *(ii)* TEVp cleavage should render L7Ae non-functional. We thus chose three candidate insertion sites for TCS in loop regions away from the K-turn binding domain and closer to the center of the L7Ae sequence (**Fig. 2.1a**). Specifically, they were placed after amino acid residue N70 (L7Ae-CS1), P56 (L7Ae-CS2), or K77 (L7Ae-CS3), shown in red for each side.

Accordingly, we built and tested three new translation repressors (L7Ae-CS1, L7Ae-CS2, L7Ae-CS3) along with a reporter gene regulated by two repeats of the K-turn motif in its 5'UTR (2Kt-dEGFP). As shown in **Fig. 2.1b**, in the absence of TEVp (State 1) and similar to the behavior with wild-type L7Ae, the three L7Ae-CS repressors downregulate translation of the reporter gene. In the presence of TEVp (State 2 in **Fig. 2.1b**), we observed 13.4 fold derepression of CS2 and 77 fold derepression of CS3, but no appreciable response to CS1 (**Fig. 2.1c-d**). Since L7Ae-CS3 showed the highest response to TEVp cleavage, we decided to focus on this variant for the following experiments. The circuit harboring L7Ae-CS3 also showed efficient and sustained repression/derepression over a long-time frame (96 hours) (**Fig. 2.1e**). Following these interesting results, we used this new protease dependent translation system to create several circuits, including an HCV protein sensor (Section 2.2.2), a cascade and a switch (Section 2.2.3).

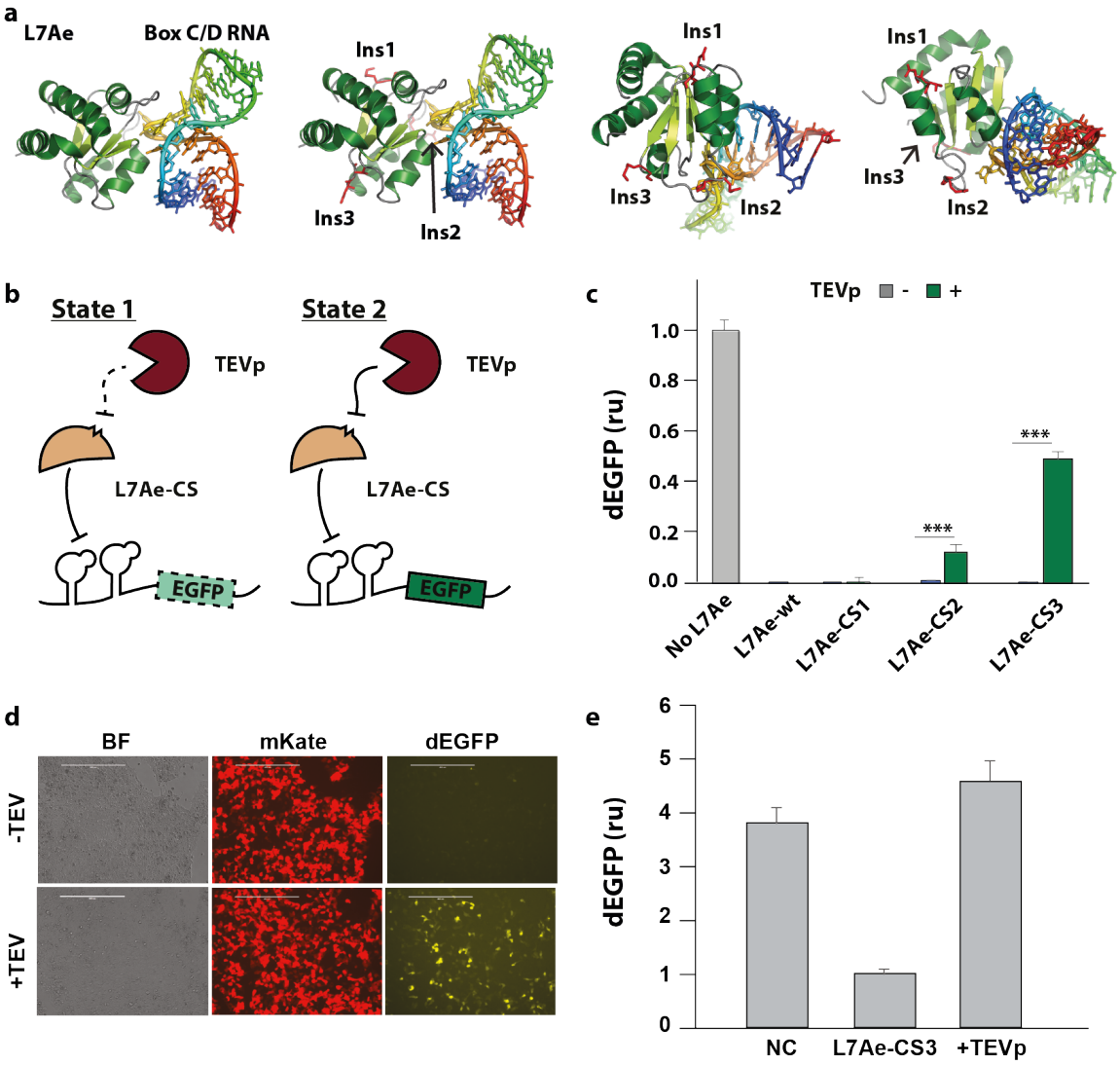


Figure 2.1: **Engineering a TEVp responsive L7Ae. (a)** Structure of L7Ae binding to box C/D RNA that forms a K-turn motif. TCS insertion sites in L7Ae were placed after amino acid residue N70 (L7Ae-CS1), P56 (L7Ae-CS2), or K77 (L7Ae-CS3), shown in red for each site. Protein structure was determined in [109] (PDB id: 1RLG). L7Ae-CS structure visualization was performed with PyMOL (https://pymol.org/2/). **(b)** Schematics of L7Ae with inserted TCS translational regulation. In the absence of TEVp, L7Ae-CS binds and repress the K-turn motifs in the 5'UTR of mRNA target (EGFP OFF-State 1, left). When TEVp is expressed, it cleaves the TCS, disrupting L7Ae structure and inhibiting its function (EGFP ON-State 2, right). **(c)** Flow cytometry analysis of three engineered L7Ae-CS tested in HEK293T cells in the absence or presence of TEVp. Data represent geometric mean of dEGFP normalized by transfection marker mKate expressed from the same constitutive promoter to account for different expression across the cell lines +/- SE. **(d)** Microscopy images for L7Ae-CS3 +/- TEVp. BF: bright field. Scale bars indicate 200 µm. **(e)** dEGFP levels 96h post-transfection. Data represent geometric mean and SE of EGFP normalized by transfection marker mKate. SE: standard error. ru: relative units. N = 3 biological replicates. Unpaired t-test, p-value: *** < 0.001.

2.2.2 L7Ae-CS3-based protein sensor

The HCV sensing device detects NS3, a serine protease, which is considered an attractive target for HCV therapy as it is responsible for processing the polyprotein precursor encoded by the virus and is essential for viral replication [111]. We used two single chain fragment intrabodies scFv35 and scFv162 specific for two distinct epitopes of NS3 previously reported to interfere with its viral activity [112]. These intrabodies can be efficiently wired into a transcriptional protein sensing-actuation device to detect NS3 expression in mammalian cells [12]. We designed a new device in which the sensor is based on translational regulation using L7Ae-CS3 and TEVp fused to scFv35 and scFv162 respectively (**Fig. 2.2a**). The presence of the target protein and subsequent binding of the two intrabodies brings the system components in close proximity, resulting in elevated TEVp cleavage of the RBP and derepression of the target 2Kt-dEGFP (**Fig. 2.2a**). TEVp-scFv162 was under the regulation of a doxycycline-inducible promoter (TRE promoter). In fact, as shown in [12], this allows to achieve specific output expression in the presence of NS3 and to minimize non-specific cleavage by TEVp. In [12], they demonstrated that TEVp fused to scFv162 retains protease activity. Similarly, we tested scFv35 intrabody fused to N- or C-terminus of wild-type L7Ae or L7Ae-CS3. All chimeric proteins efficiently repressed dEGFP (**Fig. 2.2b**), and both scFv35-L7Ae-CS3 and L7Ae-CS3-scFv35 responded to TEVp, rescuing reporter translation (**Fig. 2.2c**). Next, we tested the HCV sensor in HEK293T cells by expressing scFv35-L7Ae-CS3 or L7Ae-CS3-scFv35 along with TEVp-scFv162, in the presence or absence of a blue fluorescence-tagged NS3 (BFP-nNS3) [12]. dEGFP expression significantly increased in the presence of BFP-nNS3 (**Fig. 2.2d**), indicating specific protein-sensing cleavage of L7Ae-CS3 fused to scFv35 by TEVp. Of note, we observed stronger derepression with scFv35-L7Ae-CS3, perhaps due to facilitated interaction with the TEVp-cleavage site.

2.2.3 L7Ae-CS3 based multi-input information processing

We next decided to explore multi-input information processing and actuation of cellular responses. A robust biosensor should be able to detect the change of several intracellular species, as it is often needed to detect physiological changes. To this end, we need genetic

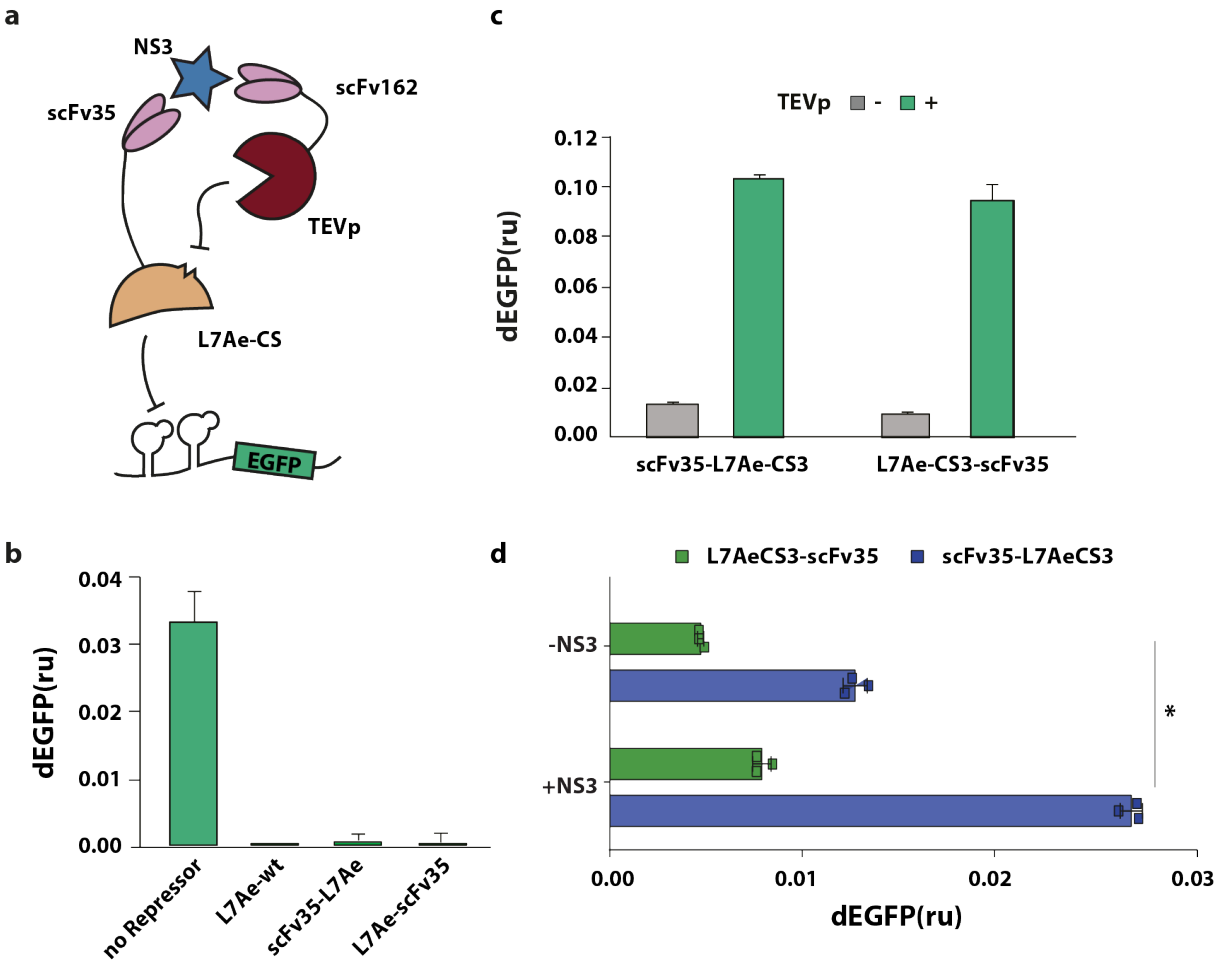


Figure 2.2: **HCV protein sensor based on L7Ae-CS3.** **(a)** Schematics of L7Ae-CS3 based HCV sensor. TEVp-scFv162 [12] and L7AeCS3-scFv35 fusion proteins are modules of a sensor for NS3 protein, which is associated with HCV virus. L7Ae-CS3 represses the mRNA target in the absence of NS3. In the presence NS3, binding of the intrabodies brings TEVp in proximity to L7Ae-CS3, resulting in cleavage of the RBP and derepression of the target gene. **(b)** Test of repression efficiency of wild-type L7Ae fused to the N- or C-terminus of scFv35 intrabody. Both scFv35-L7Ae and L7Ae-scFv35 show similar repression compared to L7Ae alone. **(c)** Test of repression efficiency of L7Ae-CS3 fused to the N- or C-terminus of scFv35 intrabody. In presence of TEVp, dEGFP levels increase, indicating inhibition of L7Ae-CS3 repression. **(d)** Test of the variants of NS3 sensor in the presence and absence of BFP-nNS316. Data collected 48 hours post transfection, represent geometric mean and standard deviation of Geometric Mean of dEGFP normalized by transfection marker mKate +/- SE. SE: standard error. ru: relative units. N=3 biological replicates. Unpaired t-test. pvalue: * < 0.05.

circuits that comprise multiple sensing units and decode the signal to one digital information which corresponds to the actuation of the desired cellular response. To address this need, we implemented a protein-based two-input three-stage repression cascade and a two-input switch.

The two-input three-stage repression cascade decodes the inputs signals with an overall logic function that is true in the presence of the synthetic short interfering RNA FF5 (siRNA-FF5, Input 1) or absence of the synthetic short interfering RNA FF4 (siRNA-FF4, Input 2). This circuit operates using both post-transcriptional and post-translational regulation (**Fig. 2.3a**). The first input to the cascade is the siRNA-FF4 that downregulates expression of TEVp through four repeats of the FF4 target site placed in the 3'UTR. To monitor TEVp modulation, we designed TEVp-2A-EYFP that couples fluorescence intensity to protease expression through a self-cleaving 2A peptide. At the next stage of the cascade (Stage 2), TEVp cleaves L7Ae-CS3, which otherwise binds K-turn motifs in the 5'UTR of RNA encoding a blue fluorescent protein (2Kt-EBFP) (Stage 2, **Fig. 2.3a**). For the second input, we placed four repeats of the target site for siRNA-FF5 in the 3'UTR of L7Ae-CS3. We tested the behavior of the circuit in HEK293T cells and observed that EBFP is repressed by L7Ae-CS3 (Stage 1), whereas knocking down L7AeCS3 activity either by siRNA-FF5 (Input 1) or by TEVp (Stage 2) restored reporter output. siRNA-FF4 (input 2) induced TEVp downregulation and resulted in rescue of L7Ae-CS3 activity and EBFP repression (**Fig. 2.3b**).

Then, we engineered a two-input switch based on protein-protein/RBP-RNA cross regulation, in which TEVp and L7Ae-CS3 repress each other (**Fig. 2.3c**). Our switch has a general cross repression topology that resembles previously described mammalian and bacterial switches [20, 37, 49]. Our switch is similar to the cascade described above, with the addition of two K-turn motifs in the 5'UTR of TEVp-2A-EYFP (**Fig. 2.3c**). Administration of either siRNA-FF4 or siRNA-FF5 was used to set the state of the switch. In the absence of siRNAs, both repressors and reporter genes remained at intermediate levels 48h post-transfection (**Fig. 2.3d**). siRNA-FF4 set the state to high L7Ae-CS3 resulting in low EYFP and EBFP; on the contrary, siRNA-FF5 resulted in the opposite state (**Fig. 2.3d-e**).

The results in this sections and in Section 2.2.2 show that L7Ae engineered to be TEVp responsive (L7Ae-CS3) is suitable to engineer regulatory networks both to sense clinically relevant biomarkers and to decode multi-input information.

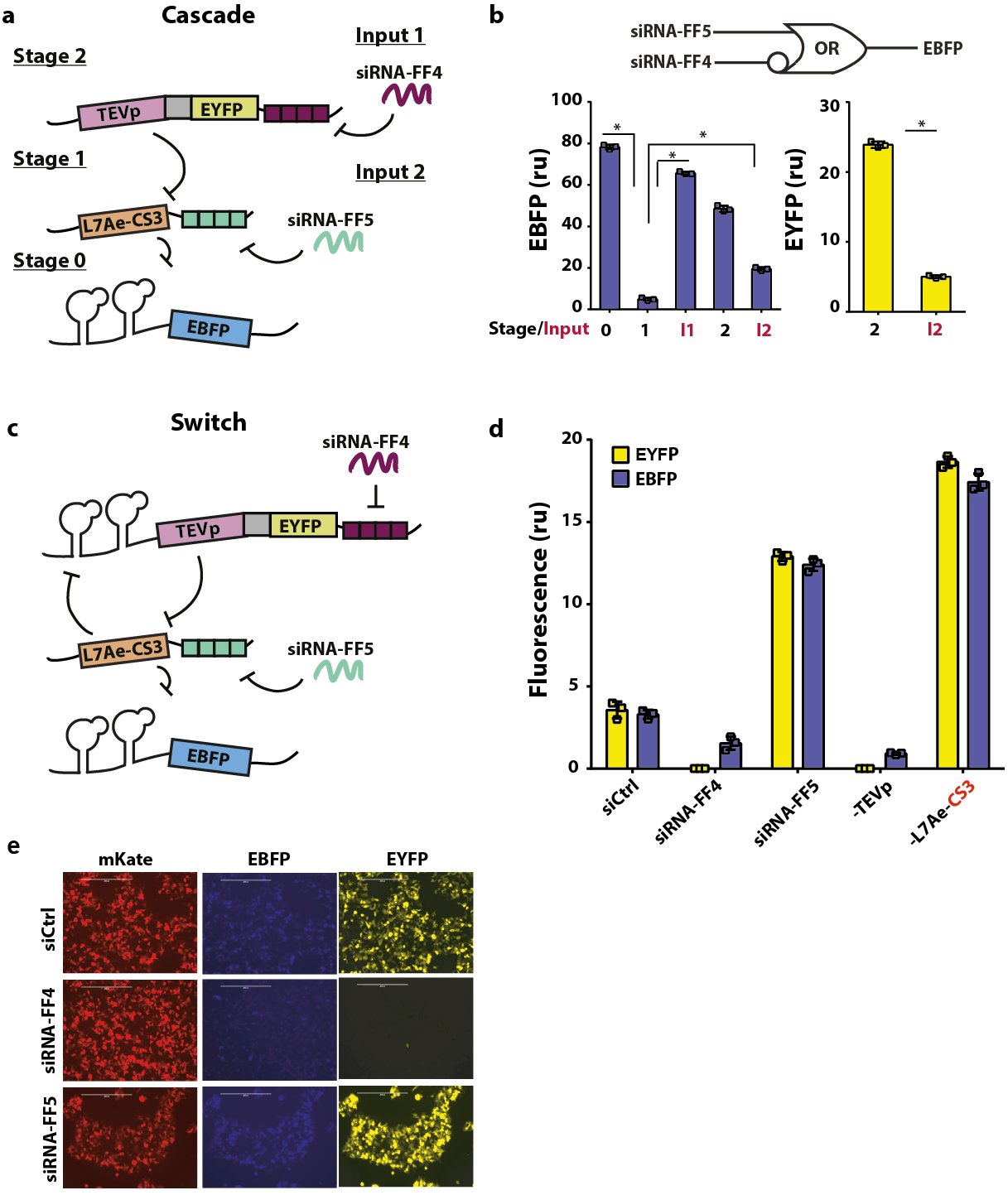


Figure 2.3: **L7Ae-CS3-based cascade and two-state switch. (a)** The cascade includes both protein-protein regulation and post-transcriptional regulation by siRNA. The L7Ae-CS3 transcript incorporates four siRNA-FF5 target sites in the 3'UTR. TEVp, which is coexpressed with EYFP, is downregulated by siRNAFF4 via target sites in its 3'UTR. EBFP is expressed in absence of L7AeCS3 (Stage 0) and repressed in the presence of the RBP (Stage 1). TEVp cleaves L7Ae-CS3, rescuing EBFP expression (Stage 2). siRNA-FF5 (Input 1) downregulates L7Ae-CS3 in the absence of TEVp. siRNA-FF4 (Input 2) knock down of TEVp results in high L7Ae-CS3 levels and thus in EBFP repression. **(b)** Top, simplified logic circuit and corresponding flow cytometry data.

Figure 2.3: The cascade implements the two-input logic "siRNA-FF5 OR (NOT siRNA-FF4)" (top). Bottom, bar charts of the EBFP and EYFP circuit outputs as additional stages of the cascade and the inputs are added to the experiment, starting from just the reporter (Stage 0), and then adding L7Ae-CS3 (Stage 1), input siRNA-FF5 (Input 1) or TEVp-2A-EYFP (Stage 2), and siRNA-FF4 (Input 2). Data represent geometric mean +/- SE of EBFP and EYFP normalized by transfection marker mKate. SE: standard error. ru: relative units. N = 3 biological replicates. Unpaired t-test. pvalue = < 0.05. **(c)** Schematics of the L7Ae-CS3-based switch. TEVp-2A-EYFP includes two K-turn motifs in the 5'UTR and four target sites for siRNA-FF4 in the 3'UTR. L7Ae-CS3 mRNA harbours 4 siRNA-FF5 target sites in the 3'UTR. To monitor L7Ae-CS3 activity, we included EBFP reporter with two K-turn motifs in the 5'UTR. siRNA-FF4 and siRNA-FF5 are used to set and reset the state of the switch. **(c)** Flow cytometry data representing network behaviour. Data represent geometric mean +/- SE of EBFP and EYFP normalized by transfection marker mKate. SE: standard error. ru: relative units. N=3 biological replicates. Unpaired t-test. pvalue = < 0.05. **(e)** Representative fluorescent micrographs of the switch showing the state of the switch regulated by siRNA-FF4 and siRNA-FF5. Scale bars indicate 200 µm. Data collected 48 hours post transfection.

2.2.4 Engineering protease-responsive MS2-cNOT7 repressor devices

The RBP MS2-cNOT7 is a fusion protein [49]. MS2 is a bacteriophage coat protein that binds to a stem-loop structure of the phage genome [113], whereas cNOT7 is a subunit of the CCR4-Not deadenylase complex that removes the poly-A tail of the transcripts triggering their degradation [114]. MS2-cNOT7 fusion protein binds the 3'UTR of mRNAs containing MS2 binding motives and cNOT7 catalyzes mRNA de-adenylation, and as described before was successfully implemented in RNA-only encoded networks.

To enable regulation by TEVp, we inserted a TCS at the junction of MS2-cNOT7 (MS2-CS-cNOT7, **Fig. 2.4a**). The readout of the system was an EGFP with 8 MS2 binding motives in the 3'UTR (EGFP-8xMS2bm). In the absence of TEVp, MS2-CS-CNOT7 showed similar repression of the target to its wild-type counterpart, whereas co-expression of the protease fully restored EGFP expression, indicating efficient and sustained disruption of RBP activity (**Fig. 2.4b**).

Thanks to the protein structure with a flexible linker between MS2 and cNOT7, the insertion of cleavage sites for cognate proteases is straight forward. Therefore, we next examined whether we could re-engineer the synthetic RBP to respond to different proteases. We built a library of MS2-CS-cNOT7 variants responsive to SuMMVp [93],

TUMVp and TVMVp [107] proteases. For these, we tested two alternative cleavage sites (Serine-S or Glycine-G in P1 [115] position of the cleavage sequence) with different affinities for the cognate protease to find the design that showed the best cleavage efficiency in addition to orthogonality to the other proteases (**Fig. 2.4c**). We did not include TEVp since previous experiments performed in our lab showed similar response when using either of the two cleavage sites. All MS2-CS-CNOT7 variants that we constructed repress EGFP-8xMS2bm translation, and the associated proteases disrupt this repression (highest efficiency obtained with TEVp and TUMVp), resulting in increased EGFP expression (**Fig. 2.4b-d**). The three-vector system with proteases and associated RBPs encoded on separate plasmids was then used to measure crosstalk between non-cognate pairs. All combinations of MS2-CS-cNOT7 and proteases were co-transfected in HEK293 cells, demonstrating that the proteases are mostly orthogonal and therefore suitable to multilayered circuit construction (**Fig. 2.4e**).

2.2.5 Engineering protease-protease interaction to generate multi-stage repression cascades

Similar to what we did for L7Ae and MS2-cNOT7 (Section 2.1.1 and 2.1.4), we hypothesized that we could extend our approach to design proteases that include a cleavage site for an orthogonal one, enabling a novel protein-protein regulation system. Towards this goal, we first analyzed TVMVp structure with PyMol (https://pymol.org/2/) and identified three insertion sites following the same criteria applied to L7Ae (Section 2.1.1). In the three insertion sites we included the TUMVp cleavage site: *(i)* TVMVp-TUCS1 between amino acid residues D26-G27, *(ii)* TVMVp-TUCS2 between amino acid residues Q119-K120, and *(iii)* TVMVp- TUCS3 between amino acid residues T173-N174 (**Fig. 2.5a**). We tested the three variants along with an MS2-cNOT7 in which we inserted the cleavage site for TVMV (MS2-TVCS-cNOT7), which in turn represses EGFP translation (**Fig. 2.6a**). Expression of TUMVp should result in TVMV-CS inhibition and restore EGFP down-modulation by MS2-TVCS-cNOT7 (**Fig. 2.6a**). We observed that one variant, TVMVp-TUCS2, interferes with MS2-TVCS-cNOT7 activity, while expression of TUMVp re-establishes EGFP translation repression (**Fig. 2.6b**).

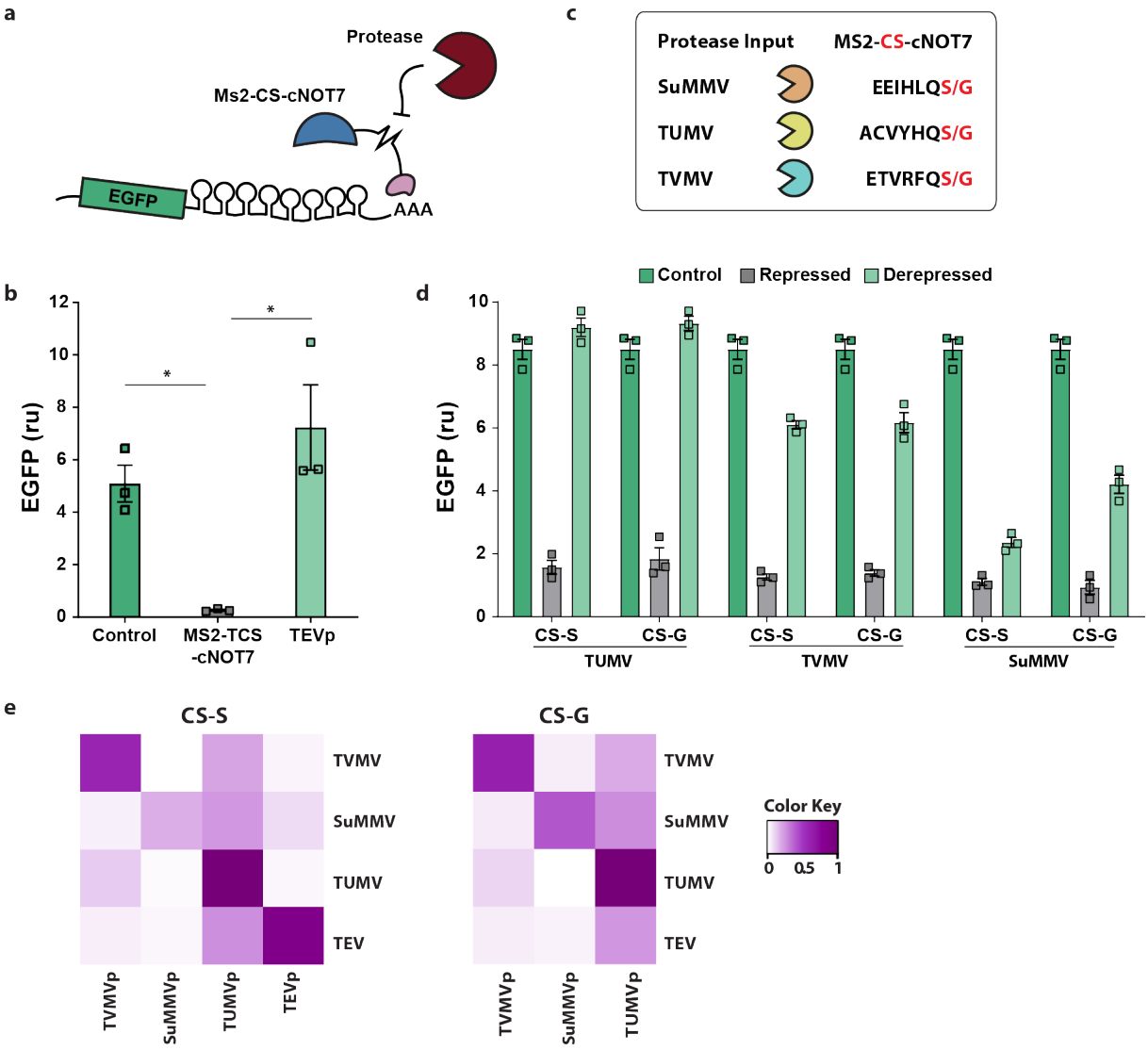


Figure 2.4: **Engineering protease-dependent MS2-cNOT7 proteins.** **(a)** Schematics of the cascade. MS2 binds its cognate sequences in the 3'UTR of target EGFP mRNA, resulting in RNA deadenylation by cNOT7. When TEVp is co-expressed, MS2-CS-cNOT7 is cleaved and the two domains are separated, rescuing EGFP translation. **(b)** Flow cytometry results. EGFP translation recovers in the presence of TEVp. Data represent geometric mean +/- SE of EGFP normalized by transfection marker mKate. SE: standard error. ru: relative units. N = biological replicates. Unpaired t-test. p-value: * < 0.05. **(c)** Proteases and cognate cleavage site sequences used to create additional regulatory MS2-CS-cNOT7 proteins. **(d)** Analysis of MS2-CS-cNOT7 variants in the absence (Repressed) or presence (Derepressed) of the cognate protease. Cleavage sites include either a Serine or a Glycine in P'1. The proteases are ordered from the most to the least efficient in EGFP derepression. Data represent geometric mean +/- SE of EGFP normalized by transfection marker mKate. **(e)** Proteases orthogonality. Regulatory effect is shown between all combination of proteases and MS2-CS-cNOT7, for cleavage sites with Serine in P1 (left) and with Glycine in P1 (right). Data represent geometric mean of EGFP normalized by transfection marker mKate. N = 3 biological replicates for all experiments. Data was collected 48 hours post transfection.

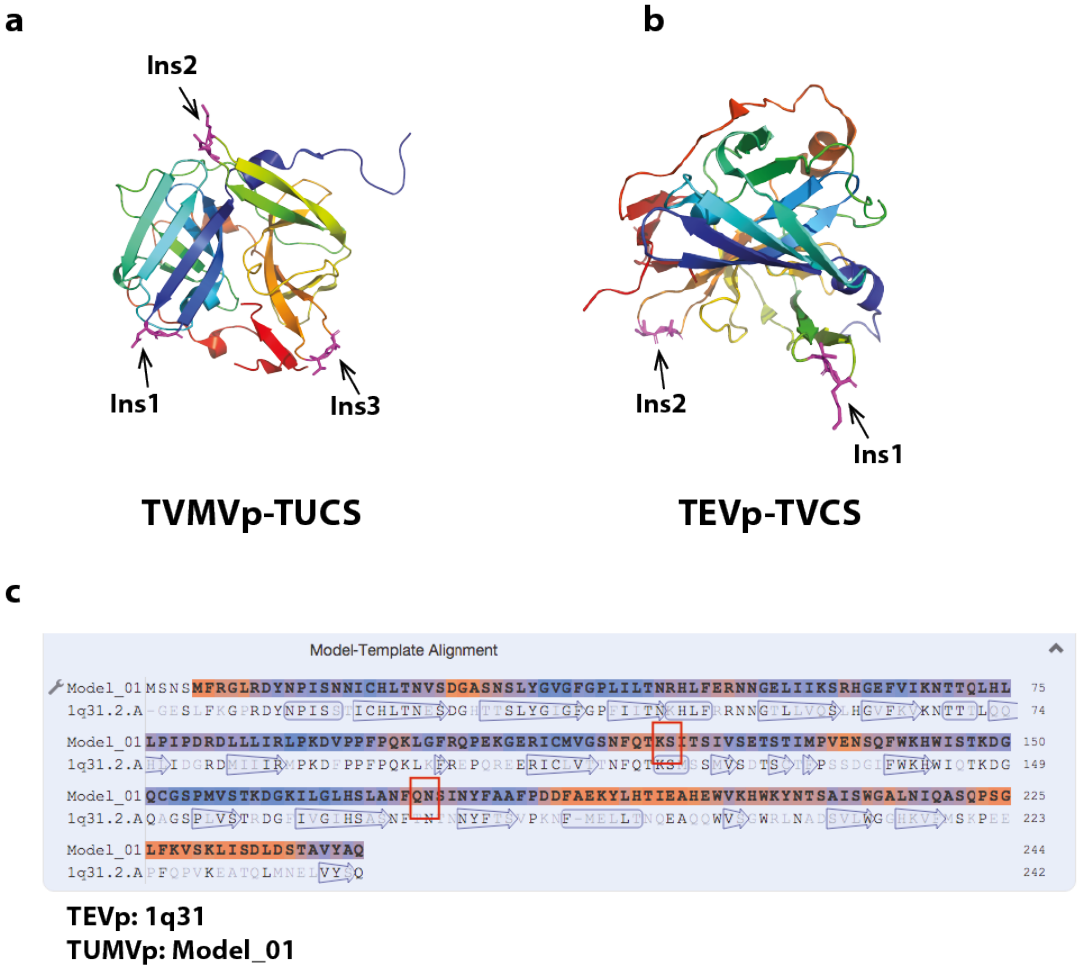


Figure 2.5: **Engineered proteases with cleavage sites.** **(a)** TVMVp with three alternative insertion points for TUMV cleavage site (TUCS). Insertion were placed between amino acid residues D26-G27 (Ins1, TVMVp-TUCS1), Q119-K120 (Ins2, TVMVp-TUCS2), T173-N174 (Ins3, TVMVp-TUCS3). **(b)** TEVp with insertion sites for TVMVp (TVCS). Insertions were placed between amino acid residues T117 and K118 (Ins1, TEVp-TVCS1) or T173-N174 (Ins2, TEVpTVCS2). Protease structure visualization was performed with pymol (http://www.pymol.org). **(c)** TUMV structure was homology modeled using swiss model and TEV2 1Q31.pdb template. Shown is model-template alignment of TUMV sequence aligned to 1Q31 with insertion sites highlighted in red (Ins1:TK, Ins2:QN)

Since TEVp and TVMVp share a high degree of structural similarity, we inserted TVMV-cleavage site (TVMV-CS) in the same amino acid regions, creating new TEVp-TVCS proteases (**Fig. 2.5b, 2.6c**). The TEVp-TVCS1 variant efficiently impaired MS2-TCS-cNOT7 repression of EGFP, and addition of TVMVp inhibited TEVp-TVCS1 function (**Fig. 2.6d**). Our results suggest an efficient insertion site within homologous proteases (amino acid region NFQX-KS, where X refers to variable amino acidic residue).

Based on this information we designed a structurally modified TUMVp with TEVp cleavage site insertion (TCS) (**Fig. 2.6e**). Since the TUMVp crystal structure is yet unresolved, we instead inferred its structure based on homology to TEVp using SWISS-MODEL (https://swissmodel.expasy.org/), an automated protein homology modelling server (**Fig. 2.5c**). We observed MS2-TUCS-cNOT7 inhibition by TUMV-TCS, with TEVp reverting the effect, albeit with reduced efficiency as compared to the other proteases (**Fig. 2.6f**).

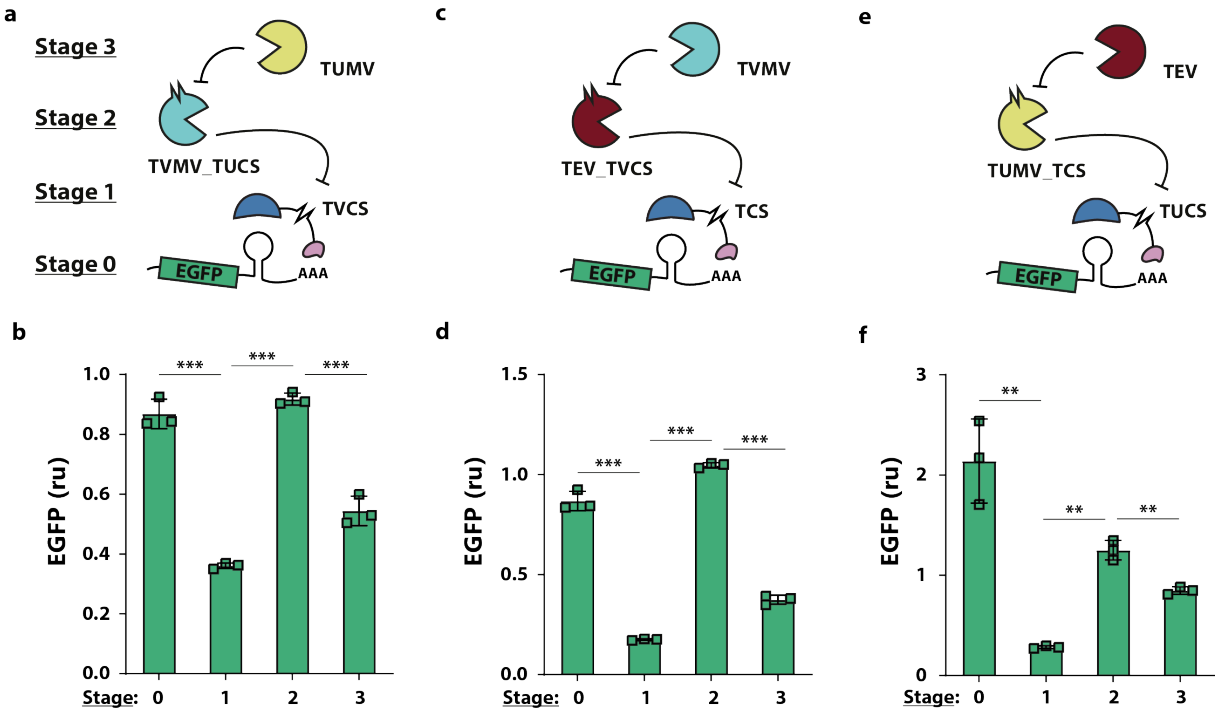


Figure 2.6: **Protease-based cascades. (a,c,e)** Schematics of the three-stage signaling cascades. EGFP translation is repressed by MS2-CS-cNOT7 (Stage 1), while RBP activity is disrupted by a protease (Stage 2), which is itself engineered to include a cleavage site for an orthogonal protease. When Stage 2 protease activity is inhibited by an upstream protease (Stage 3), MS2-CS-cNOT7 is able to repress target mRNAs. **(b,c,d)** Corresponding flow cytometry data for the geometric mean +/- SE of EGFP normalized by transfection marker mKate and rescaled to no repression values. SE: standard error. ru: relative units. N = 3 biological replicates. Unpaired t-test. pvalue: *** < 0.001, ** < 0.005. Data collected 48 hours post transfection.

2.3 Viral proteases and RBPs can be engineered to program complex regulatory motives

With our project we laid the foundation for protein-driven synthetic circuits that enable construction of tunable RNA-encoded networks. We demonstrated a set of regulatory networks including cascades, a switch and a protein sensor-actuator, showing that they operate in an orthogonal fashion combining post-transcriptional and post-translational regulation. We first engineered RBPs L7Ae and MS2-cNOT7 to include cleavage sites for cognate proteases, which exhibit robust and long-term protease-dependent regulation of target mRNA. With MS2-cNOT7 fusion protein, we created a protease-dependent system via cleavage sites inserted at the junction of the two modules. The design of L7Ae was based on analysis of the crystal structure of the RBP bound to native recognition RNA sequence, with aims to minimize alteration of wild type L7Ae structure upon insertion, to minimize interference with the ability of L7Ae to bind RNA, and maximize cleavage in presence of TEVp that render L7Ae non-functional. The L7Ae proteins with integrated cleavage sites were used to compose three different devices, including NS3 protein sensor. We used an inducible promoter to control TEVp-scFv162 expression to minimize the likelihood of random cleavage [12]. The L7Ae proteins with integrated cleavage sites were used to compose three different devices, including NS3 protein sensor. Similar to L7Ae, engineering of the proteases was based on crystal structure analysis of TEVp and TVMVp, while the cleavage site for TUMVp was inferred with a homology-based model (SWISS-MODEL). For adding new regulatory elements, computational protein design methods are becoming increasingly accurate at predicting folding and functionality of newly design proteins [116], which should facilitate the construction of new protein-based synthetic networks. We envision that our platform could also be connected to intracellular protease inputs (i.e. caspases) without requiring modifications to endogenous pathways to link changes of intracellular state to output regulation (Section 2.3.1).

The motifs we implemented have all biological relevance. Regulatory cascades can generate precisely controlled gene expression dynamics in response to environmental and intracellular signals [117]. Sequential transcriptional activation allows for optimal use of transcriptional resources, as shown both in prokaryotic and eukaryotic systems [118,

119], however they come with the risk of cross-talk with cell natural pathways. Here we report that the use of our tools for post-transcriptional and translational ordered events allow for precise and efficient control of gene expression avoiding the risk of interfering with endogenous pathways. At the same time, switches are extremely important in cell biology, from metabolic control to cell fate commitment [120]. Their "all-or-none" behavior, that is independent of input concentration, makes switches ultrasensitive and robust against noise perturbation [121, 122]. Initially, implementation of these motifs mainly relied on transcriptional regulation. Huang *et al* in 2012, showed for the first time in a prokaryotic systems that orthogonal and heterologous proteases can be employed to tune a nonlinear protein degradation circuit that shows bistable behavior [123]. With our work we show that a RBPs/proteases-based two-input switch can be implemented in mammalian cells. Finally, sensor-actuator devices are the bridge between intracellular and environmental signals and synthetic control of cellular response [12, 25]. The most common protein-based parts that have been engineered to sense molecular changes are the SynNotch receptors [124]. However, SynNotches as any other membrane receptor, sense extracellular signals. In our work, we expanded the approach first presented by Siciliano *et al* [12] - that uses intrabodies linked to a TEVp to sense intracellular markers - by using RBPs to localize all the regulation in the cytoplasm.

Overall, our system provides a strategy for composable and scalable circuit design encoded post-transcriptionally in mammalian cells, offering a useful research toolbox with potential benefits for RNA delivered *in vivo* applications such as cell-based therapies, vaccination, regenerative medicine, and biotechnological applications.

Concomitantly to our work, Gao *et al* created protein-protein regulatory circuits based on viral proteases, including TEVp, TVMVp and HCVp, which were connected to reciprocally modulate their activity (e.g. by destabilizing domains) generating regulatory cascades, binary logic gates, and dynamic analog signal-processing functions [52]. They named their system CHOMP for "circuits of hacked orthogonal modular proteases". Similar to our experimental setup, they used proteases as input to cleave target proteases inhibiting their function. The unique feature of their system is that they implemented several binary logic gates and dynamic analog signal-processing functions. To demonstrate the utility of their system, the authors linked Ras oncogene activation to the system to

trigger Caspase3 (Cas3) mediated cell killing. Specifically, Ras activation leads to reconstitution of a TEVp which in turn cleaves and converts a membrane-bound Cas3 from inactive to active state. To reduce undesired activation of Cas3 in OFF-state (leakiness of the systems), a TVMVp cleavage site placed between membrane-tag and Cas3 remove the latter from the membrane, attenuating its activation by TEVp.

The aforementioned work inspired other groups to continue to refine the design of proteases- and RBPs-based regulatory networks. Recently, Han *et al* used a split TEVp with MS2-TCS-cNOT7 to implement several logic gates and combined the design with the use of degrons to further control the RBP activity [125].

Another direction that is being explored by several research groups, is the use of phosphoregulated orthogonal signals [126–128]. As compared to the systems based on proteolytic events, the advantage of post-translational regulation by phosphorylation is that it is reversible. Phosphorylation has been used to implement synthetic network that are able to emulate the fast dynamics of cellular decision making during development, which are difficult to model with transcriptional regulation [126]. More recently, it has been used to reversibly induce dimerization of two proteins [127]. The use of phosphoregulation allows robust control of circuits outputs and stands as promising tool for therapeutic and diagnosis applications, as demonstrated by the Weiss lab at MIT who filed a patent for a phosphorylation-based miRNA sensor in which the phosphorylation of the transcription factor that drives the output is under endogenous miRNA regulation [128].

2.3.1 Future directions

Synthetic networks in mammalian cells are rapidly evolving towards precise control of gene expression. It has been recently demonstrated that synthetic mRNAs harboring cell type specific microRNAs can be used to efficiently purify human pluripotent stem cell (hPSC)-derived populations [129]. Similarly, we envision that protease responsive RBPs may represent an effective means to link detection of specific cellular states such as infection or cancer to conditional expression of output genes. For example, the SUMO1/sentrin specific protease SENP1 has a pro-oncogenic role in several types of cancer (including pancreatic ductal adenocarcinoma (PDAC), prostate cancer, hepatocellular carcinoma and metastatic neuroblastoma tissues), correlating with poor prognosis [130–132]. Ho *et*

al used SENP1 to modulate the cytotoxic protein Granzyme B (GrB) [133]. Specifically, GrB is fused at the N-terminus to the small ubiquitin like modifiers (SUMO1) cleaved by SENP1. Overexpression of the latter leads to SUMO cleavage, GrB activation and killing of cancer cells. The limit of this design is that the inhibition by these peptides works only when the activity of the effector protein depends on a free N-terminus. Nevertheless, circuits such as ours that sense SENP1 or other endogenous proteases with clinical relevance could be used to reprogram cancer cells to express immunomodulators, or trigger specific cell killing by modifying circuit output in an application-tailored fashion.

Chapter 3

Gene expression burden in mammalian cells

This chapter is about our extensive work on the characterization of gene expression burden in mammalian cells. As previously demonstrated in bacteria [64], we show that transiently expressed genes in mammalian cells compete for limited transcriptional and translational resources. This competition results in the coupling of otherwise independent exogenous and endogenous genes, creating a divergence between intended and actual function. Gene expression burden is therefore one of the greatest limitations in synthetic biology, where absolute predictability of the output is paramount. For this reason, we followed three main paths to elucidate burden and its effects in mammalian cells: first, we characterized resource competition in several experimental setups and we identified a simple circuit topology to mitigate burden (3.2); second, we investigated the molecular mechanisms that lay behind intracellular resource re-allocation as a consequence of microRNA (miRNA) action (3.3); third, we explored the global changes in the transcriptome and proteome of burdened cells, to chase the weak spots and work towards developing burden-resistant cell lines (3.4).

3.1 Gene expression burden: an overview

As anticipated in the Introduction (Section 1.2.2), unpredictable behavior of gene circuits and erratic response of the cells hampers synthetic biology potential. This has been so far observed especially when cells are loaded with high levels of exogenous material to produce [70]. Overexpression of exogenous transgenes has well-known side effects, such as resource overload, stoichiometric imbalance, promiscuous interactions and pathway modulation [134]. Transcription factors (TFs) can additionally cause "squelching"[135], or, in other words, the sequestration of the cofactors required by TFs to exert their function. Together with variability in plasmids uptake in transient DNA delivery [136, 137] and random noise in the system [58], the effects of overexpression lead to unwanted and unpredicted circuits behavior.

Several efforts have been put in place to maximize the robustness of synthetic circuits against noise perturbation, both in bacteria and mammalian cells [59, 60, 70, 138]. The ability of the circuits to achieve robust perfect adaptation relies on negative feedback motifs. Feedback controllers are known to naturally maintain the expression of genes below a certain threshold, however they are unable to finely tune the expression range below the threshold, moreover the dynamics are usually slow. To overcome this limitation, recent implementations use the integral feedback implementation [59, 60] first described by Briat *et al* in 2016 [139]. This strategy is termed antithetic integral feedback and it relies on the so-called annihilation (or sequestration) reaction between the two species that implement the controller.

Even if these circuits achieve desired output behavior, their success is only possible if the production of the two controller species relies on same resource pool (e.g. ribosome pool or polymerases pool). In fact, mRNA and protein production rates are a function of the intracellular resource availability and if the two controller species relied on different pools the stoichiometry of the production would be imbalanced [60].

According to all these observations, from the overexpression problem to the dependence of robustness of resource pools, a key problem arises: predictability of synthetic circuits in living systems is highly impaired by the dependence on the availability of intracellular resources. The competition for cellular resources has been first defined by Ceroni

et al in 2015 as gene expression burden [63]. They identified the problem of non-linear response of synthetic circuits in *E. coli* in the non-regulatory interactions generated among genes by the competition for resources necessary for their expression.

3.1.1 Gene expression burden in bacteria and yeast

After the first observation of gene expression burden in 2015 [63, 65], substantial progress towards increasing the predictability of gene expression has been made in bacteria.

Ceroni *et al* not only showed that exogenous genetic material can be expressed in unexpected ways, but also they showed that it imposes a significant burden on the cellular physiology, resulting in decreased growth rates and decreased cellular performance [63]. This has been attributed to the diversion of the pool of resources available for gene expression [140, 141] towards transcription and translation of the newly introduced synthetic payloads. These observations prompted the development of models that consider gene expression in a resource-limited context [65, 66, 142, 143] and led to approaches for mitigating the impact of resource burden in bacteria [64, 72].

One of the first modelling approaches was described by Qian *et al* in 2017, who characterized a library of activation cascades and demonstrated the burden related to their expression along with non-regulatory interactions with other genes expression in *E. coli* [66]. These non-regulatory interactions take into account availability of RNA polymerase (RNAP) and ribosomes as major limitations.

In 2018, Ceroni *et al* suggested the first approach to limit burden: they identified *via* RNA-seq the genes that are upregulated in overloaded cells and used their promoters to activate a negative feedback control on transgene expression [64], *de facto* limiting the competition for cellular resources.

Analogous studies in *Saccharomyces cerevisiae* showed that transcription and translation are limiting processes [144]. For example, the use of potent transactivators - such as the DOX-inducible rtTA - causes a squelching shortage of general transcription factors for native gene expression in yeast [145].

In mammalian cells, while performance shortcomings of synthetic circuits due to transactivator dosage and plasmid uptake variation have been observed [70], a deeper understanding of the problem of resource burden and methods for its mitigation were still

missing. Our work towards a deep understanding of gene expression burden was carried out following three main paths: *(i)* the characterization and mitigation of gene expression burden using optimal genetic circuit design (Section 3.2, in collaboration with Professor Khammash' lab at ETH and Professor Stan's lab at ICL) [67]; *(ii)* the elucidation of the key molecular mechanisms involved in miRNA-dependent resource re-allocation (Section 3.3, in collaboration with Professor Stan's lab at ICL and Dr Gabriella Viero at CIBIO, Trento); *(iii)* the understanding of the global changes induced in the intracellular environment by the imposition of genetic payloads to build controllers to reduce burden (Section 3.4, in collaboration with Professor Di Bernardo's lab at Tigem, Naples and Dr Armirotti's lab, IIT). Our extensive and detailed study paves the way towards more rational genetic circuits design to minimize unwanted burden drawbacks, such as gene expression coupling and impaired cellular physiology.

3.2 Characterization, modelling and mitigation of gene expression burden

Frei*, Cella* *et al*, Nature Communications (2020)

We characterized the effect of resource competition in several experimental setups, highlighting how both transcriptional and translational resources represent a bottleneck for reliability and predictability of synthetic circuits. We identified miRNAs as a good candidate to minimize burden (**Fig. 3.1a**). We implemented an ODE-based mathematical model that guided us towards the choice of an incoherent feed-forward loop motif that stabilizes a gene of interest against other genes' fluctuations (**Fig. 3.1b-c**). With this project we paved the way towards more critical genetic circuits design, establishing an important milestone for the field.

3.2.1 Gene expression burden characterization

To demonstrate the interplay between co-expressed genes we explored different experimental setups in H1299 cells (non-small lung cancer cells) in which we tuned an eXtra Transgene (hereafter named *X-tra*) against a fixed amount of another constitutively expressed fluorescent protein (hereafter named *capacity monitor*). The rationale is that the *capacity monitor* represents the ability of the host cells to sustain genes' production, against the extra load that we impose by delivering and tuning *X-tra*.

First, we observed that by increasing *X-tra* expression against a fixed amount of *capacity monitor*, the latter decreases (**Fig. 3.2a**). For this experiment the mKate under a CMV promoter acts as *capacity monitor*, while the *X-tra* is an EGFP under the same constitutive promoter. The equimolar ratios between the two genes span from 1:1 to 1:2.5 in steps of 0.5. Cells were analyzed 48 hours post-transfection using flow cytometry and the geometric mean of the two fluorescences was calculated to infer the level of protein expression.

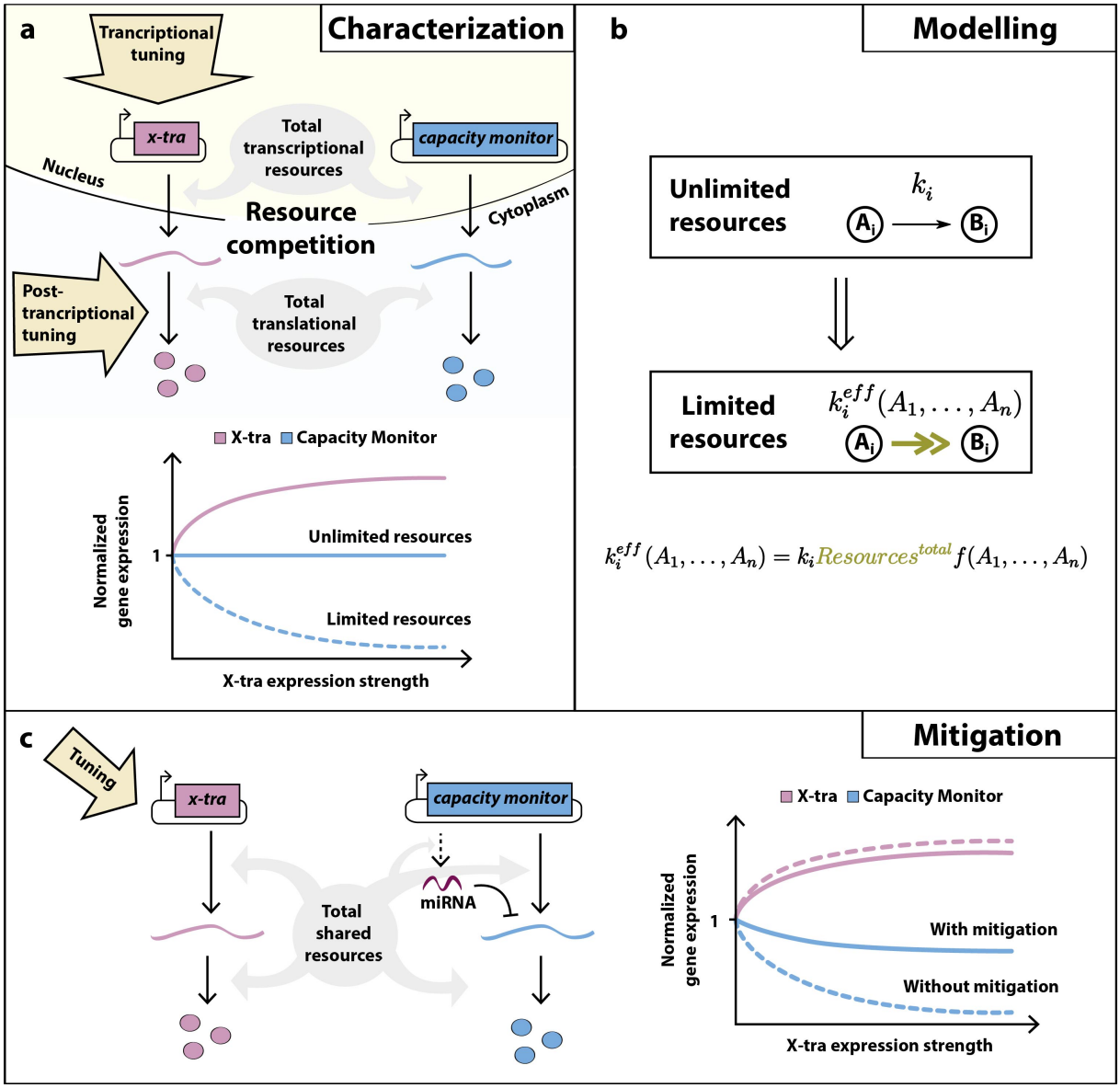


Figure 3.1: **Resource sharing and the origin of gene expression burden (a) Characterization of gene expression burden.** Expression of independent exogenous genes impacts on host cellular resources. Thus, perturbations in one gene's expression (*X-tra*) affect the expression of a second gene (*capacity monitor*). **(b) Modelling of gene expression in a resource limited environment.** Modelling of gene expression is generally performed under the assumption of unlimited resources. A simple framework enables the straightforward transformation of such a model to a system that incorporates resources explicitly. The transformation involves a simple function that scales the original reaction rate. **(c) Mitigation of gene expression burden.** A simple microRNA-based circuit motif is capable of mitigating the burden-induced coupling of *X-tra* and the *capacity monitor*.

We used the same samples to extract mRNA and measure the mRNA level of *X-tra* and *capacity monitor*. The qPCR quantification was then indicative for the transcriptional

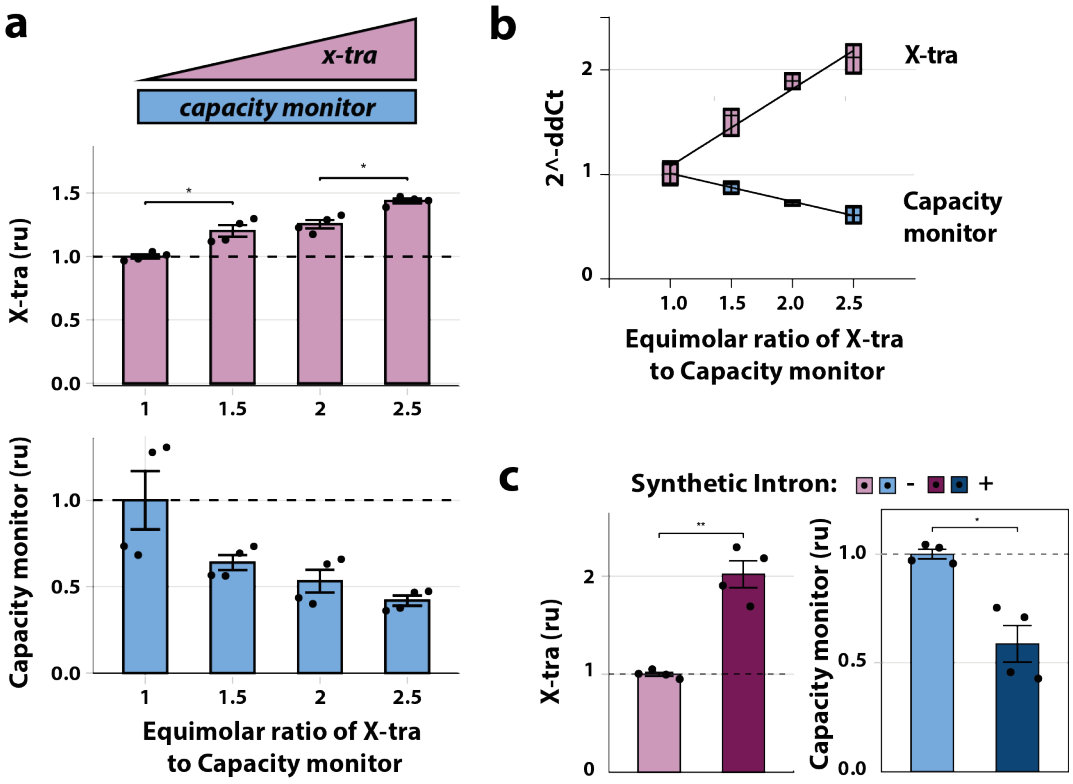


Figure 3.2: **Gene expression burden in H1299 cells. (a)** Flow cytometry results and correspnding mRNA quantification **(b)** of cells co-transfected with fixed amount of *capacity monitor* (mKate) and increasing amount of *X-tra* (EGFP; 1:1 to 1:2.5 molar ratio). As *X-tra* increases, both the protein and the mRNA levels of the *capacity monitor* decrease, due to resource competition. Transcriptional burden is detected with mRNA quantification. Data show the mean fluorescence and mRNA level normalized to their value at a plasmid molar ratio of 1. **(c)** Flow cytometry results of cells co-transfected with *X-tra* (mKate) which includes or not a synthetic intron in the 5'UTR, and *capacity monitor* (EGFP). Data show that when mKate translation is enhanced by the synthetic intron, EGFP levels decrease, therefore detecting translational burden. ru: relative units. Unpaired two-sided T-test. p-value: **<0.005, *<0.05.

burden. Consistently with the previous results, we observed a decrease in the mRNA of the *capacity monitor* as we increased the amount of *X-tra* (**Fig. 3.2b**). To tackle translational burden, we used again two constitutively expressed fluorescent proteins under the CMV promoter regulation and we inserted a synthetic intron to the 5'UTR of *X-tra* (mKate in this case) that enhances the translocation of the mRNA to the cytoplasm leading to increased translation. As compared to the control condition (no synthetic intron), the level of a co-transfected *capacity monitor* (EGFP in this case) decreases (**Fig. 3.2c**).

Further, burden was observed also when gene regulation was triggered by different promoters. Specifically, the *capacity monitor* was placed under PGK promoter regulation.

PGK promoter has weaker activity than CMV in H1299, but not in HEK293T cells (**Fig. 3.3a**). By comparing gene expression patterns in the two cells lines, we observed the same qualitative trend previously shown with the *capacity monitor* under CMV promoter (**Fig. 3.3b-e**).

As we consistently observed gene expression coupling in these and other experimental setups (data not shown, refer to [67]), we asked whether there was an impact on endogenous gene expression. First, we used the same mRNA extracted for the experiment shown in **Fig. 3.2** to measure the mRNA levels of 3 endogenous genes (CyCA2, eIF4E and GAPDH) and observed a progressive, albeit not dramatic decrease with higher amounts of *X-tra* when compared to the 1:1 ratio (**Fig. 3.4a**). To increase the sensitivity of our analysis on endogenous genes, we transfected H1299 cells with a plasmid encoding EGFP and mKate under the control of a bidirectional CMV promoter. We then sorted transfected cells according to high and intermediate levels of fluorescent markers, as well as non-transfected cells (absence of fluorescence) (**Fig. 3.4b**). We then quantified the mRNA levels of CyCA2, eIF4E, GAPDH (**Fig. 3.4c**). Notably, in transfected cells that express high and intermediate levels of EGFP and mKate, the expression of CyCA2, eIF4E and GAPDH decreases as compared to the non-transfected population.

3.2.2 Post-transcriptional regulators re-allocate resources

Post-transcriptional gene regulation (PTGR) is a molecular mechanism by which gene expression is finely tuned addressing the neo transcribed mRNAs. In mammalian cells, mRNAs are first edited by splicing, polyadenylation and capping [146]. All these steps happen in the nucleus and contribute to determine mRNA half-life and stability. After that, contribution to PTGR is given by non-coding RNAs (ncRNAs) and RNA-binding proteins (RBPs). The non-coding RNA species involved are long ncRNAs, piwi-interacting RNAs, short interfering RNAs and, above all, microRNAs (miRNAs). We hypothesized that by downregulating target gene expression, both miRNAs and RBPs could free up resources to the benefit of other genes' expression.

We tested our hypothesis using both RBPs and miRNAs. In the RBP-resources experiments we used the RBPs L7Ae, that binds to target sequences in the 5'UTR of the transcript blocking mRNA translation, and Ms2-cNOT7 that binds to target sequences

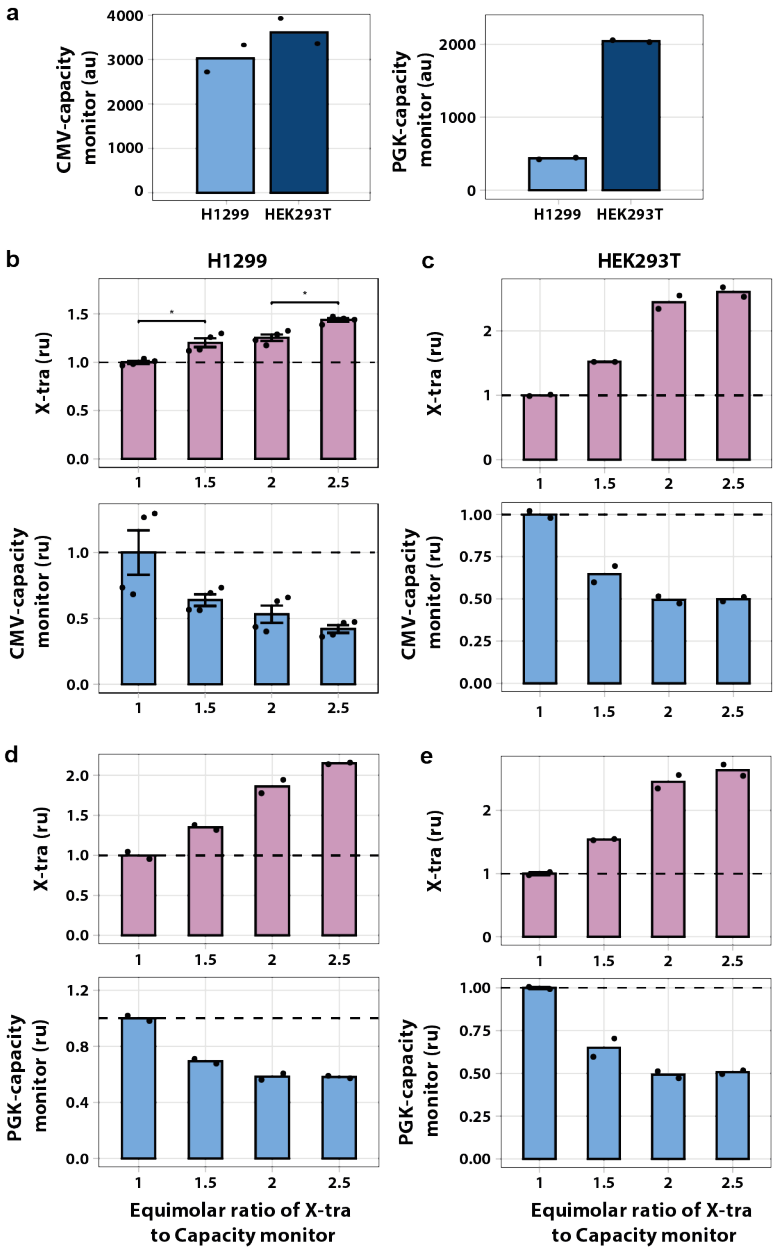


Figure 3.3: **Relation of *X-tra* and *capacity monitor* expression in H1299 and HEK293T cell lines using CMV and PGK promoter. (a)** Levels of fluorescence driven by the same promoter (CMV or PGK) differ across cell lines. Data show absolute units of capacity monitor detected by flow cytometry in 1:1 molar ratio transfection. N=2 biological replicates. Flow cytometry results of H1299 (N=4 biological replicates) **(b)** and HEK293T (N=2 biological replicates) **(c)** cells co-transfected with fixed amount of CMV-mKate (*capacity monitor*) and increasing amount of EGFP (*X-tra*) under CMV promoter regulation (molar ratio from 1:1 to 1:2.5). Flow cytometry results of H1299 (N=2 biological replicates) **(d)** and HEK293T (N=2 biological replicates) **(e)** cells co-transfected with fixed amount of mKate (*capacity monitor*) under PGK promoter regulation and increasing amount of EGFP (*X-tra*) under CMV promoter regulation (molar ratio from 1:1 to 1:2.5). Data show the geometric mean fluorescence normalized to its value at a plasmid molar ratio of 1 +/- SE. SE: standard error. au: arbitrary units. ru: relative units. Unpaired two-sided T-test. p-value: * < 0.05.

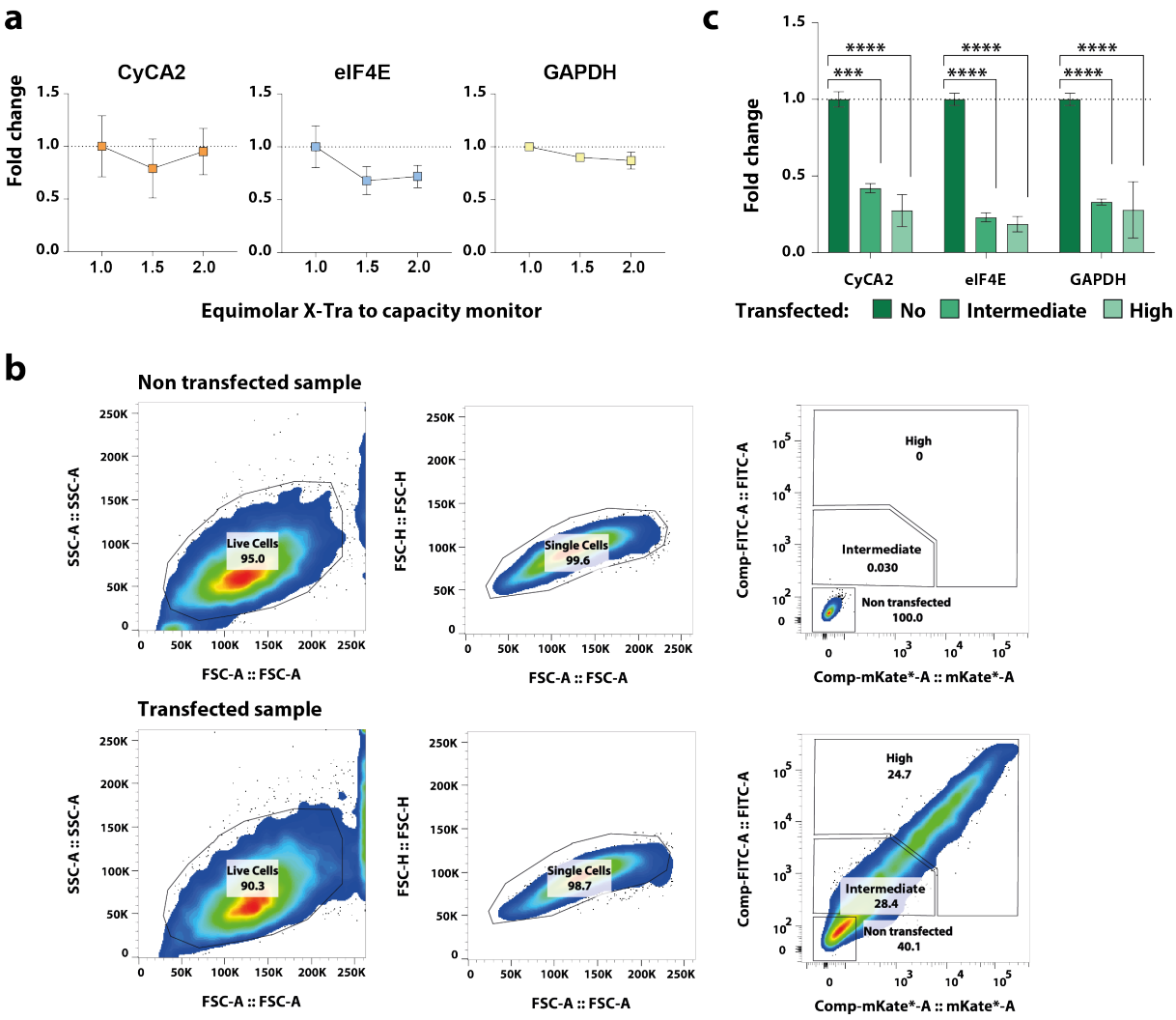


Figure 3.4: **Impact of genetic payloads on endogenous genes.** **a** Effect of *X-tra* titration on endogenous genes. We measured CyCA2, eIF4E and GAPDH mRNA levels by qPCR in the samples shown in Fig. 2c at 1.0, 1.5 and 2.0 molar ratios. Data represent the mean value normalized to the equimolar ratio of 1.0. Error bars represent the standard error, SE. N=4 biological samples for CyCA2 and GAPDH. N=2 biological samples for eIF4E. **b** Sorting strategy. H1299 cells were transfected with a plasmid encoding the fluorescent proteins EGFP and mKate, expressed from a bidirectional promoter. Cells were sorted by fluorescence intensity 48 hours post-transfection to collect non-transfected, intermediate and high transfected cells from the same transfection plate. First, gates to select live and single cells were determined (left and middle plots). Then, the threshold for fluorescent intensity was set using a non-transfected sample as reference (top, right). The two additional gates to collect intermediate and high transfected cells were created as shown in the plots on the right. **c** mRNA levels expressed from endogenous genes decrease in cells with intermediate and high fluorescence. Data show fold change +/- SE. SE: standard error. N=3 biological replicates.

in the 3'UTR of the target promoting its de-adenylation and degradation [49, 51]. In this experimental setup, the mRNA of *X-tra* is targeted by the RBPs and is co-expressed with

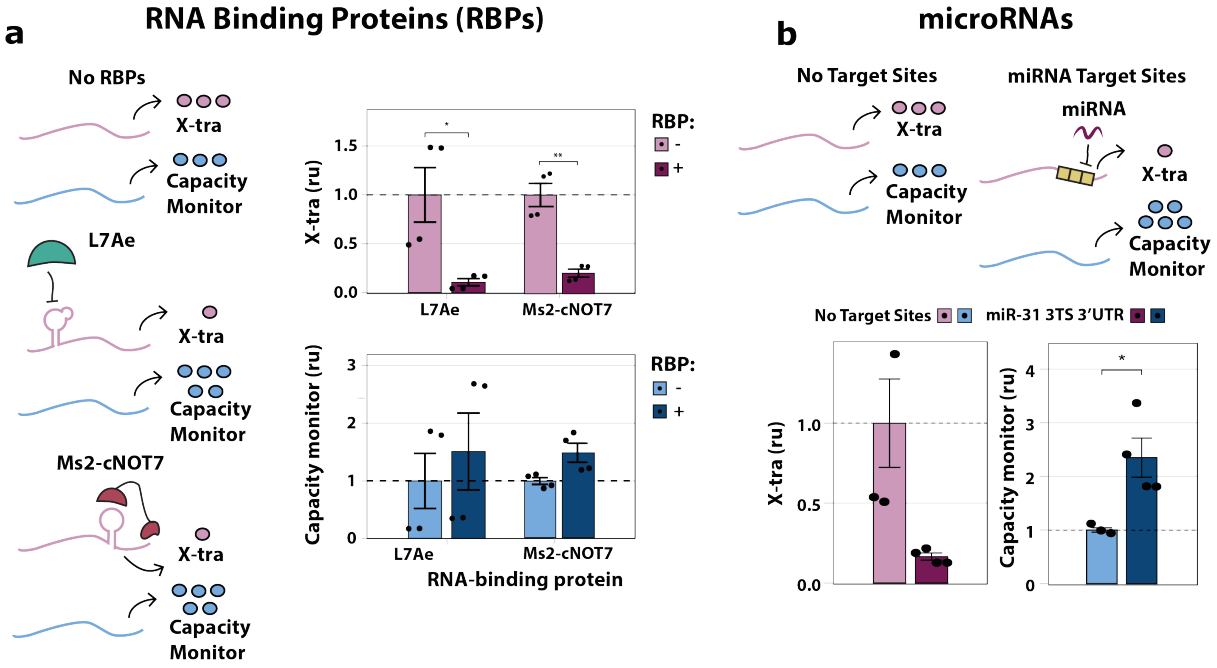


Figure 3.5: **Post-transcriptional regulators re-allocate resources in H1299 cells.** **a** Repressed *X-tra* expression leads to increased *capacity monitor* levels. N=4 biological replicates for L7Ae and N=4 for Ms2-cNOT7. **b** When X-tra is downregulated by miR-31 endogenously expressed in H1299 cells, the *capacity monitor* levels increase. All flow cytometry data were acquired 48 hours post-transfection and are plotted as mean +/- SE. SE: standard error. ru: relative units. N=4 biological replicates. Unpaired two-sided T-test. p-value: *<0.05

a second, constitutively expressed, fluorescent readout (*capacity monitor*). We demonstrated that following *X-tra* downregulation by RBPs the *capacity monitor* levels increase (**Fig. 3.5a**).

miRNAs operate by either translation inhibition or mRNA degradation, according to complete [73] or partial [147] complementarity to the mRNA target (see Section 3.3.1 for further details). To evaluate the effect of miRNA regulation on cellular resources re-allocation, we placed three perfect complementary target sites (TS) in the 3'UTR of *X-tra*, which respond to the endogenous miR-31 highly expressed in H1299 cells. The *capacity monitor* expression levels increased when the *X-tra* mRNA was downregulated by miRNAs, as compared to controls lacking miRNA target sites (**Fig. 3.5b**). Overall, our results indicate that a secondary effect of post transcriptional regulators, along with gene down-modulation, is the redistribution of the resources that allow to increase the expression of other genes.

3.2.3 Gene expression burden is cell-type independent

To demonstrate that the burden imposed by synthetic circuits is cell-type independent, plasmid titrations and co-transfections using the plasmids harboring the synthetic intron/RBPs binding sites/miRNA TS similar to the ones previously carried out in H1299, were performed also in HEK293T, U2OS, HeLa and CHO-K1 cells, obtaining similar results (**Fig. 3.6, 3.7, 3.8, 3.9**). Interestingly, even CHO-K1 cells, which are the workhorses of the biopharmaceutical industry due to their high productive capability [148], are susceptible to gene expression burden. Similarly, redistribution of resources was also observed by the RBPs L7Ae and MS2-cNot7 as well as by the endogenous miR-221 and miR-21 highly expressed in HEK293T/U2OS [73] and HeLa/CHO-K1 cells respectively.

These results confirm that burden and intracellular resources redistribution are cell-context independent. The extent of negative correlation between *X-tra* and *capacity monitor* expression, as well as the amount of repression by post-transcriptional regulators differs across cell lines; this could be the consequence of several factors, such as the relative abundance of transcriptional, post-transcriptional and translational resources.

3.2.4 Characterization of miRNA role in gene expression burden

A major advantage of miRNAs over RBPs is that they are endogenously expressed and cell line specific. Thus, their expression does not impose additional burden, and circuits' design can be easily tailored to the cell/tissue of interest. Based on the results shown in this chapter, we envision that genetic circuits that mitigate resource competition via miRNAs may be designed for any mammalian cell line with a very broad set of potential applications.

We sought to characterize the correlation between miRNA-mediated downregulation and resource redistribution, by building a library of *miRNA sensors* for miR-31, which is endogenously expressed in H1299 lung cancer cells [149]. The *miRNA sensor* is composed of the fluorescent reporter mKate with or without miR-31 TS, encoded along with the *capacity monitor* (EGFP) on a single plasmid with a bidirectional promoter

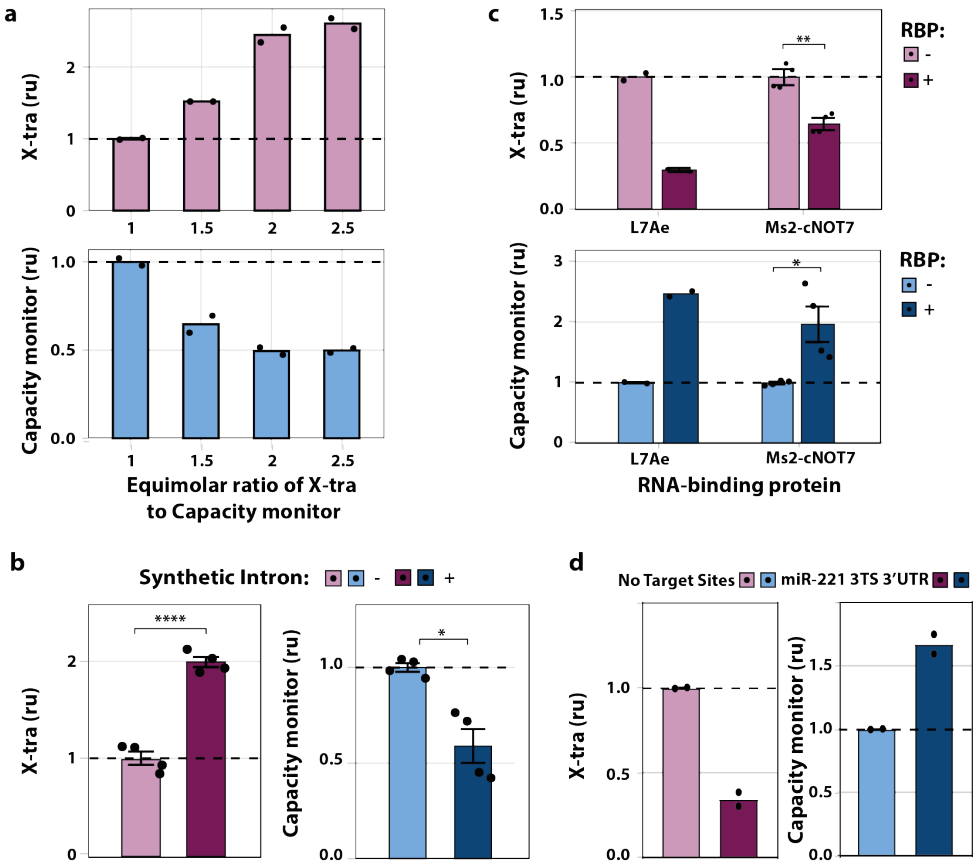


Figure 3.6: **Gene expression burden in HEK293T cells. (a)** Flow cytometry results of HEK293T cells co-transfected with fixed amount of *capacity monitor* and increasing amount of *X-tra* (1:1 to 1:2.5 molar ratio), both under CMV promoter regulation. Data show the mean fluorescence normalized to its value at a plasmid molar ratio of 1. N=4 biological replicates. **(b)** Flow cytometry results of HEK293T cells co-transfected with *X-tra* (mKate) which includes or not a synthetic intron in the 5'UTR, and *capacity monitor* (EGFP). Data show that when mKate expression is enhanced by the synthetic intron, EGFP levels decrease. Data are the mean fluorescence normalized to fluorescence values in the absence of the intron. N=4 biological replicates. **(c)** Flow cytometry results of HEK293T cells co-transfected with 2kturn-EGFP or EGFP-8xMs2 (*X-tra*) and mKate (*capacity monitor*) in presence or absence of L7Ae or Ms2-cNOT7 respectively. Data show that when *X-tra* is downregulated, the *capacity monitor* levels increase. Plot represents normalization of mean fluorescence values to the condition without RBP. N=2 biological replicates for L7Ae and N=4 for Ms2-cNOT7. **(d)** Flow cytometry results of HEK293T cells co-transfected with mKate (*X-tra*) that includes or not miR-221 target sites in the 5'UTR, and EGFP (*capacity monitor*). Data show that *capacity monitor* levels are higher when the *X-tra* is downregulated by miR-221. N=4 biological replicates. Plot represents normalization of geometric mean fluorescence values to the no target site condition. Data were acquired 48 hours post-transfection. Error bars represent the standard error. ru: relative units. Unpaired two-sided T-test. p-value: ****<0.0001, **<0.005, *<0.05.

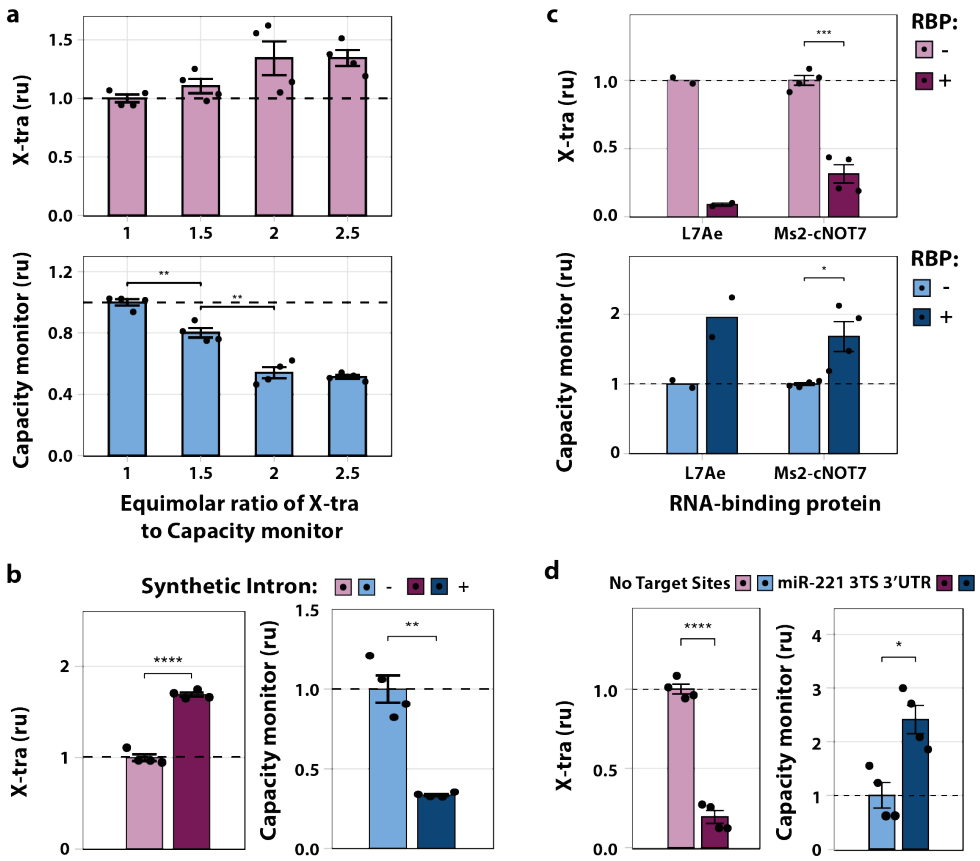


Figure 3.7: **Gene expression burden in U2OS cells.** **(a)** Flow cytometry results of U2OS cells co-transfected with fixed amount of *capacity monitor* and increasing amount of *X-tra* (1:1 to 1:2.5 molar ratio), both under CMV promoter regulation. Data show the mean fluorescence normalized to its value at a plasmid molar ratio of 1. N=4 biological replicates. **(b)** Flow cytometry results of U2OS cells co-transfected with *X-tra* (mKate) which includes or not a synthetic intron in the 5'UTR, and *capacity monitor* (EGFP). Data show that when mKate expression is enhanced by the synthetic intron, EGFP levels decrease. Data are the mean fluorescence normalized to fluorescence values in the absence of the intron. N=4 biological replicates. **(c)** Flow cytometry results of U2OS cells co-transfected with 2kturn-EGFP or EGFP-8xMs2 (*X-tra*) and mKate (*capacity monitor*) in presence or absence of L7Ae or Ms2-cNOT7 respectively. Data show that when *X-tra* is downregulated, the *capacity monitor* levels increase. Plot represents normalization of mean fluorescence values to the condition without RBP. N=2 biological replicates for L7Ae and N=4 for Ms2-cNOT7. **(d)** Flow cytometry results of U2OS cells co-transfected with mKate (*X-tra*) that includes or not miR-221 target sites in the 5'UTR, and EGFP (*capacity monitor*). Data show that *capacity monitor* levels are higher when the *X-tra* is downregulated by miR-221. N=4 biological replicates. Plot represents normalization of geometric mean fluorescence values to the no target site condition. Data were acquired 48 hours post-transfection. Error bars represent the standard error. ru: relative units. Unpaired two-sided T-test. p-value: **<0.005, *<0.05.

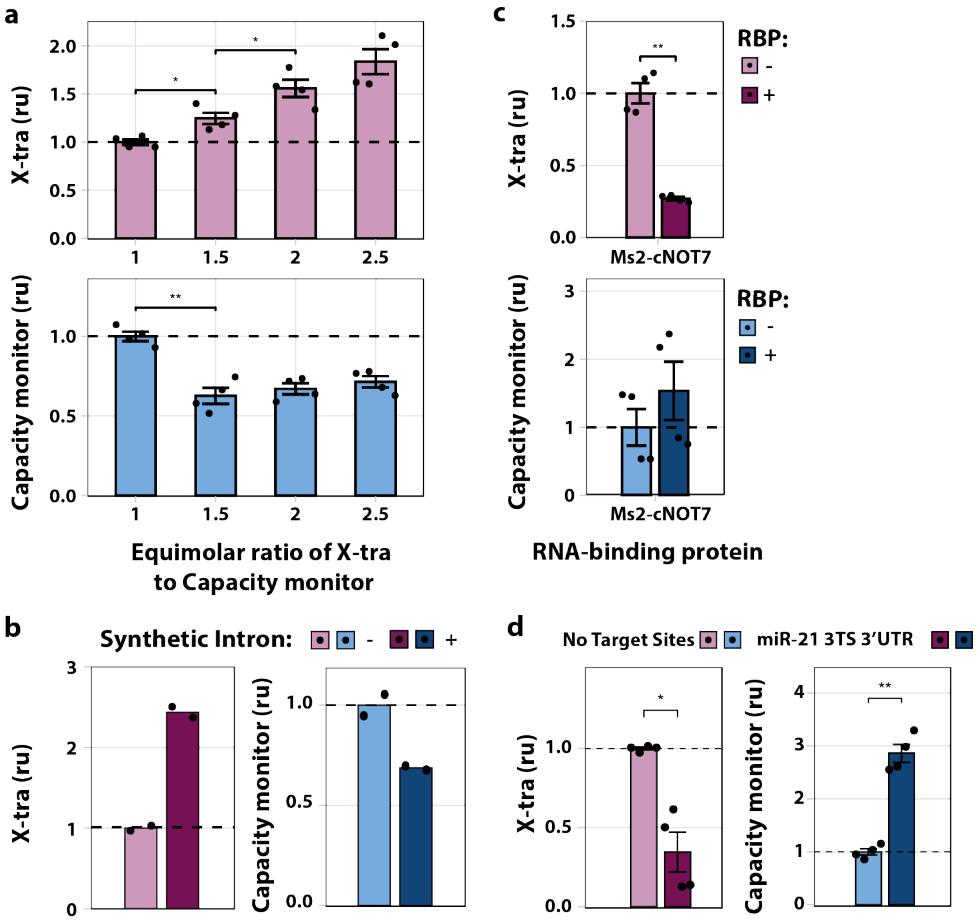


Figure 3.8: **Gene expression burden in HeLa cells.** **(a)** Flow cytometry results of HeLa cells co-transfected with fixed amount of *capacity monitor* and increasing amount of *X-tra* (1:1 to 1:2.5 molar ratio), both under CMV promoter regulation. Data show the mean fluorescence normalized to its value at a plasmid molar ratio of 1. N=4 biological replicates. **(b)** Flow cytometry results of HeLa cells co-transfected with *X-tra* (mKate) which includes or not a synthetic intron in the 5'UTR, and *capacity monitor* (EGFP). Data show that when mKate expression is enhanced by the synthetic intron, EGFP levels decrease. Data are the mean fluorescence normalized to fluorescence values in the absence of the intron. N=2 biological replicates. **(c)** Flow cytometry results of HeLa cells co-transfected with EGFP-8xMs2 (*X-tra*) and mKate (*capacity monitor*) in presence or absence of Ms2-cNOT7. Data show that when *X-tra* is downregulated, the *capacity monitor* levels increase. Plot represents normalization of mean fluorescence values to the condition without RBP. N=4. **(d)** Flow cytometry results of HeLa cells co-transfected with mKate (*X-tra*) that includes or not miR-21 target sites in the 5'UTR, and EGFP (*capacity monitor*). Data show that *capacity monitor* levels are higher when the *X-tra* is downregulated by miR-21. N=4 biological replicates. Plot represents normalization of geometric mean fluorescence values to the no target site condition. Data were acquired 48 hours post-transfection. Error bars represent the standard error. ru: relative units. Unpaired two-sided T-test. p-value: **<0.005, *<0.05.

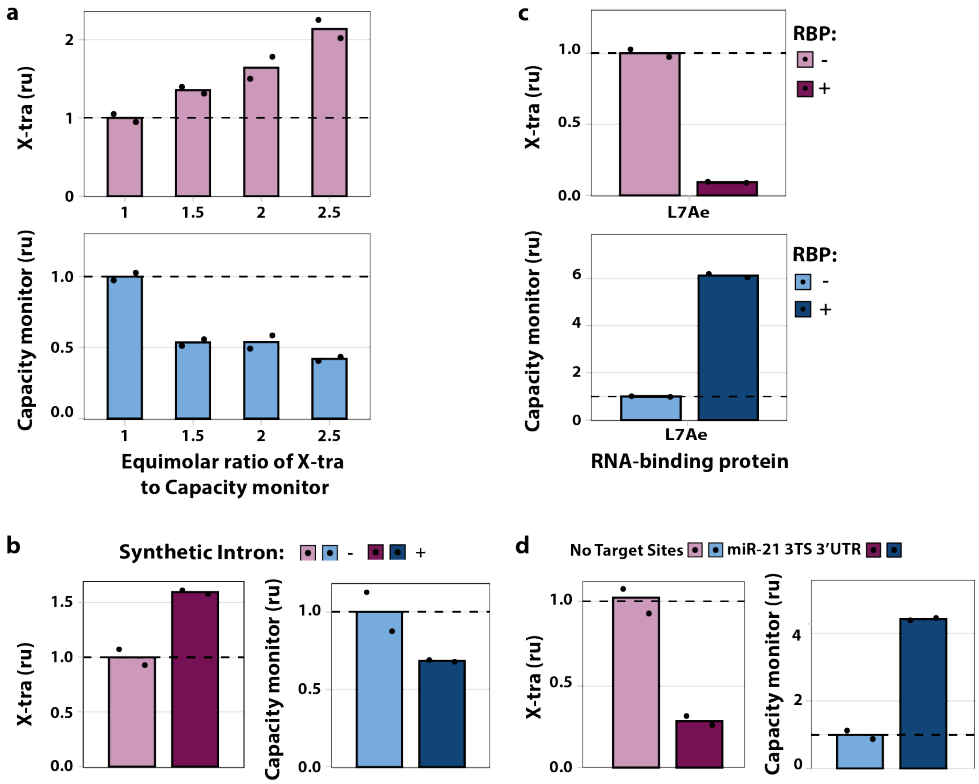


Figure 3.9: **Gene expression burden in CHO-K1 cells. (a)** Flow cytometry results of cells co-transfected with fixed amount of *capacity monitor* and increasing amount of *X-tra* (1:1 to 1:2.5 molar ratio), both under CMV promoter regulation. Data show the mean fluorescence normalized to its value at a plasmid molar ratio of 1. N=2 biological replicates. **(b)** Flow cytometry results of CHO-K1 cells co-transfected with *X-tra* (mKate) which includes or not a synthetic intron in the 5'UTR, and *capacity monitor* (EGFP). Data show that when mKate expression is enhanced by the synthetic intron, EGFP levels decrease. Data are the mean fluorescence normalized to fluorescence values in the absence of the intron. N=2 biological replicates. **(c)** Flow cytometry results of CHO-K1 cells co-transfected with 2kturn-EGFP (*X-tra*) and mKate (*capacity monitor*) in presence or absence of L7Ae. Data show that when *X-tra* is downregulated, the *capacity monitor* levels increase. Plot represents normalization of mean fluorescence values to the condition without RBP. N=2 biological replicates. **(d)** Flow cytometry results of CHO-K1 cells co-transfected with mKate (*X-tra*) that includes or not miR-21 target sites in the 5'UTR, and EGFP (*capacity monitor*). Data show that *capacity monitor* levels are higher when the *X-tra* is downregulated by miR-21. N=2 biological replicates. Plot represents normalization of geometric mean fluorescence values to the no target site condition. Data were acquired 48 hours post-transfection.

(**Fig. 3.10a**). The target sequence for the miRNA is taken from miRBase database (http://www.mirbase.org/) and is designed as fully complementary to the miRNA. We used fully complementary target sites to maximize the miRNA activity [73]. We implemented our system to have plasmids with 1 or 3 Target Sites (TS) either at 3' or 5' UTR, to gain more insights into miRNA mechanism of action depending on the number and location of TS. The negative control is a plasmid with no miRNA TS. We performed transient transfections of our genetic constructs, and measured protein levels both for *miRNA sensor* and *capacity monitor* using flow cytometry. Our results show that *miRNA sensor* protein levels decrease when miR-31 TS are present in the UTR (**Fig. 3.10b**) and the result is consistent in all the experimental setups. *miRNA sensor* levels are plotted in bar charts for a clear visualization of the protein downregulation as compared to the control. As *miRNA sensor* is downregulated, *capacity monitor* expression significantly increases (**Fig. 3.10b**), corroborating the hypothesis that genetic payloads expression depends on available resources.

To further prove that the *capacity monitor* increase responds to *miRNA sensor* downregulation by miR-31, we co-transfected miR-31 inhibitors (mirVana) and evaluated *miRNA sensor* and *capacity monitor*. The results (**Fig. 3.10c**) show that miRNA-31 inhibitor restores *miRNA sensor* protein levels, and *capacity monitor* protein level decreases accordingly, confirming that their expression levels are linked. Interestingly, the effect of the miRNA inhibitor was more pronounced with TS placed in the 3'UTR. Synthetic miRNA inhibitors bind to endogenous miRNAs in an irreversible manner [110], but differences in their action (e.g. when TS are placed in the 3' vs 5' UTR), as well as mechanistic insights into these differences, are still missing.

To confirm that our results are plasmid design independent, we tested the same system co-transfecting *miRNA sensor* and *capacity monitor* encoded on two different plasmids (**Fig. 3.11a**).

As we showed in **Fig. 3.6d, 3.7d, 3.8d, 3.9d**, the ability of miRNAs to reallocate resources is cell-type independent. To further prove it, we built a library of *miRNA sensors* specific for miR-221 and miR-21 and tested them in U2OS and HeLa cells respectively. In all cases, we confirmed the negative correlation between *miRNA sensor* and *capacity monitor* (**Fig. 3.11b-c**).

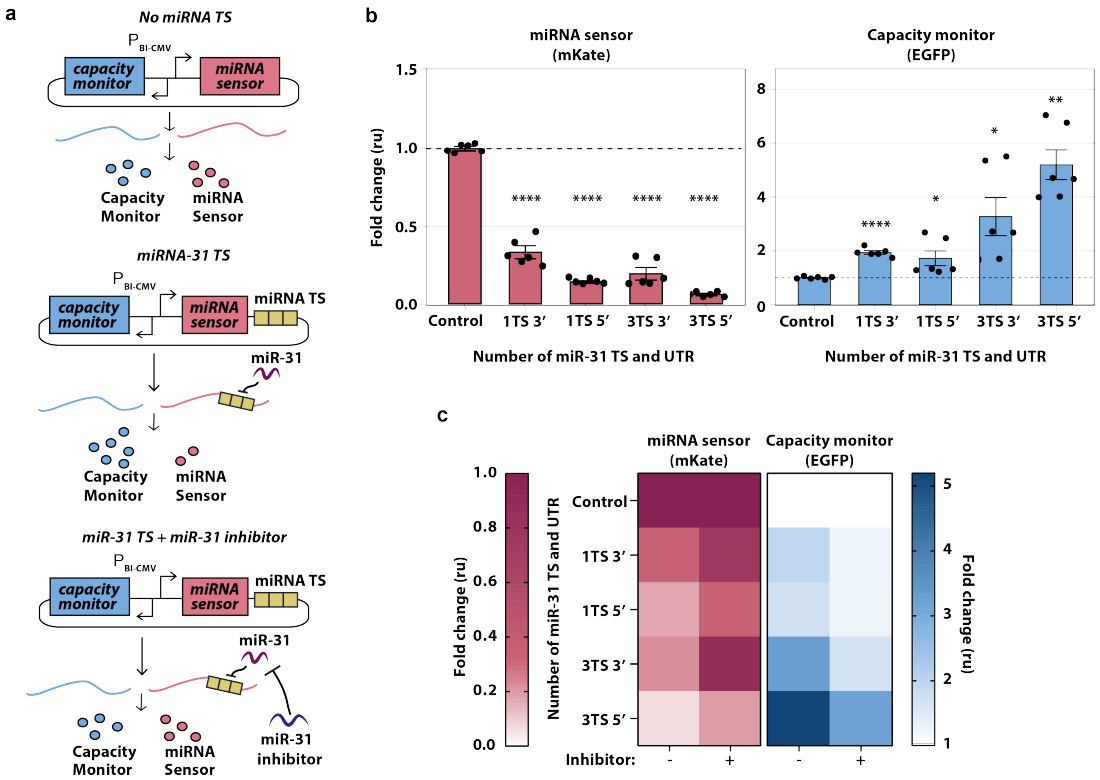


Figure 3.10: **Impact of miRNA target sites number and location on burden. (a)** Schematics of experimental design to infer miRNA-mediated cellular resources redistribution. EGFP *capacity monitor* and mKate (*miRNA sensor*) are encoded on the same bi-directional CMV promoter plasmid. 1 or 3 TS for miR-31 (TS) are added either in the 3' or 5'UTR of mKate. Control: no miR-31 TS. Hypothesis: in the absence of miR-31 regulation, *capacity monitor* and miRNA sensor are expressed to a certain level (top). In the presence of miR-31, lower miRNA sensor levels correlate with higher *capacity monitor* expression (middle). This condition is reversed by a miR-31 inhibitor (bottom). **(b)** Fold change of miRNA sensor and *capacity monitor* protein levels compared to control (set to 1). EGFP increases up to 5 fold with the strongest downregulation of mKate (3TS 5'UTR). Flow cytometry data were acquired 48 hours post-transfection and are plotted as mean +/- SE. SE:standard error. ru: relative units. N=6 biological replicates. Unpaired two-sided T-test. p-value: ****<0.0001, ***<0.0005, **<0.005, *<0.05. **(c)** When miR-31 activity was impaired by a miR-31 inhibitor, the rescue of mKate expression corresponds to reduced EGFP levels, whereas both fluorescent proteins do not vary in the control. The heat maps represent the fold change derived by flow cytometry data, calculated as the ratio between the geometric mean of 6 biological replicates and the corresponding geometric mean in the control condition.

Finally, we demonstrated that the choice of the miRNA to employ to redistribute resources has to be in agreement with the intracellular environment. We tested a miR-31 *miRNA sensor* in HEK293T and in U2OS and we could not detect any protein variation as compared to the control (**Fig. 3.11d**).

Overall, these data suggest that the number and location of TS can be tuned to achieve desired protein expression levels, therefore miRNAs can be used to develop resource-aware circuit topologies that minimize gene expression coupling.

3.2.5 Modelling and mitigation of gene expression burden

To explain the non-intuitive behavior of our genetic circuits, in collaboration with the Control Theory and Systems Biology lab lead by Prof. Mustafa Khammash at ETH and with the Prof. Stan's group at Imperial College London, we implemented a modelling framework able to capture gene expression in resource-limited environments (**Fig. 3.1b, 3.12a**, Appendix A). We successfully tested the modelling framework on some of our experimental results (**Fig. 3.12b-c**) and on experimental data presented in Lillacci *et al.* [70] (data not shown, refer to [67]).

Next, we implemented a strategy that exploits miRNAs to reduce the indirect coupling between co-expressed genes. In particular, we took advantage of the fact that miRNA production also requires pre-translational cellular resources, therefore acting as a sensor for resource availability. Thus, it is possible to reduce the coupling between genes co-expressed via a common resource pool by introducing miRNA-mediated repression of those genes (as long as the miRNA itself is also affected by the same resource pool). Since both the miRNA and the miRNA-repressed gene are affected by the availability of resources, miRNA-mediated repression implements an incoherent feedforward loop similar to previously published circuits [70, 150, 151] (**Fig. 3.13a**). Interestingly, this iFFL-based circuit constitutes a biological implementation of the miRNA circuit proposed by Zechner *et al* [152]. In this setting, the miRNA can be interpreted as an estimator of its cellular context (e.g. amount of free resources) and acts to filter out this context, thereby minimizing its impact on the output of interest.

We explored this strategy for the endogenously expressed miR-31 in H1299 cells (**Fig. 3.13b-c**) and a synthetic miRNA encoded on a plasmid (data not shown, refer to [67]). More specifically, **Fig. 3.13b** describes a strategy that exploits endogenous miRNAs to reduce the coupling of a gene of interest (GOI) to the expression level of other genes, introduced by the limitation in resources. Implementation of this strategy only requires adding target sites of an endogenous miRNA to the 5' UTR of the gene of

interest (mKate). In our experimental setup, when the copy number of a second gene (*X-tra*) is increased, resources are drawn away from the expression of mKate and allocated

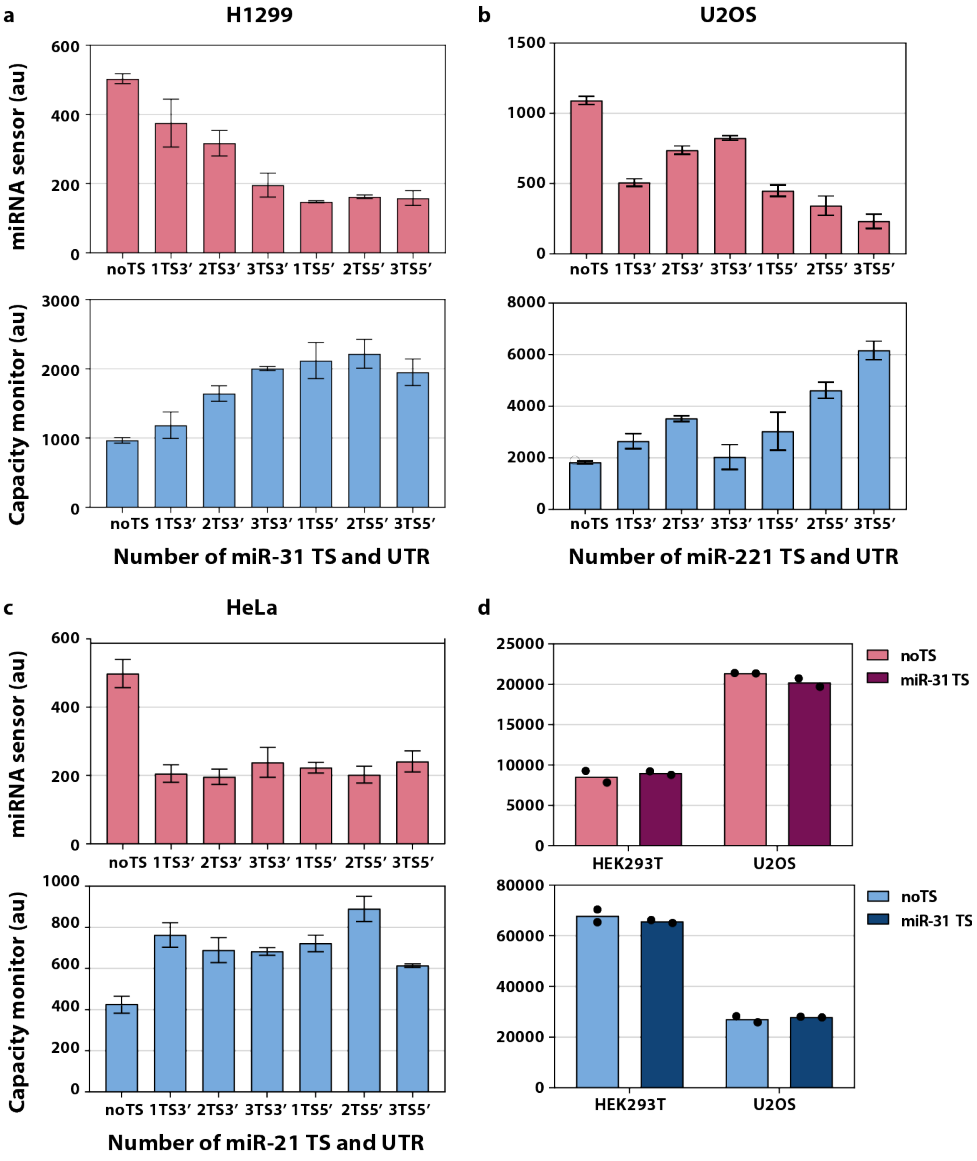


Figure 3.11: **Impact of miRNA target sites number and location on several cell lines. (a,b,c)** Flow cytometry results of miR31, miR221 and miR21 (*miRNA sensors*) co-transfected with EGFP (*capacity monitor*) in H1299, U2OS and HeLa cells respectively show that downregulation of (*miRNA sensors*) results in higher capacity monitor levels also using a different plasmid design as compared to **Fig. 3.10**. **(d)** U2OS and HEK293T cells co-transfected with the 4-TS-3'UTR *miR-31 sensor* and the (*capacity monitor*). Both cell lines do not exhibit high expression of miR-31, consistently data show that both *miRNA sensor* and the *capacity monitor* levels are comparable with and without miR-31 TS, confirming that the higher *capacity monitor* levels are indeed a consequence of miRNA activity. Data were acquired 48h post transfection and are plotted as mean fluorescence +/- SE. SE: standard error. au: arbitrary units. N $\geq$ 2 biological replicates.

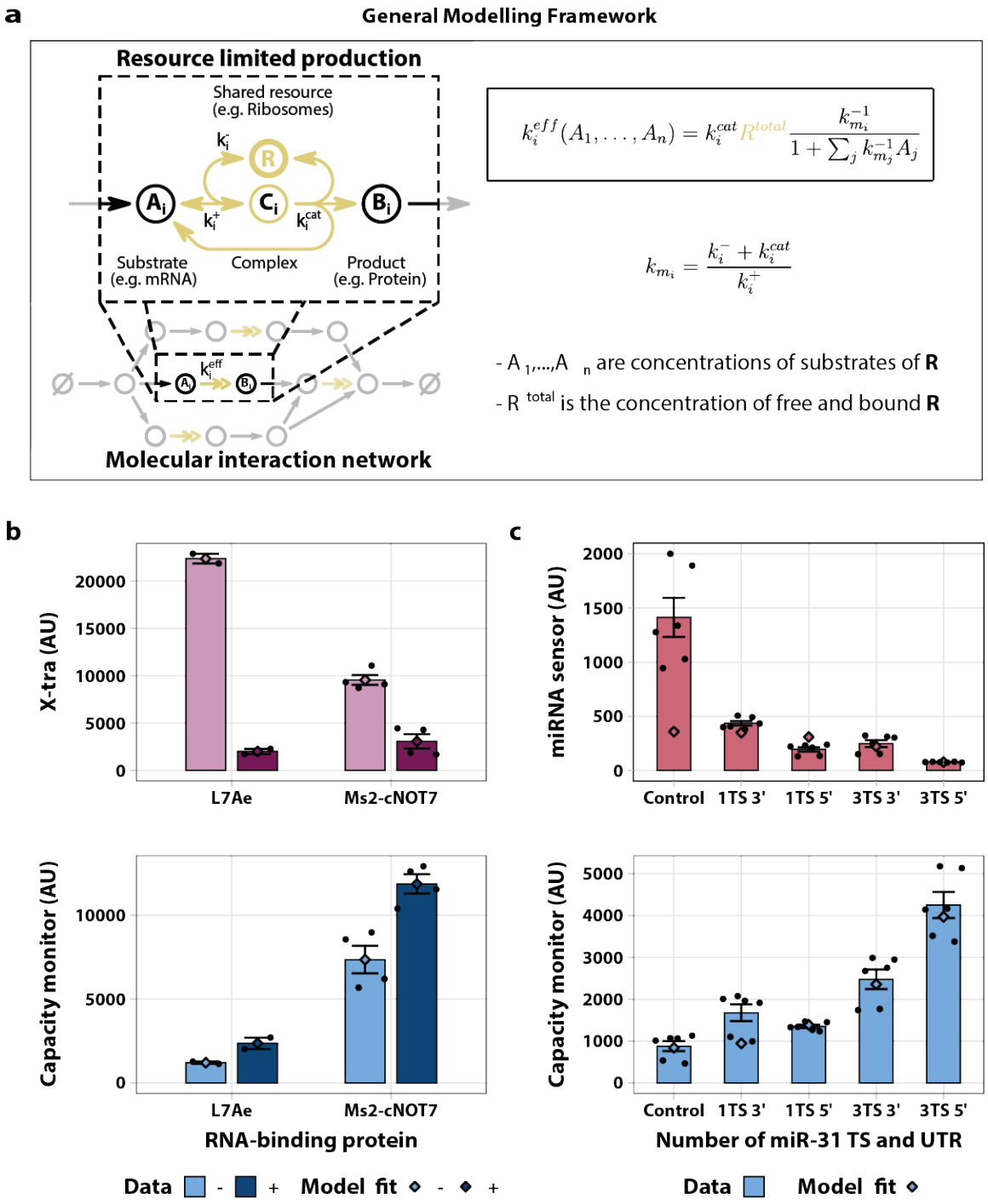


Figure 3.12: **A resource-aware mathematical modelling framework. (a)** General framework for transforming molecular interaction network models. Existing models of molecular interaction networks can be transformed to include shared limiting resources by substituting k_i, the reaction rate of a resource limited production, with $k^{eff}{}_i$ (see Appendix A.1 for the modelling derivation). Shown above an exemplary resource limited production are the detailed interactions between the substrate and the shared resource. **(b-c)** The modeling framework was applied to models of the genetic circuits used to generate the data in **Fig. 3.6c** and **Fig. 3.10b**. Detailed descriptions of these models and the parameters can be found in Appendix A.2. The data are presented as mean +/- SE. SE: standard error. au: arbitrary units.

to the expression of *X-tra*. The shift in resource allocation is expected to also affect the miR-31, which acts as a *capacity monitor*. This leads to a reduction in the repression of mKate, effectively compensating for the burden imposed by the co-expression of the *X-tra* gene. To demonstrate this mitigation approach experimentally, we co-transfected H1299 cells with increasing amounts of EGFP (*X-tra*), along with a constant amount of mKate (*GOI*) that either includes (for mitigation) or omits (no mitigation) three miR-31-TS in the 5'UTR. As expected, the expression level of X-tra approached saturation as the plasmid copy number increased, both for the targeted and non-targeted *GOI* variants (**Fig. 3.13c**). In agreement with previous results, the expression of the non-targeted *GOI* strongly decreased with increased expression of *X-tra*. Conversely, the decrease in expression of the targeted *GOI* was only about a third of that of the non-targeted variant, indicating improved adaptation to changes in resource availability (**Fig. 3.13c-d**). This observation was also captured well by a model of the system that explicitly considered resources, as described in the previous section. It should be noted that while the relative dynamic output range of *X-tra* is slightly reduced (fold change of 1.94x with mitigation, versus 2.18x without mitigation (**Fig. 3.13c**), our data show that the absolute level of *X-tra* increases about 2x in the presence of miR-31 based iFFL, *de facto* benefiting from this network topology (**Fig. 3.13e**). Analogously, miR-221-iFFL circuits specific for U2OS and HEK293T cells [153] show improved robustness to burden imposed by increasing exogenous gene load (**Fig. 3.14a-b**).

Importantly, the delivery of genetic payloads also affects the expression of endogenous genes (CyCA2, elF4E and GAPDH), as shown in **Fig. 3.4a,c**. We then sought to compare the expression of the same endogenous genes in the presence or absence of miR-31 sensor in H1299 cells. 48h post transfection of EGFP and mKate on a bidirectional plasmid, with mKate either including (miRNA sensor) or not (noTS) target sites for miR-31, we sorted cells according to high, intermediate or absence of fluorescence expression (**Fig. 3.15a**) and performed qPCR. Curiously, we observed that in cells transfected with miR-31-sensor, the decrease in the expression of the endogenous genes was much lower than in its absence (**Fig. 3.15c**). Furthermore, the expression of endogenous genes was inversely proportional to the levels of fluorescent proteins (**Fig. 3.15b**). Thus, the lower expression of endogenous genes due to the burden imposed by exogenous payloads is counteracted

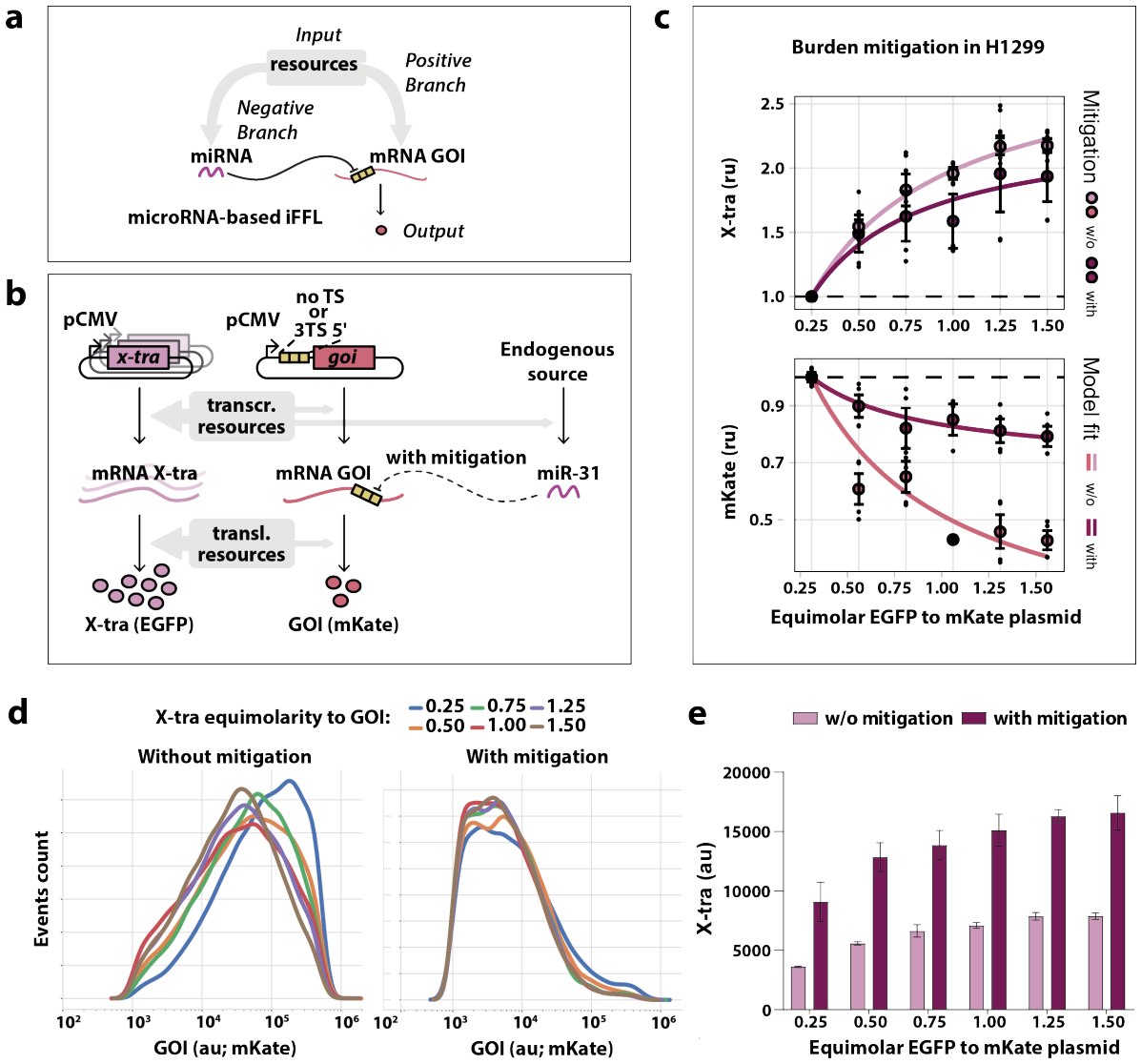


Figure 3.13: **Mitigating the effects of resource limitation with microRNA-based iFFL in H1299.** **(a)** The microRNA-based incoherent feedforward loop (iFFL) motif. **(b)** Mitigation system based on endogenous microRNA. At high copies number of the *X-tra*, resources are drawn away from the production of the *GOI* and miR-31. By sensing the resource availability and repressing the *GOI* less when there are fewer resources, the miRNA reduces the effect of limited resources. **(c)** Two plasmids were co-transfected into H1299 cells which respectively express the *X-tra* and *GOI* genes (EGFP and mKate respectively, panel b), and the molar ratio of the X-tra:GOI plasmid was progressively increased. The presence of miR-31 TS in mKate 5'UTR mitigates effects due to resource sharing. ru, relative units. The equations and parameter values for data fitting are summarized in Appendix A.3. N = 3 biological replicates. **(d)** Tolerance of mKate to increasing levels of *X-tra* gene in the absence or presence of an iFFL in which mKate includes miR-31 TS in the 5'UTR. The iFFL mitigation of resource competition is reflected by smaller shifts in GOI fluorescence at different equimolar ratios (right side). **(e)** *X-tra* expression levels in absolute fluorescence units (au). With miR-31 iFFL mitigation, the absolute X-tra expression increases about 2 fold compared to w/o mitigation. Data is plotted as mean +/- SE. SE: standard error. au: arbitrary units. N = 3 biological replicates.

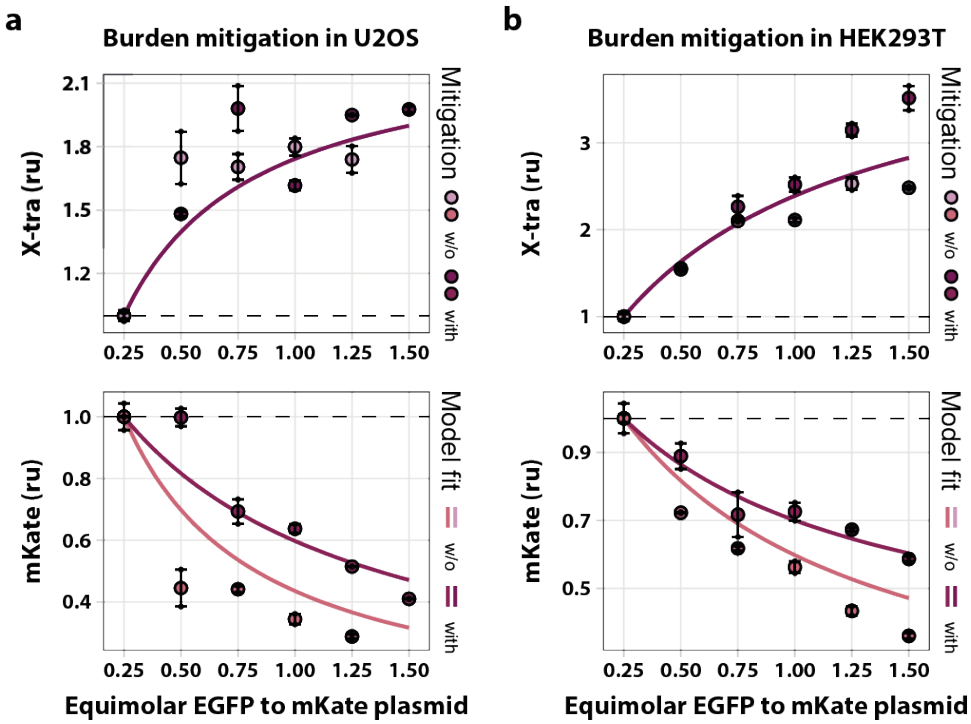


Figure 3.14: **Mitigating the effects of resource limitation with microRNA-based iFFL in U2OS and HEK293T cells. (a-b)** An iFFL whereby mKate includes miR-221 TS in the 5'UTR is less affected by the increased amount of the *X-tra* gene, as compared to the expression in the absence of miR-221 regulation. The model was unable to capture the differences in expression between the two conditions in the *X-tra* response due to the variability in the data, therefore, the two lines plotted are exactly the same (parameters listed in Appendix A.3). Data are normalized to the lowest equimolar ratio. Data were acquired 48h post-transfection and are plotted +/- SE. SE: standard error. ru: relative units. N=2 biological replicates (N = 1 for w/o Mitigation, 1.5 equimolar EGFP to mKate plasmid in panel a).

by the miR-31-sensor. To investigate whether the use of endogenous miRNAs may impair the regulation of native targets, we measured the expression of SATB2 mRNA, a natural target of miR-31 [154] in cells transfected with miR-31-sensor versus the noTS control, and observed no difference between the two conditions (**Fig. 3.15d**).

Indeed, mitigation comes at the cost of the maximal achievable expression levels for the capacity monitor. Moreover, tuning the iFFL circuit to become even less sensitive to changes in available resources will necessarily further limit the maximal expression. This trade-off is intrinsic to the iFFL mitigation strategy. Nevertheless, these results suggest that our approach can be used to mitigate resource-mediated coupling of gene expression despite cell-to-cell variability, demonstrating the portability and broad applicability of our findings. Our results demonstrate that iFFL circuits can mitigate burden from transgene

expression in mammalian cells.

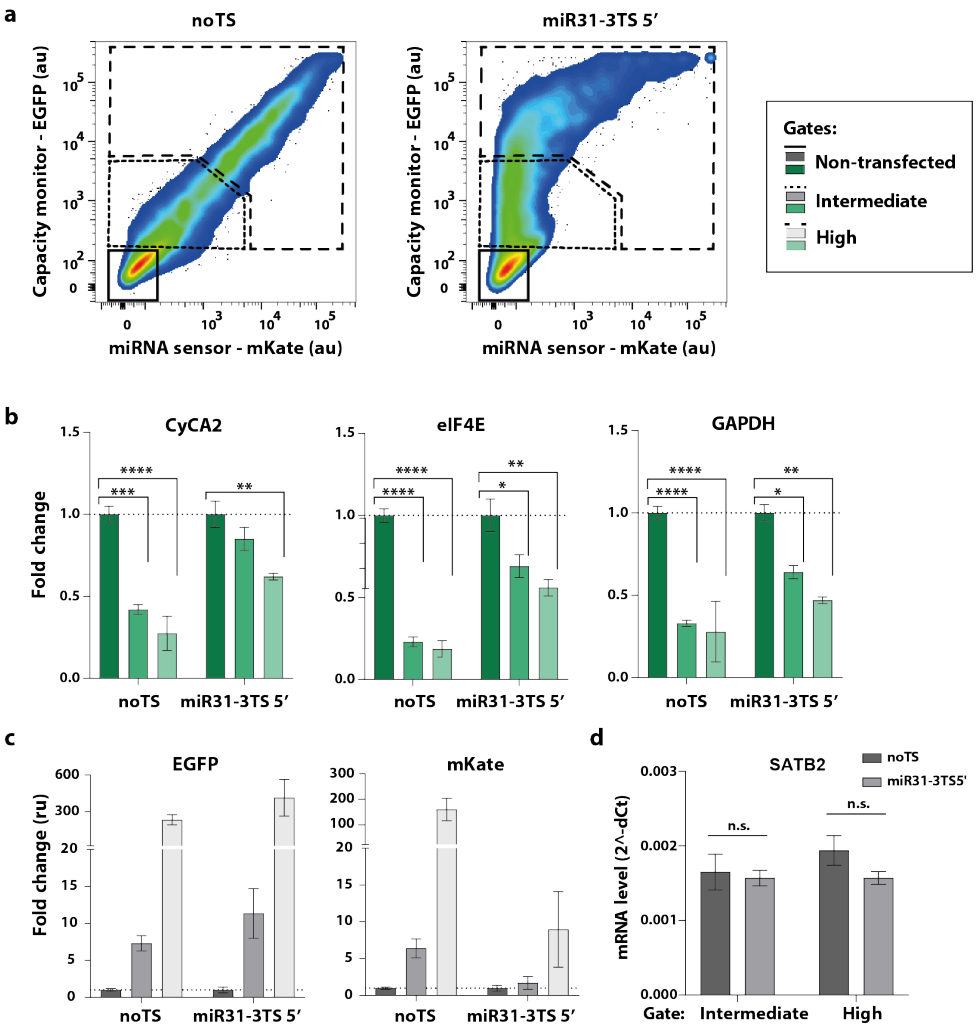


Figure 3.15: **Impact of transient plasmid transfection on endogenous genes in noTS and miR31-sensor samples.** **(a)** H1299 cells transfected with a bidirectional promoter plasmid encoding the fluorescent proteins EGFP (*capacity monitor*) and mKate (*miRNA sensor*), without (noTS, left) or with TS for miR-31 (miR31-sensor, right). Cells were sorted by fluorescence intensity 48 hours after transfection to collect non-transfected, intermediate and high transfected cells from the same transfection plate. **(b)** mRNA levels of CyCA2, eIF4E and GAPDH in sorted samples. All three endogenous genes decrease in transfected as compared to non-transfected cells. However, in cells transfected with the miR31-sensor circuit the decrease of expression is lower. mRNA levels are normalized to the non-transfected population. **(c)** Protein levels of EGFP and mKate in sorted populations. Data are plotted as the geometric mean of fluorescence +/- SE. **(d)** SATB2 mRNA level does not vary in presence of miR-31 TS. Data were collected 48 hours after transfection and represent mean +/- SE. SE: standard error. au: arbitrary units. Unpaired two-sided T-test. p-value: ****<0.0001, ***<0.0005, **<0.005, *<0.05. N = 3 biological replicates.

3.3 Molecular mechanisms involved in miRNA-dependent resource redistribution

Unpublished results

To deepen our understanding of how miRNAs mitigate unintended gene expression coupling, we sought to investigate the molecular mechanisms involved in miRNA-dependent resource re-allocation. The aim was to identify the key players involved to provide further elements to optimize gene circuits performance.

3.3.1 miRNAs mode of action

miRNAs are small non-coding RNAs, that act post-transcriptionally usually by base pairing to the 3'UTR of mature mRNAs and affecting gene expression [155]. However, the role of miRNAs is exploited also through target sequences at the 5'UTR [156] or by simultaneously binding target sites (TS) at both 3' and 5'UTR [157]. Their binding to target sequences leads to the degradation of the transcript and/or the repression of mRNA translation [158]. Once miRNAs are exported in the cytoplasm and become mature, they are loaded into a protein complex called miRISC. The mature miRNAs drive the miRISC to the mRNAs. A molecular cascade begins and leads to mRNA degradation and translational repression (**Fig. 3.16**), but the exact molecular mechanism and the dynamics of degradation and translational repression are poorly understood [159].

miRNAs are active in the form of ribonucleoprotein complexes miRISCs, in which AGO2 and GW182 are the best characterized protein components. AGO2 binds GW182, which functions as scaffold protein that triggers the recruitment of deadenylation and decapping factors [159]. Once mRNA is decapped and deadenylated, it is unstable and the exoribonuclease 1 (XRN1) is engaged for the final degradation.

In eukaryotic cells, translation initiation requires the activity of 13 initiation factors (eIFs) and represents the limiting step in translation. miRNAs inhibit initiation cutting off the cap at the 5' end of mRNAs. Moreover, we hypothesize that when miRNAs bind to the 5'UTR of the transcript, translation is blocked by steric hindrance (**Fig. 3.16**).

The mechanism of action depends also on the target site complementarity. In mammalian cells, miRNA binding sites are not fully complementary and they rather bind

to a seed region that is 7/8 bp long [160]. When this imperfect pair match is created, deadenylation and decapping occur and XRN1 is recruited. On the contrary, mRNAs in plants contain fully complementary binding sites [161], such as the TS in our system. The perfect base pairing causes a neat cleavage of the mRNA by action of the Argonaute protein itself, inducing RNA degradation and recruitment of both XRN1 and the Exosome for 3'-5' and 5'-3' mRNA degradation respectively. In animals, this alternative mechanism of action is described only in HoxB8 gene [147]. In our experimental setup, we use fully complementary TS to maximize miRNA activity on the *miRNA sensor*, however future studies will need to address if resource re-allocation changes with partially complementary TS like in physiological systems.

3.3.2 Hypotheses of resource re-allocation mechanisms

Based on this knowledge, to explain the resource re-allocation we formulated two hypotheses: *(i)* as miRNAs cause degradation of the target mRNA and block its translation, there are more ribosomes available to translate the *capacity monitor* **Fig. 3.10a-b**); *(ii)* as we induce a massive mRNA degradation of the *miRNA sensor*, a queuing effect on the mRNA degradation machinery builds up, slowing down the degradation process of the other transcripts. In support of our hypotheses, it has already been proven in bacteria that mRNA decay can be programmed to modulate intracellular resource allocation channeling translational resources from endogenous pathways to synthetic circuits [162]. Recently, it was proven that limited ribosome pools can lead to impaired translational profiles also in mammalian cells [163]. Moreover, queuing for biochemical processing has already been shown for proteins tagged for degradation by the proteasome [164], and this could apply also the mRNA species, as hypothesized by Gambardella *et al* [165].

Finally, we hypothesized that when miR-TS are placed in the 5'UTR, ribosomes are unable to assemble onto the transcript due to steric hindrance of the miRISC complex [166, 167]. On the contrary, when miR-31 TS are in the 3'UTR of the transcript, gene expression reduction may be mainly due to mRNA degradation. This would explain the observations in Section 3.2.4 in **Fig. 3.10a** and **Fig. 3.11a-c**, where *miRNA sensors* that harbor TS in the 5'UTR exhibit higher repression of the fluorescent reporter (lower mKate level).

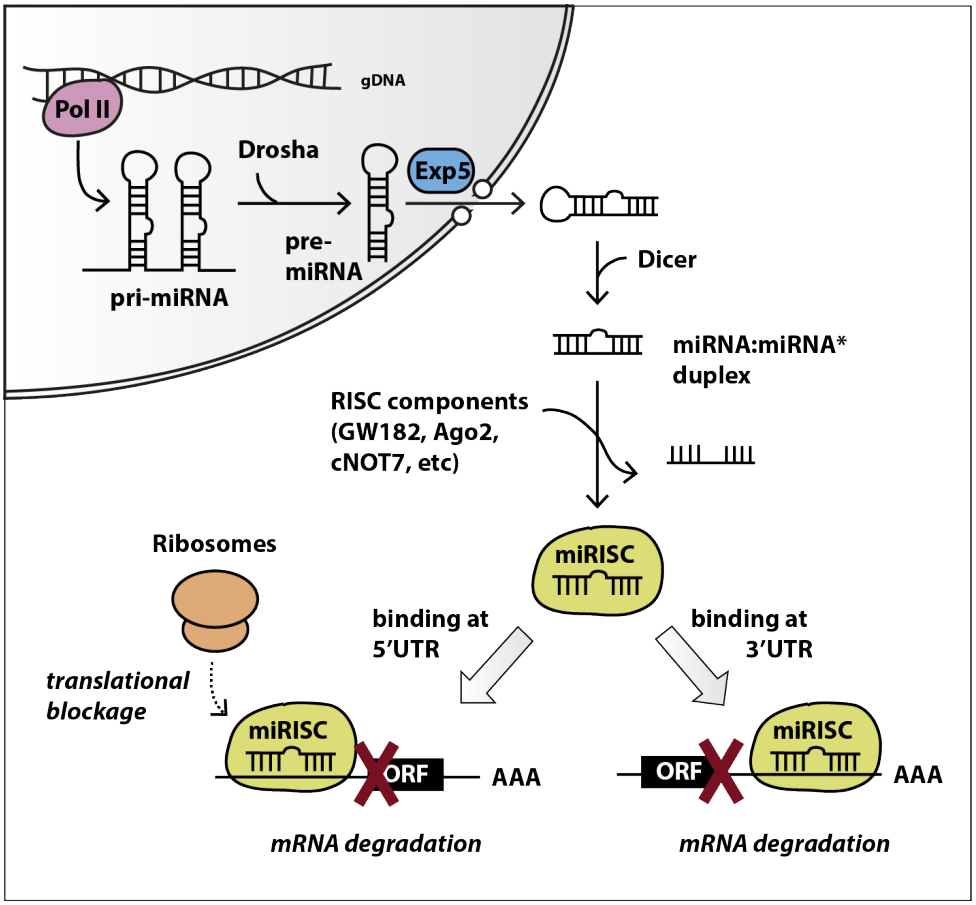


Figure 3.16: **Schematic representation of miRNA biogenesis and mechanism of action.** miRNA genes are transcribed inside the nucleus (pri-miRNA) by RNA Polymerase II (Pol II) and are usually located in the introns of the genes. Pri-miRNAs are then processed by Drosha, a ribonuclease III enzyme, and exported in the cytoplasm thanks to the Exportin 5 (Exp5). Once in the cytoplasm, miRNA precursors are further processed by Dicer, another ribonuclease III enzyme, and finally one strand of the miRNA duplex (miRNA:miRNA*) is loaded into the miRNA-induced silencing complex (miRISC). miRISC associates with the mRNA target either binding at the 3'- or 5'-UTR resulting in mRNA degradation and translational repression plus mRNA degradation respectively.

3.3.3 miRNA-dependent resource re-allocation leads to altered translational profiles

To investigate the first hypothesis (increased ribosome availability), in collaboration with Dr. Gabriella Viero (Università di Trento), we ran a polysome sequencing on cells transfected with the bidirectional promoter plasmids (**Fig. 3.10a**) with 0 or 3 miR-31 TS either at the 3' or 5'UTR. Polysomes are mRNAs coupled with more than 2 ribosomes, meaning that the mRNA is actively translated. Polysome sequencing is the technique used to study the translational profiles of transcripts of interest in different experimental

conditions [168]. First, cells are treated with cycloheximide to "freeze" ribosomes on the transcripts. Then, cellular lysates are centrifuged on a sucrose gradient and fractions are collected while the λ_{254} absorbance profile is acquired (**Fig. 3.17**). The first peak contains free cytosolic light components (RNPs), and the subsequent peaks include ribosomal subunits (40S and 60S) and monosomes (80S), all associated with non-translating particles [169]. The remaining peaks of the profile represent polysomes, which sediment with high sucrose concentrations and contain the RNAs associated with ribosomes. Therefore, heavy fractions (no. 7-12) are the ones where we find actively translated transcripts. The heaviest fractions (no. 11-12) could also contain transcripts upon which ribosomes have started stalling (no active translation).

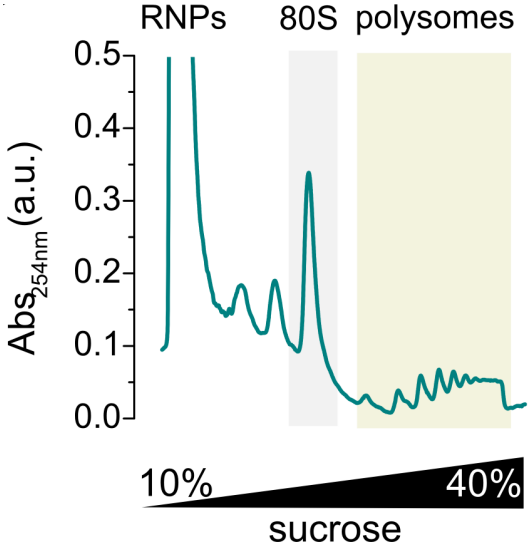


Figure 3.17: **Sucrose gradient absorbance profiles of a non-transfected H1299 cells lysate.**

Our experimental results revealed modified translational profiles for the *capacity monitor* (EGFP, exogenous) and GAPDH, eIF4E and CyCA2 (endogenous), suggesting enhanced translation of these genes in the samples transfected with plasmids with miR-31 TS as compared to the control (**Fig. 3.18**). In all plots the grey line corresponds to the 0 miR-31 TS control, and the colored line corresponds to either 3 miR-31 TS at 3'UTR (plots on the right) or 3 miR-31 TS at 5'UTR (plots on the left). The shift towards the polysomal fractions is more pronounced in the 3TS 5'UTR sample (iFFL configuration, refer to **Fig. 3.13**), consistent with our hypothesis that the miRISC complex prevents ribosomes from binding to the *miRNA sensor/GOI* due to steric hindrance, leading to a re-allocation of ribosomes among other transcripts.

In the 3TS 5'UTR case, the effect is significant for all the 4 genes analyzed (**Fig. 3.18a**): EGFP mRNA amount decreases significantly in the non polysomal fractions (no. 6 corresponding to 1 ribosome only) and increases in the 9th and 10th fractions; GAPDH, eIF4E, CyCA2 are significantly enriched in the 9th fraction, meaning that are being more actively translated. GAPDH in the Control sample seems enriched in the 12th fraction; this might also correspond to a stalling of ribosomes on the transcript rather than to an active translation.

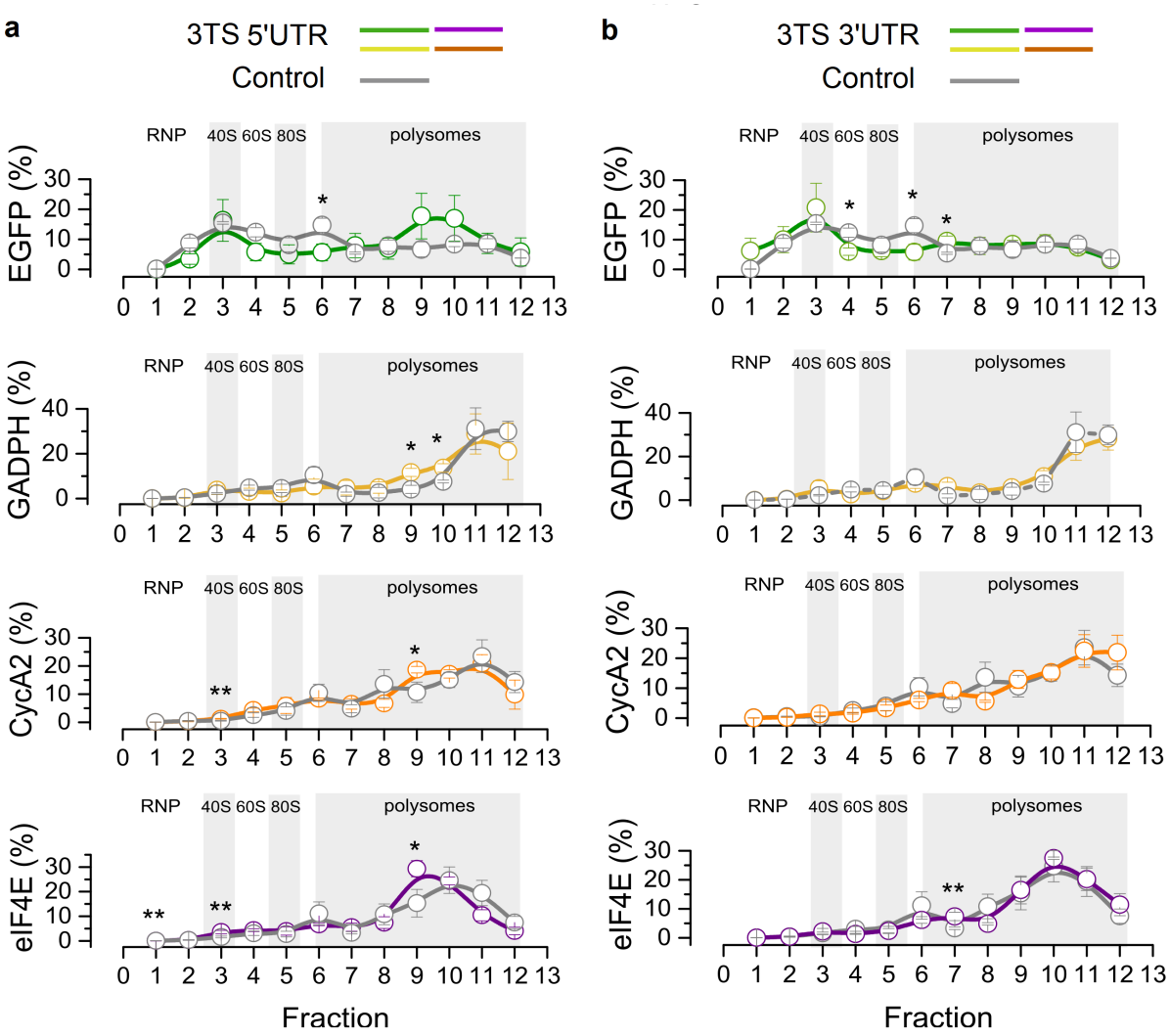


Figure 3.18: **Polysome sequencing reveals modified translational profiles in H1299 transfected with miR-31 sensors.** Polysome sequencing was performed on H1299 cells transfected with a bidirectional promoter plasmid encoding for EGFP (*capacity monitor*) and mKate (*miRNA sensor*) with miR-31 TS either in the 5' **(a)** or 3' UTR **(b)**. As control, we used the same plasmid lacking the miR-31 target sites. All data were acquired 48 hours post transfection and are plotted ± SE. SE: standard error. N = 3 biological replicates. Unpaired two-sided T-test. p-value: * < 0.05, ** < 0.005.

In the 3TS 3'UTR case, the shift towards the polysomal fraction is significant only for GAPDH and eIF4E, but a slight enrichment in the 7th fraction can be appreciated for all genes (**Fig. 3.18a**).

3.3.4 miRNA activity increases the half-life of non-target transcripts

To investigate our second hypothesis of queuing effect on the mRNA degradation machinery, we measured mRNA half-life upon transcription inhibitor 5,6-dichloro-1-beta-D-ribobenzimidazole (DRB) treatment [170] of endogenous genes (**Fig. 3.19**). We focused on CyCA2 and eIF4E, because DRB treatment was not effective on transfected genes and housekeeping genes like GAPDH.

We co-transfected H1299 cells with a plasmid encoding mKate with 0 or 3 miR-31 TS either at the 3' or 5'UTR (*miRNA sensor*) and with a constitutively expressed EGFP (*capacity monitor*). Over a time-frame of 4 hours we observed that CyCA2 and eIF4E mRNAs take longer to be degraded if cells are transfected with a *miRNA sensor* flanked by miR-31 TS as compared to the control (0 miR-31 TS) (**Fig. 3.19a-b**). More specifically, the two mRNA species decrease immediately after treatment only in the control, while in the 3TS at 3' UTR and 5' UTR, they remain almost stable for 1:30 hours after treatment and undergo a significant decrease at 2-3 hours. The qualitative trend of 3TS 3' and 3TS 5' samples is very similar, however when TS are at 3'UTR the delay in mRNA degradation seems further enhanced. Our results are in agreement with our hypothesis of a prominent queuing effect on the mRNA degradation machinery as a consequence of the massive degradation of the target (*miRNA sensor*) by miR-31.

Overall, we conclude that both ribosome availability and queuing effect on mRNA degradation play a role in unintended gene coupling and enhanced (*capacity monitor*) expression following (*miRNA sensor*) downregulation by miR-31 activity.

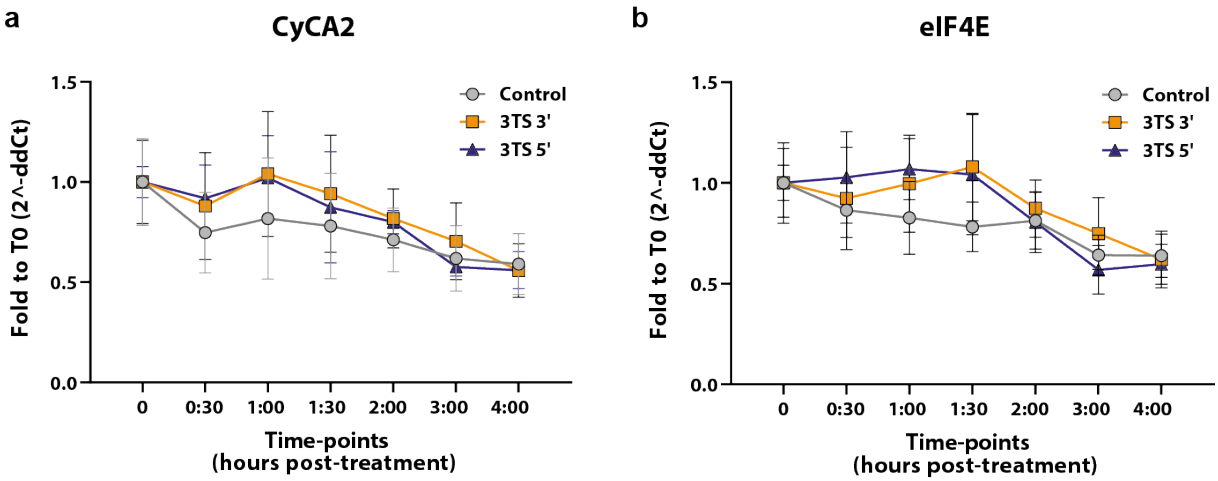


Figure 3.19: **CyCA2 and eIF4E mRNA half-life measurement upon DRB treatment.** mRNA half-life and degradation dynamics measurement of CyCA2 **(a)** and eIF4E **(b)** upon DRB treatment (50 µM). Cells were co-transfected with *miRNA sensor* and *capacity monitor* plasmids. The presence of miRNA TS on *miRNA sensor* plasmid likely cause a delay in the cellular mRNA degradation machinery, therefore endogenous genes could take longer to be degraded. Results support our hypothesis. All data were acquired 24 hours post transfection and are plotted +/- SE. SE: standard error. N = 6 biological replicates; N=2 biological replicates for Time-points = 0:30 and 1:30.

3.3.5 Modelling the molecular mechanisms involved in resource re-allocation

Based on the experimental findings described above, we decided to further dig into our hypotheses implementing a reaction-based model that describes miRNAs mechanism of action and its impact on resource allocation. This part is in collaboration with Professor Guy-Bart Stan and his group at Imperial College London. Our model is inspired by the modelling framework described in Section 3.2.5 that accounts for gene expression in resource-limited environments [67], but it is aimed at precisely capturing the individual contributions of the different resource pools.

We considered three pools of limited free resources: transcriptional, translational and mRNA-degradation resources. In our model these are represented by Pol II (Pol), ribosomes (R) and mRNA-degradation complex (X). For simplicity, we included both the XRN1 and the exosome - that play a role in mRNA degradation (Section 3.3.1) - under the X pool. The size of the free resource pools is a function of the resource demand and, consistently, they correspond to the difference between the total amount of molecules and the molecules sequestered by the other reactions.

miRNA TS location at 3' or 5'UTR was modelled through two different miRNA species - namely mi_3 and mi_5 - that are produced and degraded with the same reaction rates (Appendix B.2.2), but that exert two different effects on the target and, indirectly, on the free resource pools (Appendix B.2.3). Both the first and the second trigger mRNA degradation, but only the second causes translational blockage. The complete list of species and reactions in the model can be found in Appendix B.1.1 and B.1.2 respectively.

As described in Appendix B.2, all reactions follow mass-action kinetics. mRNA transcription and protein translation were both modelled in three steps (Appendix B.2.1, reactions B.3-B.6): *(i)* formation of an initiation complex ($C_{d,j}$ and $C_{m,j}$) given by the reversible binding of a Pol II or a Ribosome (Pol or R) to their respective binding sites (d_j or m_j); *(ii)* beginning of translation with the transition from initiation complex to elongation complex ($N_{d,j}$ and $N_{m,j}$) and concomitant release of the binding sites d_j or m_j; *(iii)* termination of the elongation phase by releasing the product species (mRNA, represented by m_j and protein, represented by p_j) and the bound resources (Pol and $\gamma_j R$). We assumed that the translating ribosomes evenly distribute upon the transcripts, therefore the number of ribosomes per transcript is:

$$\gamma_j = \frac{|N_j|}{|m_j + C_{m,j}|} \tag{3.1}$$

where N_j is the total number of translating ribosomes and the sum of $m_j + C_{m,j}$ is the number of transcripts undergoing translation. This tripartite model allows to capture the re-allocation of transcriptional and translational resources with single molecule resolution. Finally, mRNA degradation is catalyzed by the mRNA-degradation complex X at k_{Xm} plus k_d rates (reactions B.8-B.9), while proteins undergo unspecific degradation at $\lambda_{p,j}$ rate (reaction B.10).

mi_3 and mi_5 are produced according to the transcription model described above and undergo unspecific degradation at λ_{mi} rate (reaction B.11-B.13). When present in the system, they both have the ability to bind the target mRNA at k_b rate to induce its degradation. The difference between the two miRNA species is that mi_5 can bind only if the ribosome binding site (RBS, m_j) is free due to steric hindrance with bound ribosomes; while mi_3 can bind both when m_j is free or bound by a ribosome (initiation complex C_i, reactions B.14, B.20, B.21). In all the aforementioned cases, the binding of a

miRNA recruits the mRNA-degradation complex sequestering resources from the free X pool with k_{X3} and k_{X5} rates respectively (reactions B.16, B.25, B.26). They have different values, specifically k_{X5} is smaller as the main effect of mi_5 when it binds to its target is translational blockage; while we expect mi_3 to trigger higher target mRNA degradation, in agreement with the higher mRNA degradation delay observed on endogenous genes when the *miRNA sensor* has 3 miR-31 TS at 3'UTR (Fig. 3.19). miRNAs free up $\zeta_{q,j}R$ ribosomes for each degraded transcript thanks to target assisted degradation by X at k_d rate, as described by reactions B.18-B.19, B.27-B.28. Similarly to what described for γ_j, $\zeta_{q,j}$ is the number of ribosomes on each transcript bound by a miRNA, so assuming that the total bound ribosomes evenly distribute among the transcripts:

$$\zeta_{q,j} = \frac{|K_{q,j}|}{|L_{q,j}|} \tag{3.2}$$

where $K_{q,j}$ is the total number of bound ribosomes and $L_{q,j}$ is the number of transcripts bound by miRNAs. mi_5 further enhances ribosomes re-allocation by avoiding further ribosomes from binding to the RBS m_j of the mRNA target ($L_{5,j}$), which cannot initiate translation anymore. A complete list of the model reactions is listed in Appendix B.2.4.

To summarize, we defined the model reactions starting from phenomenological observations of our system and then integrating mechanistic molecular aspects, even if simplified. To differentiate the effects of miRNAs action when the TS are at the 3'UTR or at the 5'UTR, we simplified the system, assuming that at 3'UTR only degradation happens, while at 5'UTR translational blockage as effect of steric hindrance also plays a role. Next, we will run model simulations and we will determine rate values starting from what is already reported in literature.

3.4 Changes in the intracellular environment due to gene expression burden

Unpublished results

To have a broader picture of the global effects that genetic payloads have on host cells, we performed RNA sequencing (RNA-seq) and proteomic experiments (in collaboration with Dr. Andrea Armirotti, IIT) on transfected and non-transfected H1299 cells in different transfection conditions. Specifically, we ran three transient co-transfections in H1299 with equimolar ratio of: *(i)* a constitutively expressed *capacity monitor* (EGFP) and a *miRNA sensor* (mKate) with 3TS for miR-31 at 3'UTR or (ii) at 5'UTR; *(iii)* negative control of *miRNA sensor* lacking the miRNA-TS and *capacity monitor* (see plasmid design in **Fig. 3.20**). Transfected cells were sorted by fluorescence intensity to collect non-transfected cells (control), intermediate, and highly transfected cells (sorting strategy illustrated in **Fig. 3.4b** and **Fig. 3.20a**). We sorted two different populations for the transfected cells to gather information on the impact that different levels of plasmid uptake (and therefore of transgene expression) have on the intracellular machinery. Data analysis on sorted populations was tailored to study both the impact of different amounts of genetic payloads delivered into cells on the intracellular environment - distinguishing between Non-Transfected (NT), Intermediate and High transfected cells - and the extent of the impact depending on genetic circuit design - comparing no *miRNA sensor* repression (noTS) and miRNA-dependent repression (miR-31-3TS 3' and miR-31-3TS 5').

3.4.1 Delivery of genetic payloads into host cells affects endogenous genes expression

The differential gene expression analysis on the RNA-seq results was carried out comparing the High, Intermediate and NT cell populations within the same sample to understand the contribution of different levels of plasmid uptake on the transcriptional landscape. NT population was used as control in the comparison with Intermediate and High population; Intermediate population was used as control in the comparison with High population.

To select the differentially expressed genes (DEGs), we filtered the results first ac-

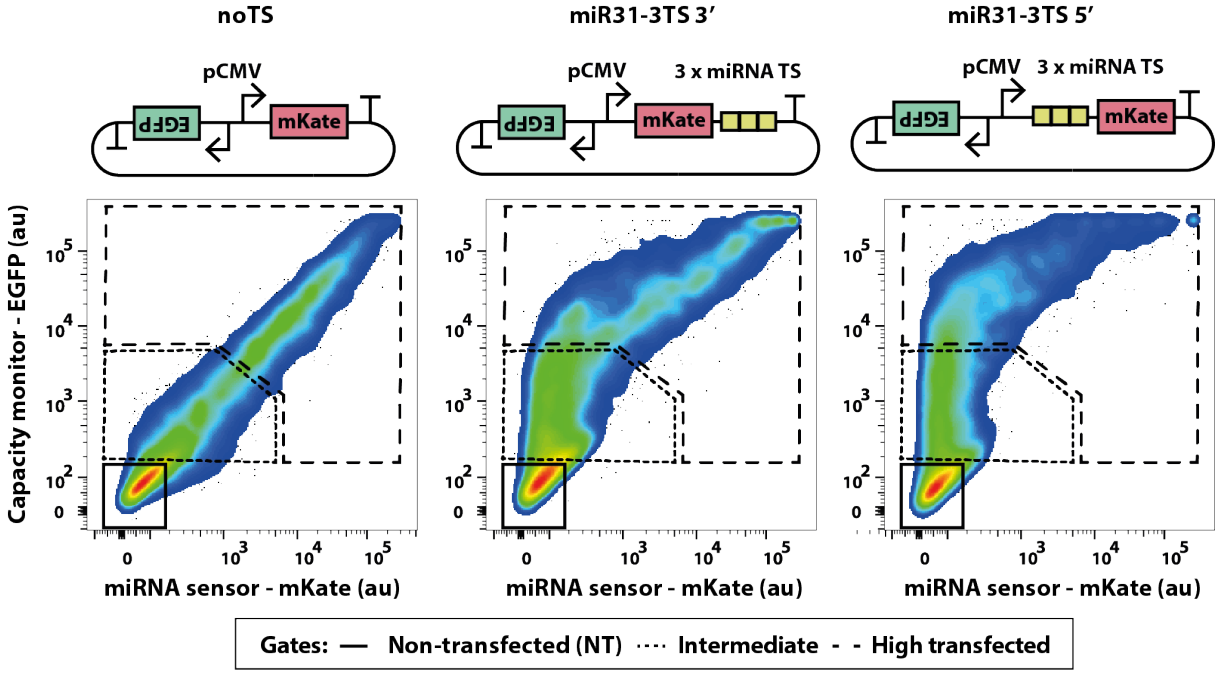


Figure 3.20: **Gating strategy to sort populations with different gene expression burden.** H1299 cells were transfected with a bidirectional promoter plasmid encoding the fluorescent proteins EGFP (*capacity monitor*) and mKate (*miRNA sensor*), without (noTS, left) or with 3TS for miR-31 either at the 3'UTR (middle) or 5'UTR (right). Cells were sorted by fluorescence intensity 48 hours after transfection to collect non-transfected (NT), intermediate transfected and high transfected cells from the same transfection plate.

cording to the False Discovery Rate (FDR) < 0.05 in order to shortlist only the statistically significant changes (**Table 3.1**), and then by the absolute logarithm of the fold change ($|\text{Log}_2(\text{FC})|$) ≥ 0.5, to exclude changes that would hardly have a biological significance considering their magnitude.

Table 3.1: **Number of DEGs in each transfection condition according to the RNA-seq analysis.** NT = Non-transfected; Intermediate = Intermediate transfected; High = high transfected.

Sample	Comparison	FDR < 0.05	LogFC ≥ 0.5	LogFC ≤ -0.5
noTS	High vs NT	1106	271	235
	Intermediate vs NT	800	206	135
	High vs Intermediate	4	3	1
3TS 3'	High vs NT	420	150	83
	Intermediate vs NT	37	14	11
	High vs Intermediate	73	17	23
3TS 5'	High vs NT	314	123	75
	Intermediate vs NT	0	0	0
	High vs Intermediate	132	24	52

We observed that the noTS sample exhibits the highest number of DEGs. This result may indicate a stress the cells are undergoing, in agreement with the strongest gene burden we had previously observed (Section 3.2 and 3.3). Regarding the comparison between TS locations, if they are located at the 5'UTR the impact on gene expression is further reduced (lower amount of DEGs). Further, the analysis shows that a high amount of plasmid uptake results in the highest number of DEGs as compared to the NT control population in all the samples. Cells transfected with the miR-31 sensor showed a lower degree of DEGs, possibly indicating that less stress was induced by transgene expression. This is especially true for low plasmid quantity (Intermediate *vs* NT) and, to a lower extent, also for higher amount of plasmid uptake (High *vs* Intermediate).

Similarly to the RNA-seq data, we performed a differentially expressed proteins analysis comparing the proteomics results among the different populations within the different transfections or samples (**Table 3.2**). The number of peaks (no. peaks) corresponds to the number of proteins detected in each sample and it varies between 2065 and 2668. We considered FDR ≤ 0.1 as statistical cutoff to select the differentially expressed proteins (DEPs). We set a less stringent threshold for the FDR as compared to the RNA-seq data, because a different statistical test was applied (Section 6.6.2) and the number of significant comparisons was otherwise very low. We considered biologically significant the changes with $|\text{Log}_2(\text{FC})| \geq 0.5$.

Table 3.2: **Number of DEPs in each transfection condition according to the proteomics analysis.** NT = Non-transfected; Int = Intermediate transfected; High = high transfected.

Sample	No. Peaks	Comparison	FDR ≤ 0.1	Log(FC) ≥ 0.5	Log(FC) $\leq$ -0.5
noTS	2668	High vs NT	185	37	64
		Int vs NT	134	10	77
		High vs Int	94	33	16
3TS 3'	2055	High vs NT	121	46	23
		Int vs NT	33	10	14
		High vs Int	102	31	30
3TS 5'	2515	High vs NT	90	59	14
		Int vs NT	54	29	14
		High vs Int	61	29	18

Giving a closer look at the numbers reported in the table, we can see that the number of DEPs (FDR ≤ 0.1) is higher in noTS samples as compared to 3TS 3' and 3TS

5' samples. Therefore, also for protein expression, it seems that having miR-31 TS at the either at the 3'UTR or at the 5'UTR of the exogenous gene is beneficial.

3.4.2 The transcriptional landscape of burdened cells

To gain more insights about gene dysregulation, we investigated the DEGs between sorted populations using Venn diagram to detect the common DEGs and scatter plots to represent their fold change correlation **Fig. 3.21a, 3.22a, 3.23a**. The principal component analysis (PCA) revealed how the 3 sorted populations distribute in the 2 dimensional Cartesian space for the three different transfection conditions (**Fig. 3.21b, 3.22b, 3.23b**). In the noTS samples the distance between non-transfected (NT) and transfected populations (both Intermediate and High) is clear (**Fig. 3.21b**), while when 3TS for miR-31 are added at the 3'UTR the distance between the populations is shrunk (**Fig. 3.22b**). Finally, when the TS are placed in the 5'UTR, the High transfected population is the most distinguished from the others (**Fig. 3.23b**). It should be noted that the effect of gene expression burden on gene expression is not huge. In fact, the principal component 1 (PC1) is a variable that relates to the sample-to-sample variability: note how the 4 biological replicates (4 symbols with the same shape) equally distribute along the x axis. Nevertheless, distinct population patterns can be observed along the y axis (principal component 2, PC2).

We represented the distribution of the differential gene expression results with volcano plots for all the comparisons within the three transfection conditions (**Fig. 3.21c-e, 3.22c-e, 3.23c-e**). Light blue dots (top left quadrants) represent the significantly downregulated genes (Down); red dots (top right quadrants) represent the significantly upregulated genes (Up); grey dots (below the horizontal blue dashed line) represent the genes that do not undergo significant change (noDEG, FDR < 0.05); black dots (between the two vertical dashed lines and above the blue dashed line) represent the genes that undergo significant change but too little to be considered biologically relevant ($|\text{Log}_2(\text{FC})| \leq 0.5$). Consistently with the previous representations of the results, the most abundant and significant DEGs are observed in the High *vs* NT population comparison (middle plots). While the impact of gene expression burden is progressively mitigated by the inclusion of 3TS for miR-31 at 3'UTR or at 5'UTR, a closer look into the differentially

expressed genes reveals a common pattern. DDIT3 and ATF3 are the DNA Damage Inducible Transcript 3 and the Activating Transcription Factor 3 respectively. The first is activated by endoplasmic reticulum stress, and promotes apoptosis [171]; the second is involved in the complex process of cellular stress response [172]. Both of them are always upregulated - significantly 4 out of 8 times for ATF3 and 6 out of 8 times for DDIT3 - with the exception of ATF3 in noTS Intermediate *vs* NT **Fig. 3.21c** and DDIT3 in 3TS 5'UTR Intermediate *vs* NT comparison **Fig. 3.23c**. In both cases the Log_2(FC) of the genes is slightly negative but very close to 0. KLF10 and OSGIN1 are the Kruppel Like Factor 10 and the Oxidative Stress Induced Growth Inhibitor 1 respectively. Both of them are negative regulators of cellular growth: the first is a transcriptional repressor involved in the inhibition of cancer growth [173]; the second encodes an oxidative stress response protein that regulates cellular apoptosis by inducing cytochrome c release from mitochondria [174]. Both these genes are upregulated in transfected cells, significantly most of the time. Finally, PLD1 and PRKAR1B are the Phospholipase D1 and the Protein Kinase CAMP-Dependent Type I Regulatory Subunit Beta respectively. The first plays a role in signal transduction and subcellular trafficking [175]; the second is involved in many cellular events, including ion transport and metabolism [176]. Both of them are significantly downregulated in transfected samples.

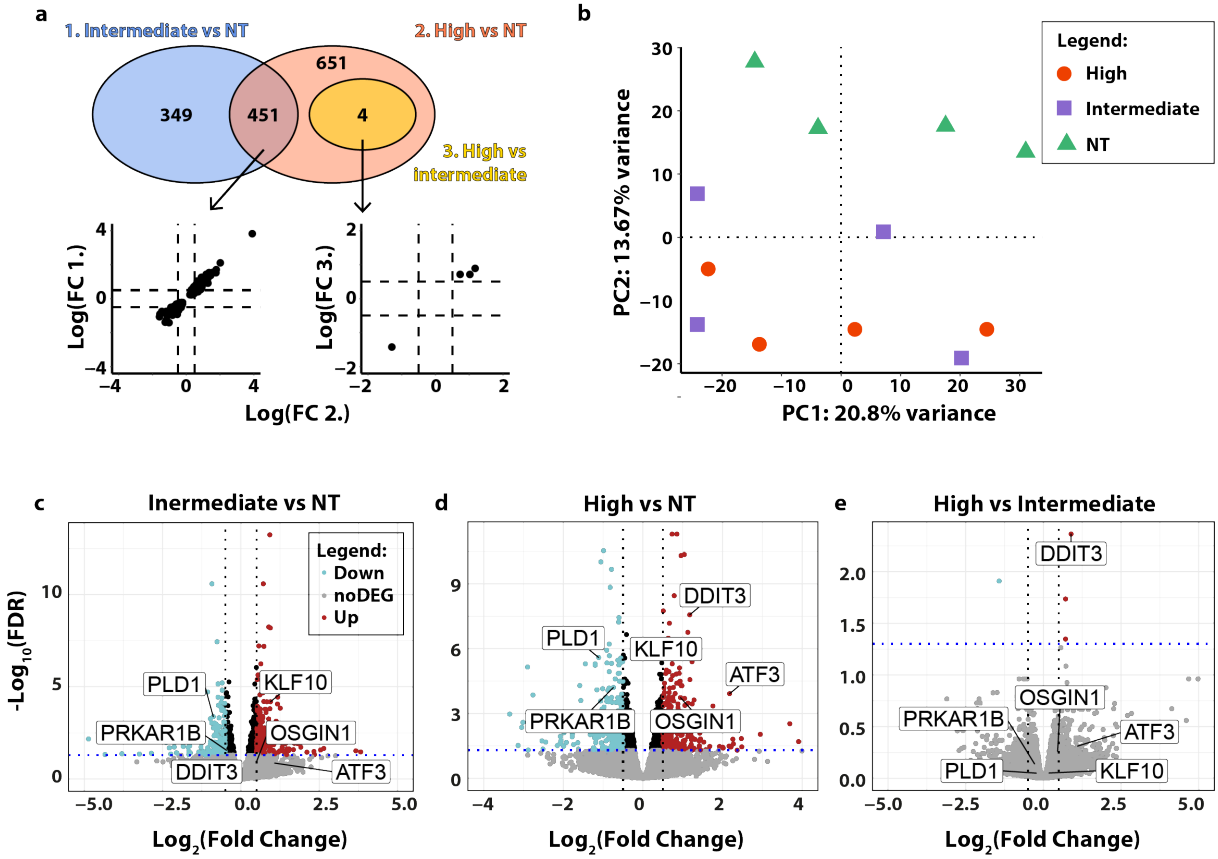


Figure 3.21: **DEGs in noTS sample.** **(a)** Intersection of the DEG among the three comparisons. Top, the Venn plot highlights a big overlap between Intermediate *vs* NT and High *vs* NT confirming that with this plasmid design the cellular machinery is impacted even with little DNA amount. Bottom, scatter dot plots show the positive relation between the Log_2(FC) of the common genes in the different comparisons. **(b)** PCA plot of the 4 biological replicates. The NT population (green triangles) is distinguished from the transfected populations (High and Intermediate). Volcano plots of genes distribution in Intermediate *vs* NT **(c)**, High *vs* NT **(d)** and High *vs* Intermediate **(e)** comparisons. Light blue dots (top left quadrants) represent the significantly downregulated genes (Down); red dots (top right quadrants) represent the significantly upregulated genes (Up); grey dots (below the horizontal blue dashed line) represent the genes that do not undergo significant change (noDEG, FDR < 0.05); black dots (between the two vertical dashed lines and above the blue dashed line) represent the genes that undergo significant change but too little to be considered biologically relevant ($|Log_2(FC)| \leq 0.5$). Highlighted are transcription factors related to stress response (DDIT3 and ATF3), negative regulators of cellular growth in response to stress (KLF10 and OSGIN1), genes involved in signal transduction and cellular metabolism (PLD1 and PRKAR1B). The latter are downregulated. Data were collected 48 hours post-transfection. N = 4 biological replicates.

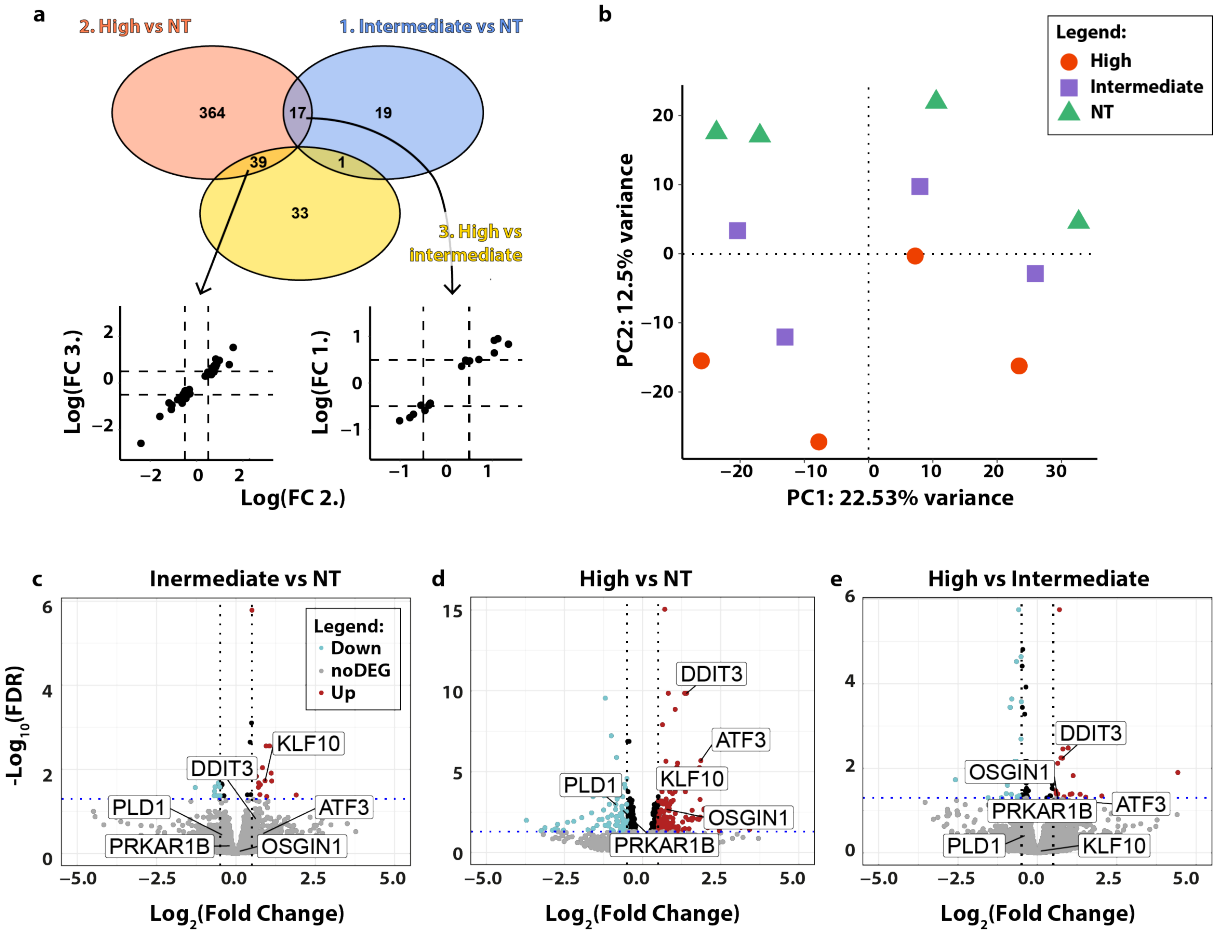


Figure 3.22: **DEGs in 3TS 3'UTR sample.** **(a)** Intersection of the DEG among the three comparisons. Top, the Venn plot highlights comparable overlaps between Intermediate *vs* NT / High *vs* NT and Intermediate *vs* NT / High *vs* Intermediate confirming that with this plasmid design the cellular machinery is less impacted by little DNA amount as compared to noTS condition (**Fig. 3.21a**). Bottom, scatter dot plots show the positive relation between the Log_2(FC) of the common genes in the different comparisons. **(b)** PCA plot of the 4 biological replicates. The three populations are distinguishable, but the distance of NT from transfected cells is narrower as compared to noTS condition (**Fig. 3.21b**. Volcano plots of genes distribution in Intermediate vs NT **(c)**, High vs NT **(d)** and High vs Intermediate **(e)** comparisons. Light blue dots (top left quadrants) represent the significantly downregulated genes (Down); red dots (top right quadrants) represent the significantly upregulated genes (Up); grey dots (below the horizontal blue dashed line) represent the genes that do not undergo significant change (noDEG, FDR $<$ 0.05); black dots (between the two vertical dashed lines and above the blue dashed line) represent the genes that undergo significant change but too little to be considered biologically relevant ($|Log_2(FC)| \leq 0.5$). Highlighted are transcription factors related to stress response (DDIT3 and ATF3), negative regulators of cellular growth in response to stress (KLF10 and OSGIN1), genes involved in signal transduction and cellular metabolism (PLD1 and PRKAR1B). The latter are downregulated. Data were collected 48 hours post-transfection. N = 4 biological replicates.

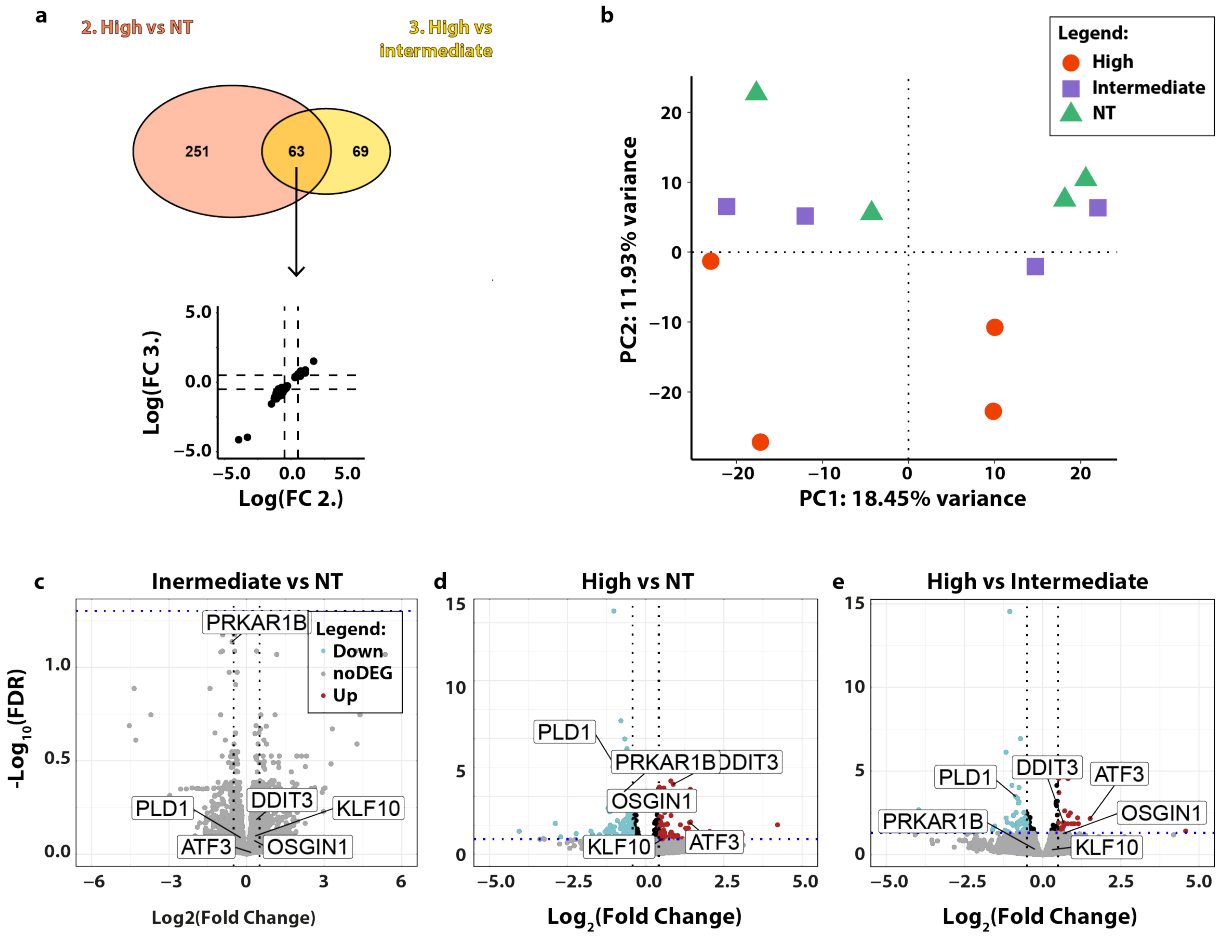


Figure 3.23: **DEGs in 3TS 5'UTR sample. (a)** Intersection of the DEG among the three comparisons. Top, the Venn plot highlights the overlap between High vs NT and Intermediate vs NT populations. The comparison Intermediate vs NT did not show any significantly DE gene, confirming that with this plasmid design the cellular machinery is less impacted by little DNA amount as compared to both noTS and 3TS 3'UTR conditions (**Fig. 3.21a**, **Fig. 3.22a**). Bottom, scatter dot plots show the positive relation between the Log_2(FC) of the common genes. **(b)** PCA plot of the 4 biological replicates. Only High population is distinguishable from the others, consistently with what reported by the Venn plot (a). Volcano plots of genes distribution in Intermediate vs NT **(c)**, High vs NT **(d)** and High vs Intermediate **(e)** comparisons. Light blue dots (top left quadrants) represent the significantly downregulated genes (Down); red dots (top right quadrants) represent the significantly upregulated genes (Up); grey dots (below the horizontal blue dashed line) represent the genes that do not undergo significant change (noDEG, FDR $<$ 0.05); black dots (between the two vertical dashed lines and above the blue dashed line) represent the genes that undergo significant change but too little to be considered biologically relevant ($|Log_2(FC)| \leq 0.5$). Highlighted are transcription factors related to stress response (DDIT3 and ATF3), negative regulators of cellular growth in response to stress (KLF10 and OSGIN1), genes involved in signal transduction and cellular metabolism (PLD1 and PRKAR1B). The latter are downregulated. Data were collected 48 hours post-transfection. N = 4 biological replicates.

Finally, we investigated the common DEGs between the three transfections. We focused on the comparison High *vs* NT, as they are the most diverse populations and show the highest differential gene expression profile. In the comparison, we identified 46 genes with FDR < 0.05 (**Fig. 3.24a**). By visualizing their behavior (**Fig. 3.24b**), all transfection conditions reported consistent upregulation or downregulation of the same genes. The number of downregulated genes is higher, suggesting that cells undergo a global slowdown in transcription when they are overloaded with exogenous genetic payloads. Among the upregulated genes we find protein coding genes related to stress response (e.g. ATF3 and DDIT3) and to the inhibition of cell growth (e.g. KLF10 and OSGIN1). Many differentially expressed genes are long non-coding RNAs (lncRNA or pseudogenes) and their role will require further investigation.

Overall, the results presented in this section suggest that exogenous genes delivery induces a change in the transcriptional landscape and correlates with a potential stress imposed to the cells. This unwanted effect varies depending on the plasmid design and, as already suggested by our prior studies (see Section 3.2.5), the use of endogenous miRNAs for exogenous genes regulation has the power to mitigate burden. Moreover, we interestingly observed that genes related to stress response are differentially expressed in transfected cells, consistently with what was previously observed in bacteria with similar analysis [64]. These recent results will be further investigated.

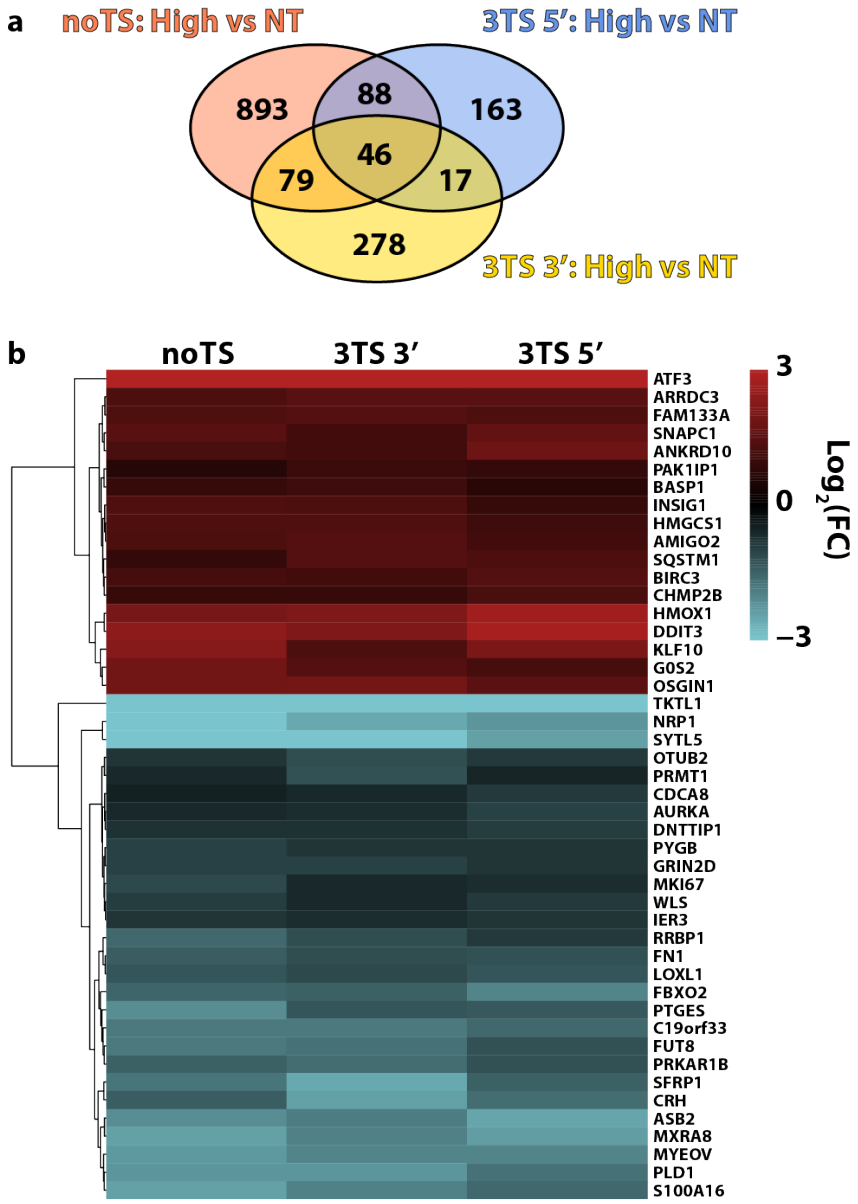


Figure 3.24: **DEGs in High vs NT populations in the three transfection conditions. (a)** Intersection of the differentially expressed genes among the three transfection conditions. Top, the Venn plot highlights the overlap between noTS, 3TS 3' and 3TS 5' transfections for the High vs NT populations comparison. The comparison reveals 46 common DE genes. **(b)** Heatmap of the differentially expressed genes. In red are the upregulated genes and in light blue the downregulated ones. The majority of the genes are downregulated and among the upregulated ones there are genes related to stress response (DDIT3 and ATF3) and arrest of cell growth (KLF10 and OSGIN1). Data were collected 48 hours post-transfection. N = 4 biological replicates.

3.4.3 Differential protein expression analysis confirms observations at gene level

Similarly to the RNA-seq data, we compared the 3 sorted populations within the same experimental conditions (or samples, **Table 3.2**). We visualized common protein regulation among the different comparisons in the three transfection conditions (**Fig. 3.25a,c,e**). The comparisons among noTS sorted populations show a good degree of similarity, especially between High *vs* NT and both Intermediate *vs* NT and High *vs* Intermediate populations. In 3TS 3' and 3TS 5' the number of DEPs is reduced and the intersections reveal that the highest correlation is between High *vs* NT and High *vs* Intermediate populations, confirming that only high plasmid uptake leads to a significant change in protein expression. The common DEPs between High *vs* NT ∩ High *vs* Intermediate and High vs NT ∩ Intermediate *vs* NT populations are positively correlated (**Fig. 3.25a,c,e** bottom). However, the common proteins between Intermediate *vs* NT and High *vs* Intermediate populations show inverse correlation (**Fig. 3.25a,c,e** bottom middle plots). This counterintuitive but consistent behavior requires further investigation.

The PCA shows that, as for gene level (**Fig. 3.21b, 3.22b, 3.23b**), the differences between the populations are little and are overcome by the replicate-to-replicate variability: the principal component 1 (PC1, x axis) implies replicate heterogeneity as the 5 replicates of each population occupy all the horizontal space. Nevertheless, we can appreciate population clusters also along the vertical space (PC2, y axis). In the noTS condition (**Fig. 3.25b**), the NT population in 2 out of 5 replicates is the most distant from the other samples (bottom left of the Cartesian space). In the 3TS 3' condition, the differences between the populations are shrunk and they appear all equally distant (**Fig. 3.25d**), as already observed at gene level (**Fig. 3.22b**). Finally, in the 3TS 5' condition (**Fig. 3.25f**), the High population is the only one clearly distinguishable from the others, once again in accordance with RNA-seq data (**Fig. 3.23b**).

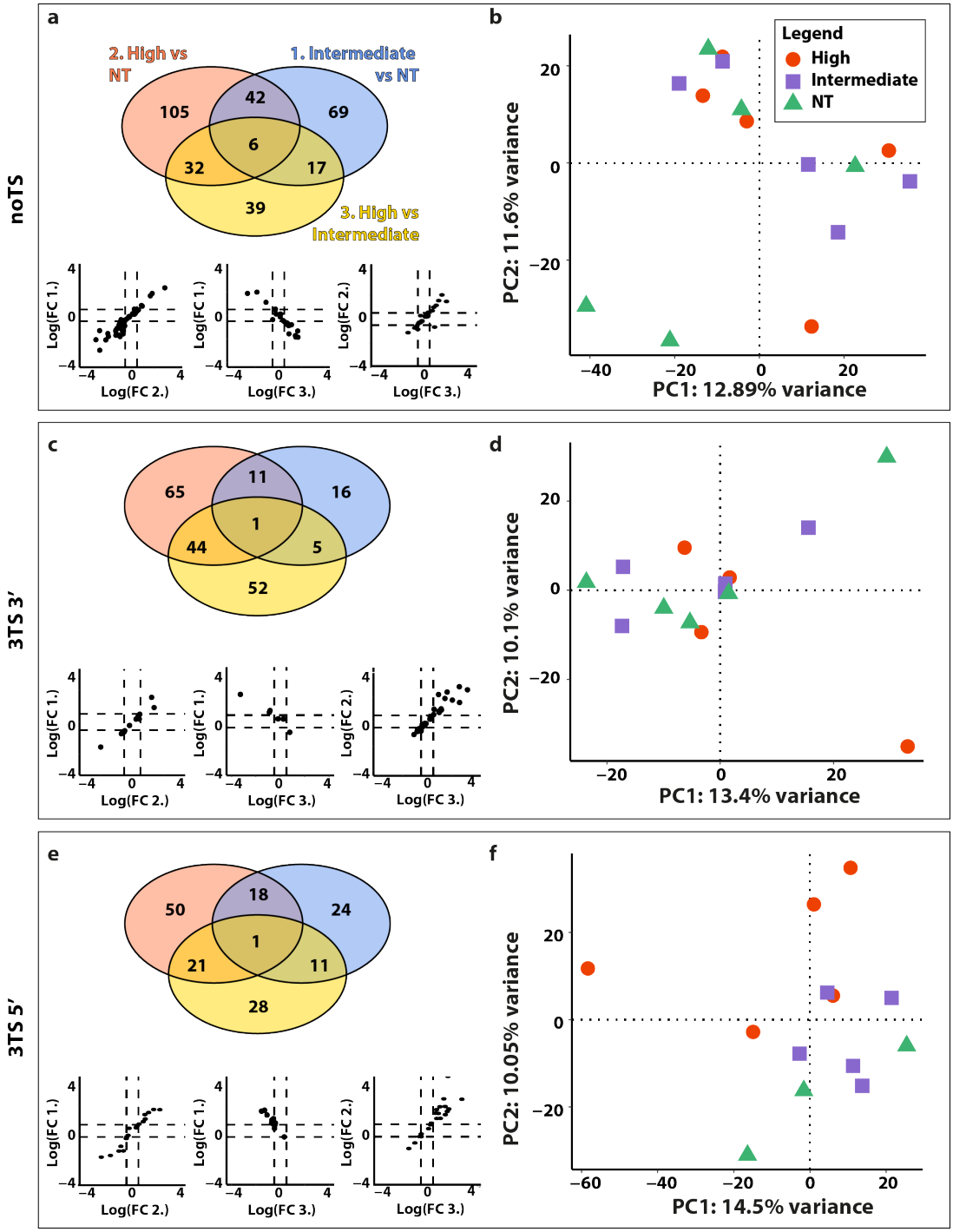


Figure 3.25: **DEPs in the three transfection conditions. (a,c,e)** Top, Venn plot highlights the overlap of the DEPs among the different populations. Bottom, scatter dot plots show the relation between the Log_2(FC) of the common genes. **(b,d,f)** PCA plots of the three transfection conditions. N = 5 biological replicates; N = 3 for 3TS 5' NT population.

Finally, we looked at the intersection between the different transfection conditions focusing on High *vs* NT population comparison, as we did for the differential gene ex-

pression analysis (**Fig. 3.26**). The Venn plot shows very little share of DEPs between 3TS 5'UTR and the other two transfection conditions, consistent with the fact that it is the plasmid design that generates the smallest gene expression burden. The other two transfection conditions share 18 common genes.

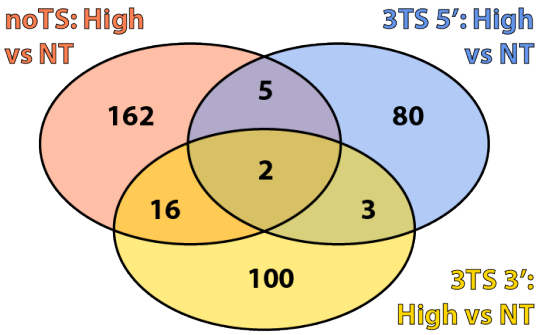


Figure 3.26: **Intersection of DEPs in High *vs* NT population comparison among the three transfection conditions.**

In the future, we will analyze the intersection between RNA-seq and proteomics to investigate whether there is a good correlation between DEGs and DEPs, or if the involvement of post-transcriptional regulation in controlling gene expression causes the differential expression of different factors at the gene and protein levels [177, 178].

3.4.4 RNA-seq reveals dysregulated pathways related to stress response and cellular growth

So far we focused on the identification of individual genes and proteins that are differentially expressed due to gene expression burden, without considering the system as a whole. To understand if gene expression burden may have an impact on cellular metabolism and pathways expression we did the Gene Set Enrichment Analysis (GSEA) on RNA-seq gene expression analysis results. GSEA is a computational method that determines whether two biological states exhibit statistically significant differences based on the gene expression profile [179]. It is a functional class scoring (FCS) tool and, as such, it considers the gene-level statistics or the Log_2 of the fold changes for all genes from the differential expression analyses and then checks whether gene sets for specific biological pathways are enriched among the positive or negative fold changes. The hypothesis of FCS methods is that despite significant changes in individual genes (that will be detected via over-representation analysis methods), weaker but coordinated changes in sets of functionally

related genes (pathways) can also have significant effects. Thus, rather than setting an arbitrary threshold to identify significantly DE genes, all genes are considered in the analysis. The gene-level statistics from the data-set are aggregated to generate a single pathway-level statistic and statistical significance of each pathway is reported. Pathways are classified with the Gene Ontology (GO) domains and annotations, which inform about their role and the molecular species involved. When over-representation analysis (ORA) is employed, often times there are hundreds of statistically significant GO terms per data-set. Comparing enriched categories between a large number of analyses and identifying the term within the GO hierarchy with the most connections is challenging. Furthermore, ascertaining biological themes representative of the samples can be highly subjective from the interpretation of the enriched categories. The advantage of GSEA over ORA is that by considering all the genes variations the overall results on the pathways identification is more robust.

We began by the analysis of the noTS transfection condition (**Fig. 3.27**). For each comparison among the 3 sorted populations, we identified clusters of biological processes with different behavior due to higher dosage of exogenous gene expression (**Fig. 3.27a-c**). To be noted that only the 30 most significant pathways according to the FDR are displayed in the figures for each comparison. As expected, protein translation related pathways, as well as RNA and protein catabolic processes are upregulated, consistent with the higher number of genes in the system due to transfection. Importantly, High *vs* NT populations comparison reveals a cluster of stress response pathways which are highly upregulated. On one side we observe the aforementioned effects in the comparison between transfected *vs* NT populations, but on the other side, in **Fig. 3.27c** we can see that cell division and cytoskeleton organization are inhibited by high amount of transgene expression. Detail of the most interesting dysregulated pathways is reported in **Fig. 3.27d**. Here we show both the enriched biological processes and cellular components, according to GO annotation. Again, the pattern shows enhanced protein translation and genes catabolism - supported by an upregulation in ribosomal and proteasomal components - and decreased cellular division - supported by lower amount of mitotic spindle and cytoskeleton components (such the microtubules). Interestingly, pathways related to unfolded protein response and transcription in response to stress are upregulated.

In 3TS 3'UTR transfection, the pathways related to protein translation are downmodulated in Intermediate *vs* NT and High *vs* NT population comparisons (**Fig. 3.28**). Similarly to the noTS condition, in 3TS 3'UTR transfection high transgene expression triggers transcriptional activation in response to stress (**Fig. 3.28b**), and decreased mitosis (**Fig. 3.28c**). We hypothesize that the negative modulation of the cell respiration pathways observed in **Fig. 3.28b-c**, is due to reduced protein production and mitosis, two high ATP demanding processes in the cell. Nevertheless, this is in contrast with the fact that transfected cells should be producing more proteins, demanding high ATP amount. A more detailed representation of significantly dysregulated pathways confirms a downregulation of protein production processes and of cell division (**Fig. 3.28d**). Gene expression burden results in upregulated transcriptional response to stress and in endoplasmatic reticulum stress. Finally, an increase in the cellular components of the p-bodies is consistent with the high mRNA degradation induced by miRNA action.

Lastly, in 3TS 5'UTR transfection the number of dysregulated pathways per comparison is lower than in the other transfection conditions (**Fig. 3.29a-c**). Intermediate transfected cells population shows a mild increase in protein translation (**Fig. 3.29a**), which is more enhanced in High transfected population (**Fig. 3.29c**). Once again, stress response pathways emerge in High transfected population (**Fig. 3.29b**). Looking at a more detailed representation of the pathways (**Fig. 3.29d**), we further appreciate that none of them is dysregulated in Intermediate *vs* NT populations comparison, confirming the negligible burden imposed by this plasmid configuration when plasmid uptake is low (Intermediate transfected population). In general there is a downregulation of the pathways associated with cell division and an upregulation of the pathways associated with protein translation.

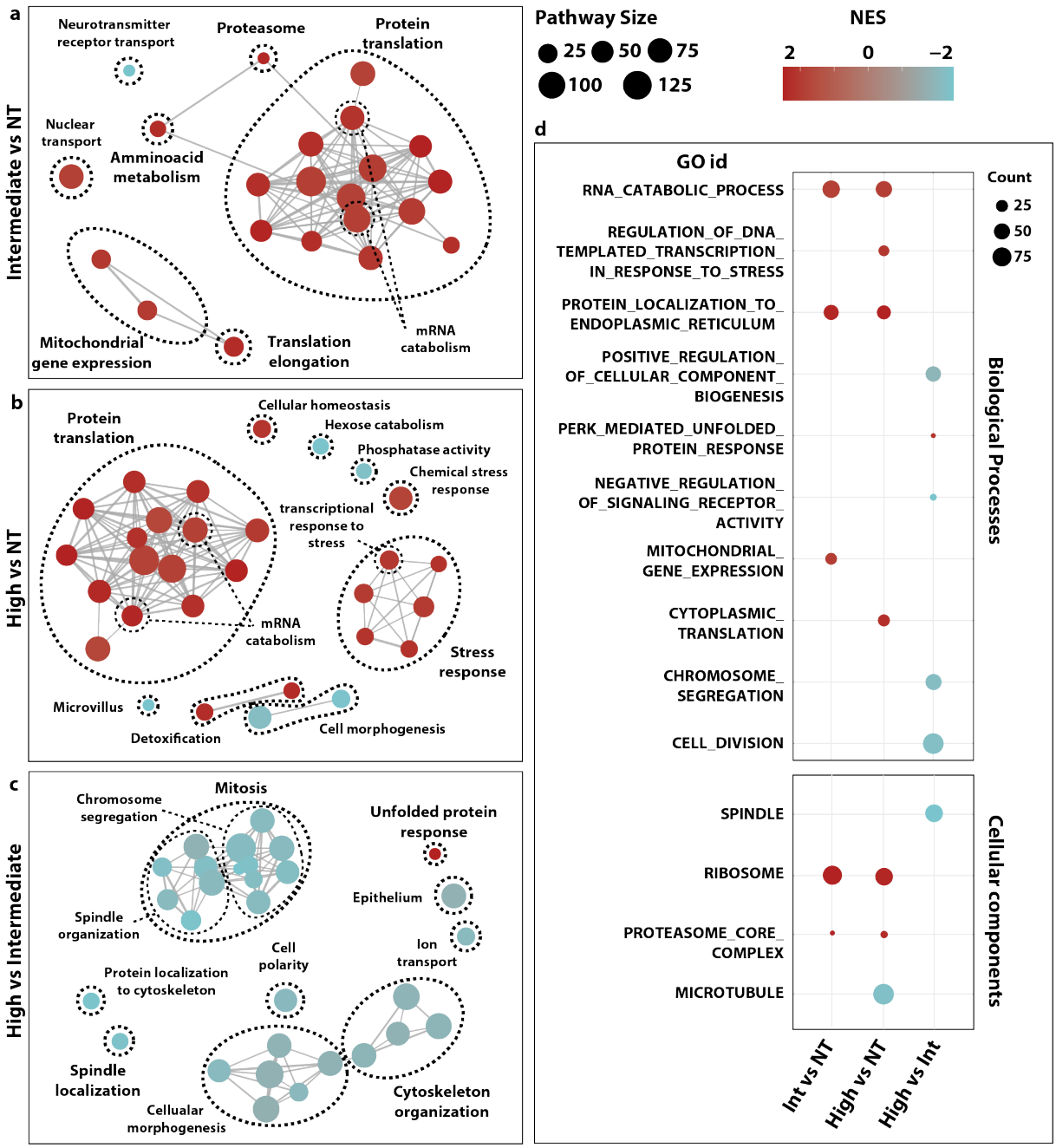


Figure 3.27: **GSEA analysis of RNA-seq data of noTS transfection.** **(a,b,c)** Enriched pathway clusters are highlighted for Intermediate *vs* WT, High *vs* WT and High *vs* Intermediate comparisons respectively. Upregulated pathways are marked in red; downregulated pathways are marked in blue. Size of the dots represents the number of genes involved in the pathway (size of the pathway). **(d)** Representation of a subset of significantly enriched pathways across the population comparisons. Size of the dots represents the genes count in the comparison. FDR < 0.05. N $= 4$ biological replicates.

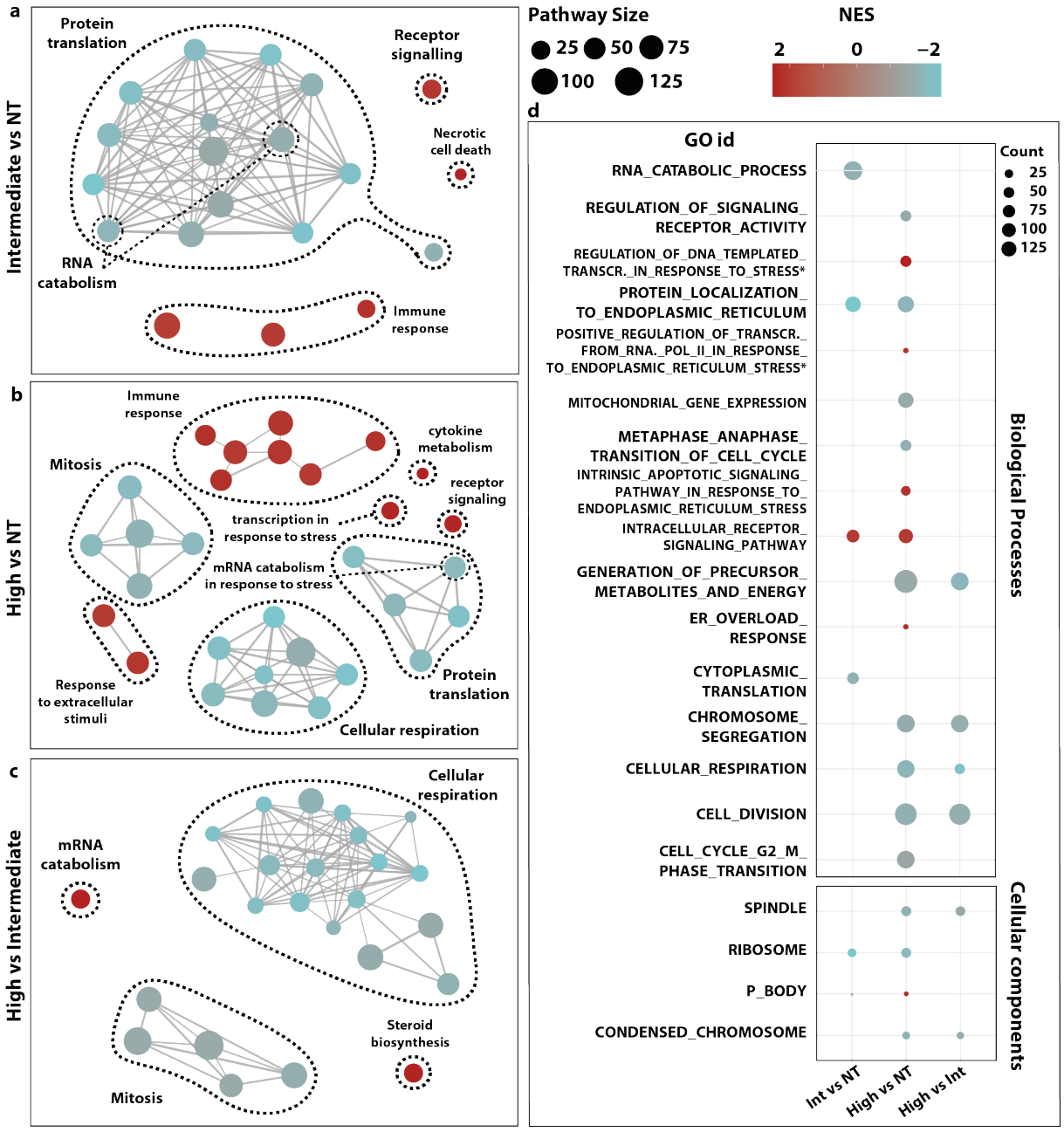


Figure 3.28: **GSEA analysis of RNA-seq data of 3TS 3'UTR transfection. (a,b,c)** Enriched pathway clusters are highlighted for Intermediate *vs* WT, High *vs* WT and High *vs* Intermediate comparisons respectively. Upregulated pathways are marked in red; downregulated pathways are marked in blue. Size of the dots represents the number of genes involved in the pathway (size of the pathway). **(d)** Representation of a subset of significantly enriched pathways across the population comparisons. Size of the dots represents the genes count in the comparison. FDR < 0.05. N $= 4$ biological replicates.

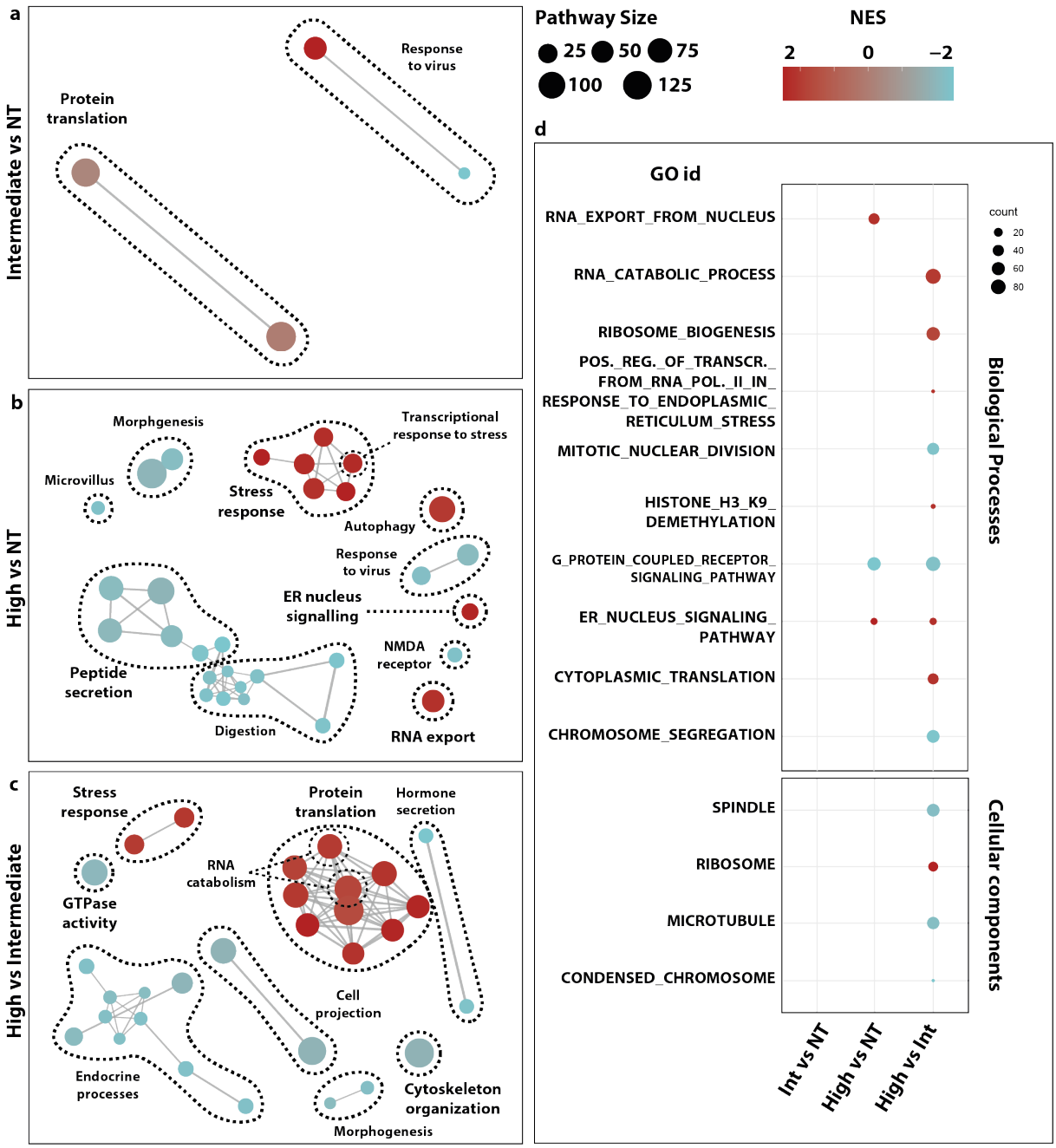


Figure 3.29: **GSEA analysis of RNA-seq data of 3TS 5'UTR transfection. (a,b,c)** Enriched pathway clusters are highlighted for Intermediate *vs* WT, High *vs* WT and High *vs* Intermediate comparisons respectively. Upregulated pathways are marked in red; downregulated pathways are marked in blue. Size of the dots represents the number of genes involved in the pathway (size of the pathway). **(d)** Representation of a subset of significantly enriched pathways across the population comparisons. Size of the dots represents the genes count in the comparison. FDR < 0.05. N $= 4$ biological replicates.

Finally, we checked for enriched pathways common for all the transfection conditions focusing on the High *vs* NT population comparison (**Fig. 3.30**). We identified 4 pathways, all of which exhibited consistent behavior across the three transfections (**Fig. 3.30a**). Three of these pathways are correlated with response to stress and are upregulated: the *Tumor Necrosis Factor mediated signaling pathway*, the *regulation of DNA templated transcription in response to stress*, and the *NIK NF-Kb signaling*. The only downregulated pathway relates with response to neurotransmitters, which is not biologically significant in H1299 cells, therefore it might result dysregulated as unnecessary genes are downmodulated by the cell in response to gene expression overload. However, this hypothesis will require further investigation. It is also the only pathway that does not share genes with the other three (**Fig. 3.30b-c**). From the network map it can be noted the presence of ATF3 and DDIT3, identified by the differential gene expression analysis are stress-related genes upregulated in all transfection conditions. Casp4 is a pro-apoptotic gene [180], which supports the idea that transfection blocks mitosis and cellular growth. We focused on the pathway related to the transcriptional response to stress and displayed it with the GSEA plots for the three transfections (**Fig. 3.30d**). The GSEA plots show the running Enrichment Score (ES) for the gene set as the analysis walks down the ranked gene list. The black bars in the middle part of the plot correspond to where the members of the gene set appear in the ranked list of genes, and they are ranked according to the Log_{10} of FDR values adjusted by the sign of fold change in this way: on the left (correspond to red) there are the most upregulated genes and on the right (correspond to blue) there are the most downregulated. The running ES shows that this gene set is positively enriched as there is a distinct peak at the beginning (left part) of the ranked list, especially in noTS and 3TS 5'UTR samples (top and bottom plots).

Overall, the results reflect the conclusions depicted from the differential gene expression analysis: gene expression burden activates stress response and inhibits cellular growth.

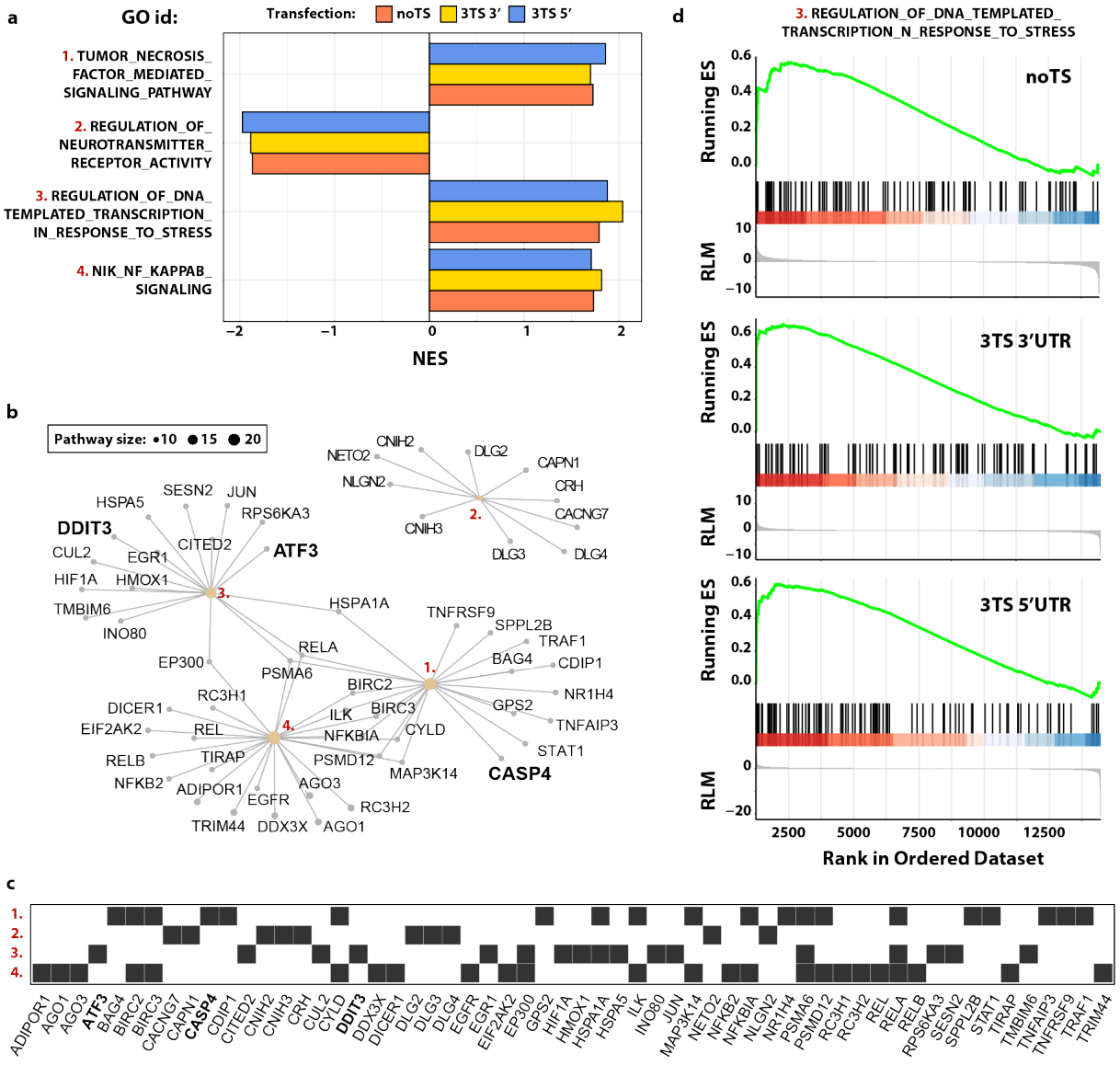


Figure 3.30: **Common enriched pathways among transfection conditions in the High *vs* NT comparison as for RNA-seq data. (a)** Bar plot reveals the same trend of the 4 dysregulated pathways in the 4 samples. NES: normalized enrichment score. **(b,c)** Detailed network and genes' list of the 4 pathways. The stress response-related transcription factors DDIT3 and ATF3 have been highlighted. Same for the pro-apoptotic gene Casp4. **(d)** GSEA plots of the *regulation of DNA templated transcription in response to stress* pathway in the 3 transfections. ES: enrichment score. RLM: ranked list metric. FDR < 0.05. N $= 4$ biological replicates.

3.4.5 Dysregulated pathways according to proteomics data.

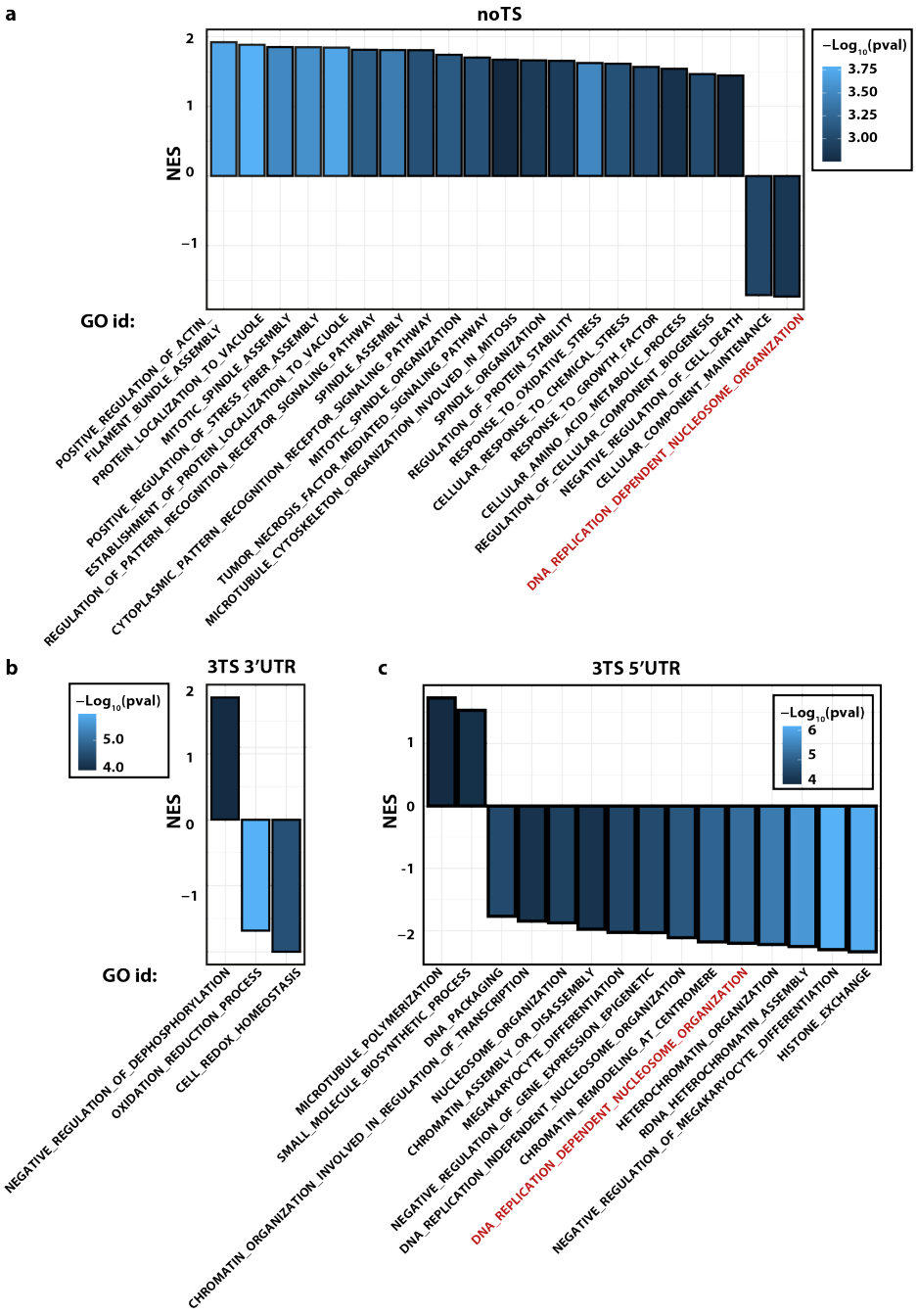


Figure 3.31: **GSEA analysis on proteomics data in the three transfection conditions for the comparison High vs NT population.** (a,b,c) Bar plots of dysregulated pathways in noTS, 3TS 3'UTR and 3TS 5'UTR transfections respectively. Height of the bar corresponds to NES and color corresponds to -Log_{10}(pvalue). In red is highlighted the only common pathway NES: normalized enrichment score. FDR < 0.1. N $= 5$ biological replicates.

We followed up by running the GSEA analysis on the proteomics data for the

High *vs* NT population comparison in the three transfection conditions (**Fig. 3.31**). Only pathways with FDR < 0.1 were considered significantly dysregulated. The higher number of dysregulated pathways was once again in the noTS transfection condition, while the lowest number in the 3TS 3'UTR transfection. Little correlation is found among the three transfections. The only shared pathways is the *DNA replication-dependent nucleosome organization* which is downregulated both in noTS and in 3TS 5'UTR samples (highlighted in red). This pathway is correlated with DNA replication, suggesting that cells are dividing less. In 3TS 5'UTR transfection, this hypothesis is corroborated by other downregulated pathways such as *Chromatin remodeling at centromer*, while in noTS sample pathways related with mitosis are upregulated. The contrasting results will need to be investigated further.

3.4.6 The intersection between the *omics*

Finally, we started exploring the intersection between proteomics and RNA-seq GSEA results with the noTS transfection condition for High *vs* NT population comparison (**Fig. 3.32**). The intersection shows only 2 common pathways: the *Tumor Necrosis Factor mediated signaling pathway* and the *cellular response to chemical stress*. The latter does not seem to have much to do with potential consequences of transfection, however it leads to apoptosis, like other pathways dysregulated at gene level (**Fig. 3.30**). The first one instead appeared dysregulated in all transfection conditions at gene level and is related to stress response. Further focus on the *omics* intersection will give interesting insights on the systemic effect of gene expression burden.

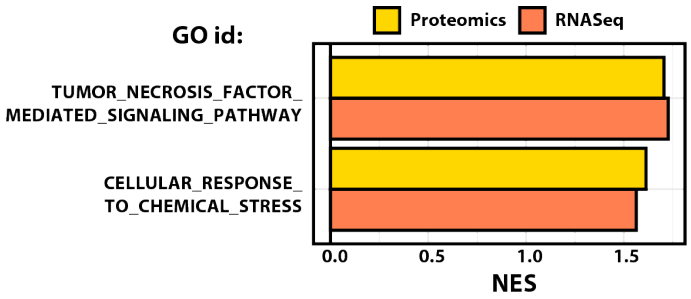


Figure 3.32: **Intersection between GSEA results on RNA-seq and proteomics data.** NES: normalized enrichment score. Proteomics: N = 5 biological replicates; FDR $<$ 0.1. RNA-seq: N = 4 biological replicates; FDR < 0.05

3.5 Resource pools: a challenge for synthetic biology

The main result shown in this Chapter is that mammalian cells are susceptible to burden imposed by genetic payloads and this leads to both unwanted gene expression coupling and to impaired endogenous gene expression. We showed that the effect is cell-type independent and it is due to limited resource pools both at transcriptional and translational level Section 3.2. We demonstrated that the sharing of limited cellular resources represents a general bottleneck for the predictability and performance of transiently transfected synthetic circuits in mammalian cells, with important consequences for mammalian synthetic biology and biotechnology applications. Due to resource limitations, transient heterologous gene expression results in the coupling of independent exogenous genes and affects the expression of endogenous ones.

We dug into several aspects of gene expression burden showing that: *(i)* burden mitigation can be achieved using resource aware circuits design (Section 3.2); *(ii)* important regulatory mechanisms, such as post-transcriptional regulation by miRNAs, have both a direct effect on the target and an indirect opposite effect on the co-expressed genes (Section 3.3); distress caused by transgenes delivery causes the impairment of genes and proteins expression in host cells, with important consequences on intracellular pathways (Section 3.4).

3.5.1 Monitoring and mitigation of gene expression burden

With our project centered on the characterization and modelling of gene expression burden we presented a simple mean by which burden and cellular production capacity can be quantified: the *capacity monitor* gene used in the study efficiently revealed resource shortage in different experimental setups both at transcriptional and translational level.

We described a modelling framework that captured the indirect interdependence of gene expression in a resource-limited context and suggested that an incoherent feedforward loop (iFFL) is a particularly well-suited circuit motif for mitigating burden effects. As described in Section 1.3, computational simulations can efficiently direct experiments towards the right intuition. The incoherent feedforward loop (iFFL) itself is one of the core gene regulatory motifs in biology, and unsurprisingly it has served as inspiration for many

synthetic genetic circuits that exploit its adaptation properties [70, 137, 150, 151]. We adopted a miRNA implementation of iFFL circuits for the purpose of burden mitigation. Previously, synthetic miRNA-based iFFLs have been demonstrated to increase robustness to gene dosage variability [70] and gene expression noise [150, 181]. The implementation of these circuits features the use of endogenous miRNAs as elementary components of the engineered iFFL device, a versatile hybrid design that allows burden mitigation to be achieved across different cell-lines with minimal resource requirements. To our knowledge this is the first time that endogenous miRNAs are used to perform logic computation instead of just serving as input to the system [50, 76].

Regardless of their origin, miRNA-based iFFL circuits were shown here to decouple the expression of both exogenous and endogenous genes. This prompted us to investigate the molecular mechanisms involved in miRNA-dependent resource re-allocation (Section 3.3). Contextually to our study, Jones *et al* showed that an implementation of iFFL could alternatively be achieved using endoribonucleases [68]. Their study demonstrates that CasE endoribonucleases can be employed to mitigate burden in a highly efficient and tunable manner by changing number and location of recognition sites, similarly to our miRNA configuration. We opted for a miRNA-based approach as they are either endogenously expressed or encoded intronically (synthetic miRNAs variant, data not shown, refer to [67]) and therefore their expression channels a negligible amount of resources. Moreover, iFFL circuits that exploit endogenous miRNAs enable cell-type specificity [76, 129], whereas synthetic miRNAs enable portability of circuits across different cell lines. Flexibility at the sequence level allows scaling up to many orthogonally operating circuits. The specificity of a miRNA can be easily engineered to target any synthetic or endogenous gene without the need to engineer the target itself enabling programmability [182–184]. Finally, tunability of repression strength can be easily achieved both through the number and the placement of the targets, and can be used to enhance adaptation to variations in resource availability.

One of the limitations of our system is that stronger repression will yield lower expression levels of the gene of interest (*GOI*, **Fig. 3.13c**). This tradeoff is unavoidable, and is an inherent limitation to all implementations of iFFL-based burden-mitigation circuits [68]. Another potential limitation of miRNA-based iFFL circuits is the diversion of

endogenous miRNAs from native targets to synthetic ones. Although this is not what we observe in our miR-31 iFFL (**Fig. 3.15d**), this may however give rise to an inevitable trade-off similar to what has been observed for competing endogenous RNA (ceRNA) [185]. ceRNAs are known to naturally regulate other RNAs by competing for miRNA-binding. To attempt to remedy this, one could use partially complementary TS, which would decrease the affinity of the miRNA to the target and diminish the competition. However this would make the system less efficient and potentially decrease the mitigation effect. Alternatively, the incorporation of multiple TS that respond to different highly expressed miRNAs would distribute the competition between multiple miRNAs and reduce the detrimental effects on their native targets.

Besides iFFLs, negative feedback motifs can also be used to mitigate resource burden [70, 186], as was shown in a series of studies in *E. coli* [63, 64, 66, 72]. While negative feedback circuits possess well-established robustness properties, iFFL circuits have several advantages for burden mitigation. In particular, iFFL circuits are considerably simpler to implement and easier to tune than negative feedback circuits, which usually require more components and can become dynamically unstable if not properly designed and tuned. In terms of dynamic response, iFFL circuits are also generally faster in rejecting disturbances like a sudden change in resource availability. Indeed iFFL regulation responds to the disturbance itself, while negative feedback begins to act only after the impact of the disturbance on the regulated output has been detected.

Ultimately, the goal of gene circuit engineering is the creation of cell lines that stably express circuits of interest. Our work focused on the effects of limited resources as induced by transient transfection, but in the future it will be necessary to investigate if similar effects also occur in the context of genomic integration of highly expressed genes. Moreover, while using a transiently transfected *capacity monitor* enables the quantification of cellular expression capacity by providing a comparative measure of the geometric mean of free resources in a burdened population relative to a minimally burdened baseline population, stable integration of the capacity monitor would permit a more direct measure of free cellular resources. Using burden-aware designs, cell therapies that rely on finely-tuned expression and secretion of therapeutic molecules can now be engineered with resource-aware circuits. Our findings suggest that, when choosing a host cell line,

one of the key factors to consider should be its transcriptional and translational capacity, not only in terms of productivity, but also in terms of the ability of the cells to maintain their fitness while performing their engineered function [144].

Our findings not only pave the way to more realistic output predictions and optimal synthetic construct design in mammalian cells, but also offer a portable design capable of enhancing the insulation of transgene expression. This will contribute to the development of robust-by-design mammalian synthetic circuits, with important implications for basic science and applications in industrial biotechnology and medical therapy.

3.5.2 Resource re-allocation by post-transcriptional regulation

Our observations on the ability of post-transcriptional regulators to affect co-expressed gene's level, prompted us to investigate the molecular mechanisms involved. Our experiments were designed to verify our hypotheses of ribosome redistribution and queuing up of for mRNA degradation.

The results shown in Section 3.2.4, reveal that EGFP expression always correlates with mKate downregulation, also when mKate levels are restored with miR-31 inhibition. We demonstrated that with a higher number of miRNA target sites the downregulation of the target gene is more pronounced, as well as when the TS are at 5'UTR, probably because the binding of the miRISC complex avoid ribosome binding by steric hindrance.

We confirmed this hypothesis with polysome sequencing. We observed modified translational profiles suggesting enhanced translation both for transfected genes (EGFP) and for endogenous genes (CyCA2, eIF4E, GAPDH). Of note, GAPDH is commonly used as internal reference for qPCR and western blot experiments as its expression is considered particularly stable across samples [187, 188]. Here we show that also the most robust genes are hit by resource re-allocation. The effect is stronger with stronger mKate degradation (3TS 5'UTR).

mRNA half-life measurement upon DRB treatment indicated that there is also a queuing effect on the mRNA degradation machinery on endogenous genes. Unfortunately, DRB treatment was ineffective on transfected genes (EGFP) and GAPDH, therefore future investigation to confirm that also transfected genes are affected will be needed.

With our experiments, we investigated the molecular mechanisms involved in re-

source re-allocation dependent on miRNA action when we add fully complementary TS to our exogenous genes. However full complementarity of the TS is observed in plants, while in animals it is far more common to find partial complementarity (SEED region of 7/8 nt) [147]. We speculate that the same molecular mechanisms play a role in systems with partial complementarity of the targets, however this will need experimental validation in the future.

Starting from our experimental results and the miRNAs mechanism of action reported in literature, we implemented a reaction based model to recapitulate our experimental findings and gather more insights into the system's dynamics. We are currently testing our model, running simulations to evaluate the qualitative trend of our species expression, as measure of the ability of the model to accurately describe our system. Once we will have concluded this part, we will run the the sensitivity analysis on our rates to determine which ones are more determinant for the dynamic species in our model, with specific focus on mRNA (m_j), protein (p_j) and miRNA (mi_q). The sensitivity analysis increases the understanding of the dependency between rates and species. At the same time, it further detects errors in the model, if unexpected and senseless dependencies are shown. For example, we expect our dynamic species to be affected by their own production and degradation rates and we expect the parameters that describe miRNA dynamics to have a great impact on the whole system, including the size of resource pools. On the other hand, we expect other parameters, such as protein degradation rate (λ_j) to have almost no impact. We envision that our model will offer a useful starting point to implement a tool to simulate and predict optimal genetic circuits design that use endogenous miRNAs for output regulation.

The success in proving the presence of cellular burden in mammalian cells and in outlining some mechanisms involved, pushed us to challenge a physiological model to see whether non-regulatory interactions are part of the cellular pathways to gain new insights in cellular functions. For this very early-stage project, we chose a compartmentalized system to narrow down the space and the amount of cellular components involved. Starting from the evidence that miR-181a selectively maturates in hippocampal dendrites, as described by Sambadan *et al* [189], we decided to inhibit its expression to cause an upregulation of its targets. The upregulation of the targets would supposedly

drive a redistribution of available ribosomes on the transcripts. One of the main targets of miRNA-181a is CamKIIa, an important synaptic modulator.

We verified that CamKIIa effectively increases upon miR-181a inhibition by immunofluorescence (IF) on mice hippocampal neuron transfected with miR-181a inhibitor, a scramble inhibitor and a control with no inhibitor (**Fig. 3.33a**). This experiment was performed in collaboration with Dr Barberis' lab (IIT).

To proceed with this study, we are optimizing the experimental conditions to appreciate the co-localization of transcripts stained with single-molecule-FISH probes (smFISH) and ribosomes stained with IF. We are optimizing the microscopy conditions to

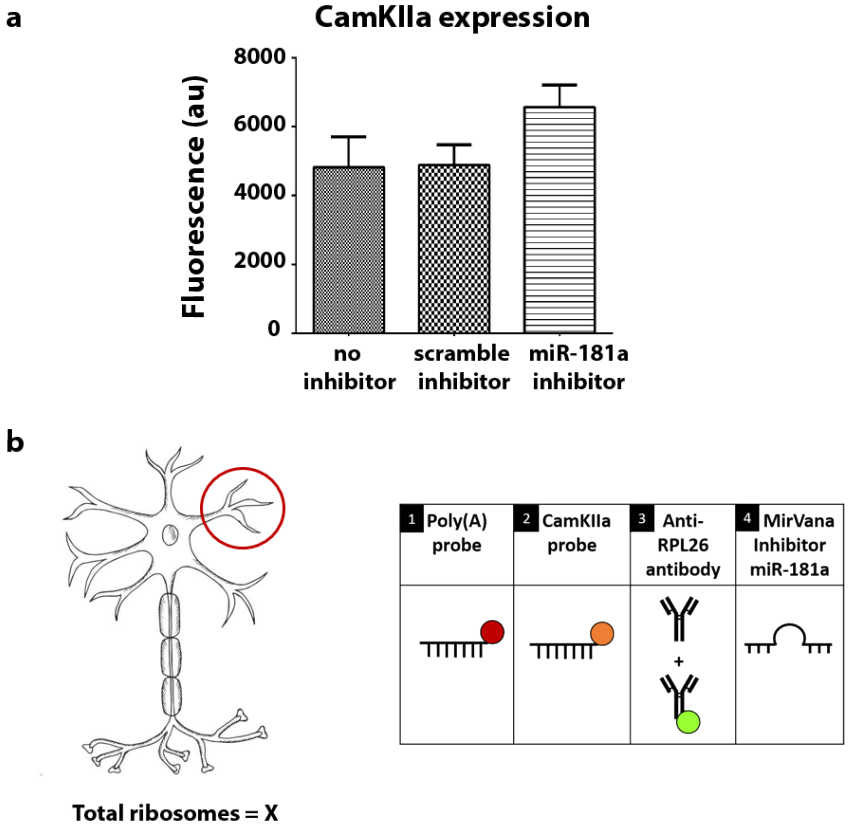


Figure 3.33: **Study of resource redistribution in physiological conditions. (a)** Fluorescence intensity in dendrites of CamKIIa stained with Alexa-568. Data were acquired 48 hours post transfection and are plotted +/- SE. N = 3 biological replicates. **(b)** Schematic representation of the experiment to study resource re-allocation in mouse hippocampal neuron dendrites. Neurons will be transfected with miR-181a inhibitor and stained with smFISH probes – one set for all transcripts (poly(A) probe) coupled with Quasar-670 and one set specific for CamKIIa coupled with TAMRA – and with anti-ribosomal protein L26 primary antibody plus secondary antibody coupled with Alexa-488.

allow the measurement of ribosomal signal coupled with transcripts. What we expect is to have a decrease of the ribosomal signal per transcript as we upregulate CamKIIa expression by inhibiting miR-181a. Optical super-resolution microscopy technologies - such as STED, SIM and STORM - offer good possibilities to study molecules co-localization by fluorescence detection [190–192]. A schematic representation of the experimental setup is provided in **Fig. 3.33b**.

3.5.3 Towards burden-free systems

Finally, our work on the global effect of gene expression burden on host cells gave novel insights into cellular functions in response to genetic overload.

To study the systemic impact of gene expression burden we run both RNA-seq and proteomics analysis on transfected H1299 samples. The analyses revealed both differentially expressed genes and proteins and impaired cellular pathways. We highlighted how genes involved in stress-response, cell growth and metabolism are differentially expressed in transfected cells as compared to NT cells population (**Fig. 3.21-3.26**). Specifically stress response seems enhanced, while cellular growth and metabolism seem to be slowed down. Consistently with our results from the previous projects that showed that by adding endogenous miRNA TS gene expression burden can be mitigated (see Section 3.2.5), our plasmid designs with miR-31 TS showed the lowest impact both on gene and protein expression. The lower impact was reflected by the number of DEGs and DEPs and by the amount of DNA that was required to induce the differential expression (cells in the Intermediate transfected population showed stress response only in the noTS sample). In our hands, the higher the *miRNA sensor* repression (3TS 5'UTR transfection), the lower the impact on endogenous genes and protein expression.

Besides the protein coding genes, RNA-seq revealed a large number of differentially expressed long non-coding RNAs. Their role is unknown and it will be interesting to interrogate the genomic locus in which they are located. Another hypothesis is that they are transcribed as readthrough transcripts in response to stress, as reported in [193]. However this hypothesis needs to be carefully investigated according to the methods described in their work.

We ran differential gene and protein expression analyses separately, and observed

a maximum correlation of 40% between the fold changes [177]. This is consistent with what is generally observed with the intersection between the *omics* [194]. We are currently evaluating tools to integrate our multi-omics data in a reliable manner, for example using mixOmics tool [178]. mixOmics offers a novel analytical tool particularly insightful to identify molecular signatures across those multiple data-sets. It uses a multivariate method with more relaxed assumptions about data distribution, data size and data range than univariate methods. This is an advantage in the comparison of multiple data sets that are not necessarily linearly correlated. The integration of multiple molecular profiling technologies creates novel opportunities for biomedical applications, improving the biological understanding of the cellular context [195].

We inferred global pathway dysregulation using GSEA. Interesting results came from the RNA-seq, with upregulation of pathways related protein production and stress response pathways and downregulation of pathways related to cell division (mitosis). Dysregulation of stress-related pathways is consistent with what observed in bacteria by Ceroni *et al* [64].

Unfortunately, GSEA on proteomics data does not seem to correlate well with RNA-seq results. This could be due to several reasons. The presence of outliers in some transfection condition (e.g. 2 NT replicates in 3TS 5'UTR transfection) might reduce the statistical significance of the results, causing the exclusion of otherwise significantly enriched pathways. A more careful evaluation of the data in our hands will tell if the analysis can be refined to this regard. On the other side, between gene and protein level there are post-transcriptional and translational regulation levels, which play a crucial role in gene expression in mammalian cells. It is possible that what reported by GSEA on RNA-seq data is completely re-shuffled at protein level.

According to our results, we conclude that by adding miRNA-TS to the transfected genes, the global gene and protein dysregulation caused by burden is reduced. Moreover, the stronger is the target repression, the higher the stabilization of the intracellular machinery against transgenes expression. This is an overall conclusion that can be depicted from all the three projects described in this Chapter and can be explained by the synergy of multiple factors: the stabilization of gene expression is achieved thanks to miRNAs that form a natural iFFL motive with the co-expressed targets; the action of miRNAs on

their targets frees up cellular resource that are re-allocated for the production of other genes.

This study has a two-fold outcome: not only we are studying the effects on the systemic impact of gene expression burden on host cells, as shown in Section 3.4, we could use our RNA-seq data to depict information to engineer burden-resistant cell lines, similarly to what they did in [64]. To this regard we are exploring two approaches: *(i)* designing synthetic promoters responsive to upregulated transcription factors; *(ii)* integrate burden controllers downstream upregulated genes. In both cases we referred to RNA-seq results in noTS samples comparing High *vs* NT populations for the TF and genes' choice.

For the first approach, we were inspired by the work of Ceroni *et al.* They did RNA-seq analysis on burdened bacteria to select genes upregulated by burden. They took the promoters of the genes and cloned a feedback controller of transgene expression downstream [64]. Unfortunately, this approach is not feasible for mammalian promoters, as their length is of several kbp. Therefore, we decided to build synthetic promoters containing the minimal YB-tata promoter and 5 repetitions of the selected TF responsive elements. The responsive elements for the identified transcription factors are inferred from the Transcription Factor Target Gene database (http://tfbsdb.systemsbiology.net/). We selected 5 TF - namely ATF3, DDIT3, PPARG, MAFF and WT1 - and for each of them several predicted binding sites variants. As proof of concept we cloned a red fluorescent protein (mCherry) downstream the synthetic promoters. Ideally, the more plasmidic DNA we transfect, the higher the gene expression burden and the higher the mCherry activation should be (**Fig. 3.34a**).

For the second approach, we selected three of the most upregulated genes in High *vs* NT comparison in noTS transfection and designed gRNAs to integrate a controller downstream these genes. This approach is very similar to the one described by Origel Marmolejo *et al* to monitor the unfolded protein response [196]. Similarly to the previous approach, we are initially integrating a fluorescent reporter as shown in **Fig. 3.34b**. In both cases, we use fluorescent signal as proof of concept to easily visualize reporter activation in response to gene expression burden increase. In future, the fluorescent genes will be exchanged with a feedback controller of exogenous transgene expression or of

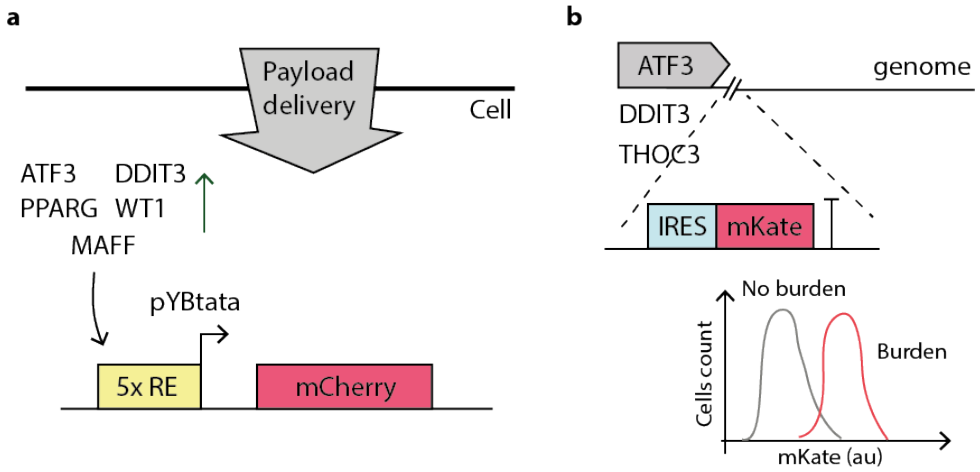


Figure 3.34: **Experimental design of burden controllers in mammalian cells. (a)** Synthetic promoters activated by TF highly expressed in burdened cells. RE: responsive element. **(b)** dCas9-driven integration of IRES-mKate reporter downstream upregulated-by-burden genes. Burdened cells should turn red upon genetic payloads delivery and burden induction.

endogenous genes that can help to maintain normal cellular physiology.

This project paves the way to the full understanding of the consequences of gene expression burden in mammalian cells and provides key information to engineer burden-resistant systems. Critical for these achievements is the power of the *omics* analyses applied to synthetic biology. It is not surprising that an increasing number of synthetic biology groups focuses on RNA-seq, proteomics, epigenomics, etc for their projects [197–201]. For an even deeper investigation of the effects of gene expression burden on host cells, in the future it will be interesting to run single cell RNA-seq (scRNA-seq), for fine-grained assessment of cell-to-cell variability in response to burden [202].

Chapter 4

miRNA-based cell classifiers

Cell classifiers are decision-making synthetic circuits that enable *in vivo* cell-type classification [74]. As such they share the basic architecture of the sensor-actuator devices: sensor modules that determine timing and localization of actuator modules activation.

Input to this class of synthetic circuits can be any molecular signature unique to a cell type or cluster [74, 203]. Due to their versatility and broad expression, the most common implementations use miRNAs as input to the classifier. miRNA profile constitutes a unique fingerprint not only of different cell types, but also of different cell states (e.g. T-cells in naive *vs* activated state [204]). Their robustness in discriminating among cellular types (e.g. cancer *vs* normal cells [76]) has made promising tools for biomedical applications.

In this chapter, I will present how miRNA expression data alone are not sufficient to identify good miRNA candidates to build a classifier. I will then present a workflow for the implementation of a classifier with enhanced robustness, that integrates information about miRNA abundance and activity. Our workflow will hopefully allow the implementation of a reliable tool to drive experimental design for discriminating different cancer cell-types.

4.1 Previous implementations of miRNA-based cell-type classifiers

For the first time in 2011, Xie *et al* showed that cell-type classifiers could be implemented in mammalian cells [76]. They found that by using a customizable set of endogenous microRNAs, they could trigger a cellular response only if the expression level of the miRNAs matched a predetermined profile of interest. Their classifier was able to trigger apoptosis in HeLa without affecting non-HeLa cells.

To discriminate among a broader set of cell types expanding their approach, more powerful computational tools were needed to identify both good input candidates to the circuits and the best circuit design to process the signal to activate cellular response. To this regard, throughout the years several works have focused on the implementation of robust algorithms that can predict classifier architectures [74, 75, 205].

In 2017, Mohammadi *et al* [75] implemented a work-flow in a two-step process. They first implemented a mechanistic model of the classifier to predict the output according to different input miRNA levels. Optimal parameters values for the model reactions were set in order to maximize the divergence between output on- *vs* off-state. In the second step, they used these parameters to identify the best fitting miRNAs according to miRNA expression from a data set. They validated their approach on predicted data sets showing perfect prediction of classifier design. However, as pointed out by Becker *et al*, it is unlikely that the perfect circuit design for maximal classification efficiency can be always identified when working with real data sets [205]. To this regard, Becker *et al* followed a multi-step workflow for classifier optimization, which allows to obtain globally perfect and imperfect (in case when a perfect classifier does not exist) classifiers in a short time. Their approach relied on binarized miRNA data sets using Answer Set Programming (ASP) [206]. Thanks to this approach they could compare between optimal classifier designs, identifying the key features.

In 2019 Nowicka *et al*, embraced a new approach to cell classifier design. They applied the concept of Distributed Classifiers (DC), consisting of simple single circuits that decide collectively according to a threshold function [74]. The central idea is to create a complex classifier from a population of weak or simple classifiers [207]. They

validated their computational method on real world data, demonstrating high prediction accuracy.

Despite the accuracy of the aforementioned approaches in running predictions on simulated or experimentally available data-sets, they lack the fundamental validation of the miRNAs activity in the cell types of interest. This information is essential to implement a tool that can efficiently predict the behavior of a miRNA-based cell classifier, and therefore, that can be used to efficiently guide the experimental design.

4.2 Building a miRNA-based cell classifier that accounts for miRNA activity

Unpublished results

As already demonstrated, the reported abundance of the miRNAs in the cells does not necessarily correlate with their activity [73]. The number of targets of a given miRNA in the cell type of interest [185], the compartmentalization of the cell and the spatial expression of the miRNA [189] and the expression level of a plethora of molecular components that assist miRNAs in exploiting their function [208] are also determinant for miRNA activity. To prove this point we show that relying on simple miRNA expression data sets is not enough to predict miRNAs activity and therefore their suitability to classify different cell-types. To address the discrepancy between miRNA expression level (both inferred by existing data-sets and experimentally validated) and actual miRNA activity, we adopted a workflow that iterates between computational implementation and experimental validation.

4.2.1 Cell lines and miRNAs identification

We decided to implement our classifier to discriminate cancer cell lines. The Cancer Cell Lines Encyclopedia (CCLE database) provides public access to genomic data, analysis and visualization for over 1100 cell lines [209, 210]. We identified 4 cells lines of interest, namely H1299 (non-small lung cancer cell line), U2OS (osteosarcoma cell line), U87 (glioblastoma cell line) and Jurkat (lymphoblast cell line) and compared their miRNome profiles. We calculated the z-score of each miRNA as relative to the mean and variance across the 4 cell lines. The z-score describes a value's relationship to the mean of a group of values and is measured in terms of standard deviations from the mean. We selected only miRNAs differentially expressed in one of the four cell lines, meaning that its z-score in that cell line is opposite to its z-scores in the others. We assigned a score to the shortlisted miRNAs, calculated as the sum of the z-scores in the three concordant cell lines times 3 (number of cell lines). Finally, we selected about 20 differentially expressed miRNAs according to the given scores (**Table 4.1**).

Almost every miRNA has a 5p and a 3p variant that derive respectively from

the forward and the reverse strands of the pre-miRNA processed by Dicer (refer to **Fig. 3.16**). We decided to focus on the 5p variant for all miRNAs if not otherwise specified by the CCLE data set (e.g. for miR-199a-3p + miR-199b-3p and miR-520a-5p) and with the exception of miR-31 that we considered both in its 3p and 5p variants. The complete list of miRNAs and targets is reported in Appendix D).

Table 4.1: **List of differentially expressed miRNAs analyzed in this study.**

miRNAs differentially expressed in H1299	**H1299**	**U2OS**	**U87**	**Jurkat**	**Score**
miR-31	3.01464	-0.93200	-2.80328	-2.71718	-19.35740
miR-633	5.74348	-1.98512	-1.60717	-0.20357	-11.38760
miR-1973	3.81380	-1.96193	-0.46028	-0.31701	-8.21765
miR-630	2.58093	-0.63248	-1.11832	-0.60111	-7.05575
miRNAs differentially expressed in U2OS	**U2OS**	**U87**	**H1299**	**Jurkat**	**Score**
miR-135b	3.33430	-1.97324	-3.00380	-4.64106	-28.85432
miR-199a-3p + miR-199b-3p	3.17213	-1.24269	-1.99179	-0.58852	-11.46900
miR-522	1.95118	-0.18338	-0.52614	-0.75051	-4.38007
miR-520a	-1.01443	0.06523	3.02168	0.23465	9.96469
miR-29b-1	-0.88754	1.94540	2.28935	1.08612	15.96206
miRNAs differentially expressed in U87	**U87**	**U2OS**	**H1299**	**Jurkat**	**Score**
miR-335	1.34976	-2.34466	-1.86470	-2.16395	-19.11990
miR-7b	1.12064	-1.33902	-2.52878	-0.16378	-12.09470
miR-376c	1.54611	-0.77266	-0.67851	-1.21338	-7.99365
miR-301a	-2.20475	0.80995	1.21512	2.18814	12.63964
miR-340	-2.52178	0.79209	0.99284	2.52838	12.93992
miR-106b	-1.20989	1.22178	1.89700	1.81658	14.80606
miRNAs differentially expressed in Jurkat	**Jurkat**	**U2OS**	**U87**	**H1299**	**Score**
miR-744	3.99584	-1.35947	-1.35947	-1.05939	-11.33500
miR-223	9.22275	-1.33696	-1.33696	-1.03688	-11.13240
miR-216a	-2.46261	0.46306	3.61504	2.47235	19.65134
miR-100	-5.55133	2.35974	5.24610	4.37382	35.93895

4.2.2 miRNAs quantification reveals discrepancies with CCLE data set

To confirm the abundance of the miRNAs reported by CCLE database for our cell lines, we performed mature miRNA quantification by qPCR. As we aimed at reliably comparing miRNA abundance across different cell lines, we tested the stability of two possible internal references: the Small Nucleolar RNA, C/D Box 72 (SNORD72) [211] and the Small Nucleolar RNA, C/D Box 95 (SNORD95) (**Fig. 4.1**). Except for one of the two replicates of H1299 cells, the Ct values for SNORD95 were relatively stable across the 4 cell lines, with fluctuations between 20 and 21.5 Ct. SNORD72 showed slightly higher variability and a lower expression as compared to SNORD95. For these reasons we decided to stick with SNORD95 as internal reference for our analyses (**Fig. 4.2**). In any case, the analysis according to SNORD72 normalization showed similar results (Appendix C).

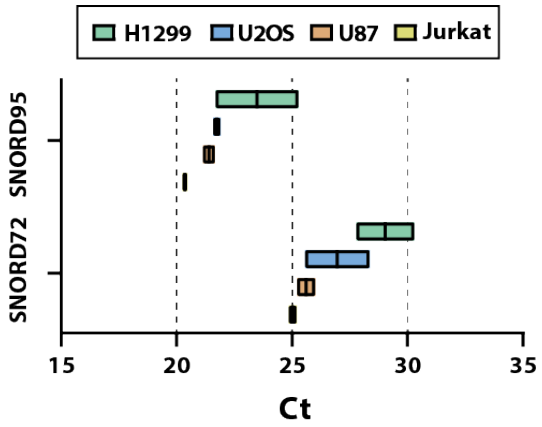


Figure 4.1: **Ct values of SNORD72 and SNORD95 across cell lines.** qPCR cycle threshold (Ct) values at which the internal references SNORD72 and SNORD95 were detected in each cell line for N = 2 biological replicates. Floating bars represent the mean Ct value and the distance between the two biological replicates.

Quantification of the miRNAs is calculated with the dCt method and 2^{-dCt} is plotted to allow comparison of relative expression among all of them. Looking at the y axis values, it possible to infer that some of the selected miRNAs are very poorly expressed, for example miR-633, miR-520a-5p and miR-376c-5p. The expression is so low that in some cases the miRNA is not detected at all in a cell line (miR-633 in H1299). Each subfigure reports the z-scores of each miRNA for every cell line plus the total miRNA score (top) and qPCR results for the corresponding miRNAs (bottom).

Fig. 4.2a shows two of the shortlisted miRNAs differentially expressed in H1299: miR-31-5p (left) and miR-633 (right). miR-31-5p nicely reflected the z-scores calculated

from the data set: high expression in H1299 and low level in the other cell lines, especially in U87 and Jurkat. On the contrary, miR-633, that should be highly expressed in H1299, was not even detected, while in the other cell lines it was, even if at very low levels.

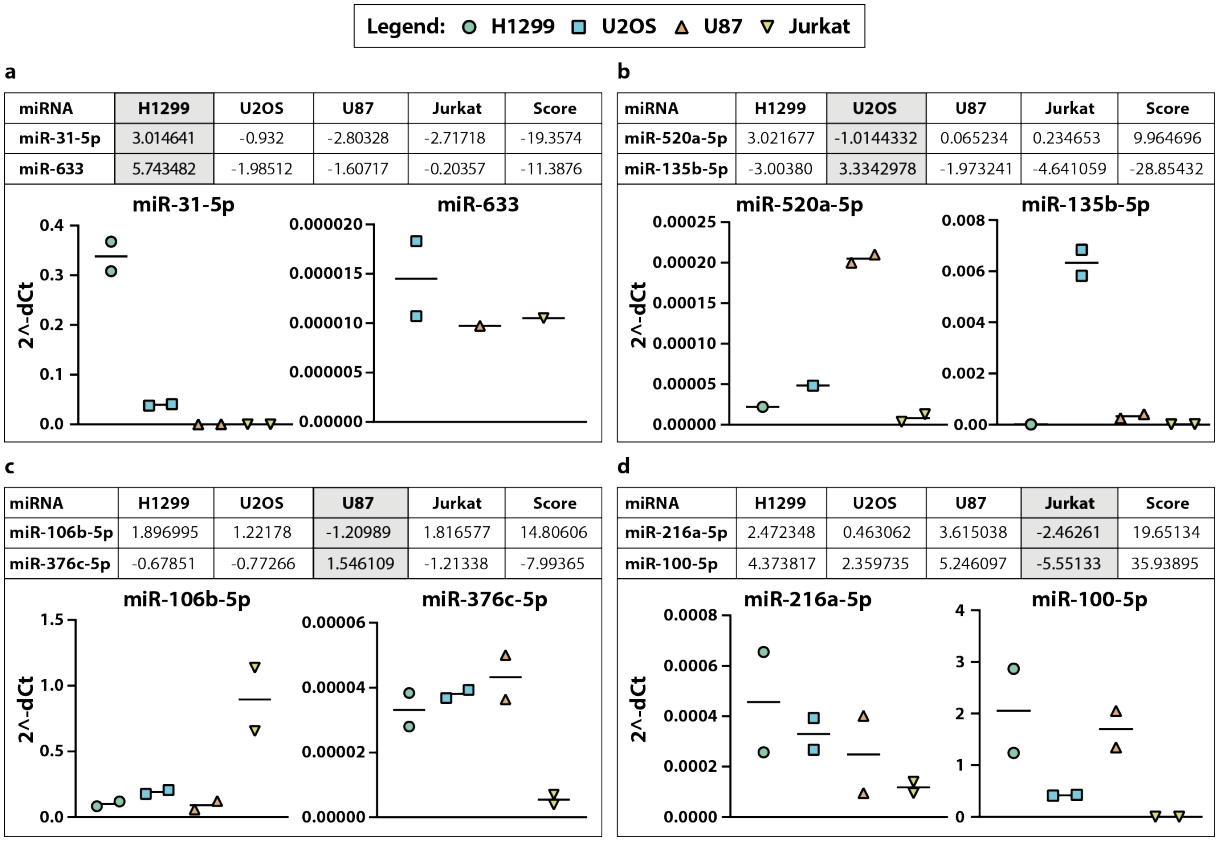
Fig. 4.2b shows two of the miRNAs differentially expressed in U2OS cells. miR-520-5p had an almost complete opposite expression pattern as compared to the calculated z-scores: U2OS was not the cell line in which the expression was the lowest and in H1299, that should have the highest amount, it was barely detected. On the contrary, miR-135b-5p expression in U2OS was higher than in the other cell lines and in H1299 it had the lowest expression reported, as much that only one of the two replicates was detected, according to what reported in the data set.

Fig. 4.2c reports two miRNAs differentially expressed in U87 cells. Here, miR-106b-5p level should be lower in U87, however experimental results showed that it had a comparable expression in H1299 and in U2OS cells. The level in H1299 should be comparable to the one in Jurkat cells, which were instead expressing it much more than the others. miR-376c-5p should be higher in U87 than in the other cell lines, which was confirmed experimentally, even if not as much as predicted by the z-scores. In Jurkat cells instead it was low as expected.

Finally, **Fig. 4.2d** displays two differentially expressed miRNAs in Jurkat cells. miR-216a-5p was indeed lower in Jurkat cells, however its level in U87 should be higher. Finally, miR-100-5p almost perfectly met the predictions, except for the level of H1299 that should be slightly lower than the one in U87.

To summarize, out of 8 miRNAs tested, only two of them satisfactorily validated the z-scores calculated from the CCLE data set: miR-31-5p and miR-135b-5p. Quantification of miR-376c-5p, miR-216a-5p and miR-100-5p confirmed that the cell line in which the miRNA is differentially expressed has the lowest/highest level of the miRNA, but the level in other cell lines was not confirmed. The remaining miRNAs were not validated.

These results, urged us to reconsider the reliability of the miRNoma data set as the only reference on which to identify candidate miRNAs. It is in fact well-known that cancer cell lines exhibit high degree of variability [212].



a

miRNA	H1299	U2OS	U87	Jurkat	Score
miR-31-5p	3.014641	-0.932	-2.80328	-2.71718	-19.3574
miR-633	5.743482	-1.98512	-1.60717	-0.20357	-11.3876

b

miRNA	H1299	U2OS	U87	Jurkat	Score
miR-520a-5p	3.021677	-1.0144332	0.065234	0.234653	9.964696
miR-135b-5p	-3.00380	3.3342978	-1.973241	-4.641059	-28.85432

c

miRNA	H1299	U2OS	U87	Jurkat	Score
miR-106b-5p	1.896995	1.22178	-1.20989	1.816577	14.80606
miR-376c-5p	-0.67851	-0.77266	1.546109	-1.21338	-7.99365

d

miRNA	H1299	U2OS	U87	Jurkat	Score
miR-216a-5p	2.472348	0.463062	3.615038	-2.46261	19.65134
miR-100-5p	4.373817	2.359735	5.246097	-5.55133	35.93895

Figure 4.2: **miRNA quantification using SNORD95 as internal reference.** Each subfigure shows the selected miRNAs and their z-scores with respect to the mean and the variance among the 4 cell lines 4.2.1. In grey is highlighted the cell line in which the displayed miRNAs are differentially expressed (top). Bottom right and left, the corresponding quantification by qPCR relative to the internal reference SNORD95, round green symbols correspond to H1299 data, blue squares to U2OS, upwards orange triangles to U87 and downwards yellow triangles to Jurkat. To compare with the quantification relative to SNORD72 see Appendix C. **(a,b,c,d)** selected miRNAs differentially expressed in H1299, U2OS, U87 and Jurkat respectively. Data represent the mean and single values for N = 2 biological replicates. Missing data points correspond to undetected values due to too little abundance of miRNA in the sample.

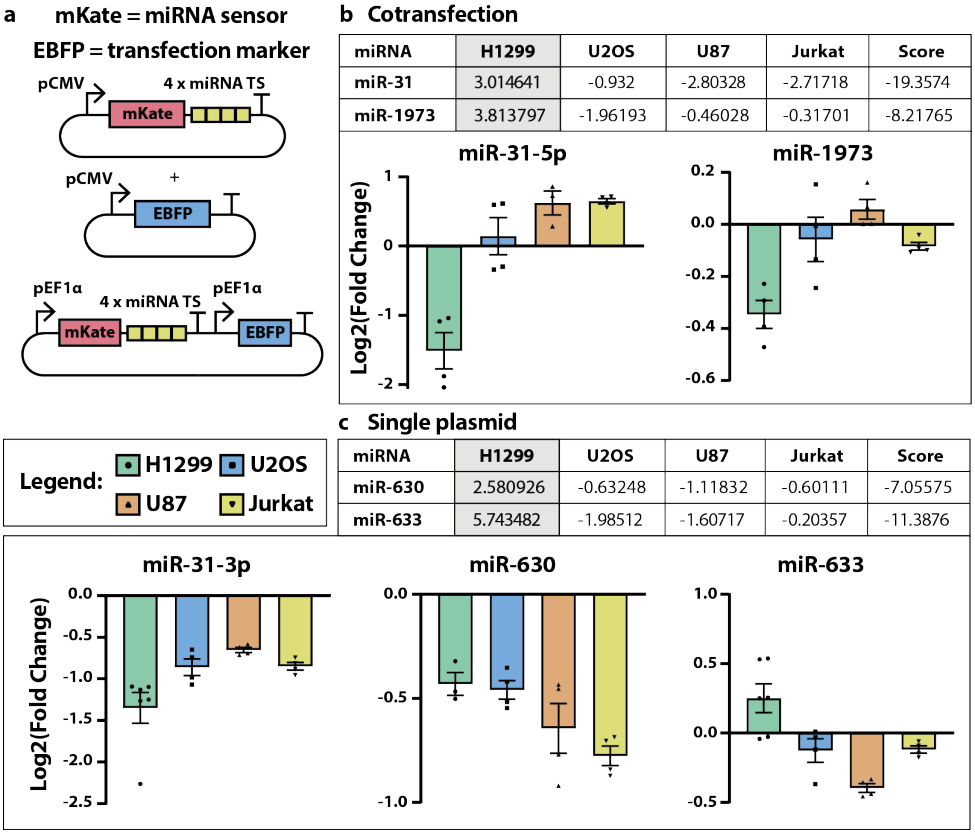
4.2.3 miRNAs activity poorly reflects miRNA abundance

Concomitantly to the qPCR analysis, we measured the activity of the selected miRNAs. To do so, we used (*miRNA sensor*) plasmids encoding for a fluorescent protein (mKate) with 4 fully complementary miRNA target sites at 3'UTR, co-expressed with a fluorescent transfection marker (EBFP). The plasmids design had two different variants: we either ran co-transfections of the *miRNA sensor* and the transfection marker under CMV promoter regulation encoded on two different plasmids (**Fig. 4.3a**, top), or we used single plasmids encoding both for the *miRNA sensor* and the transfection marker under the

EF1a promoter regulation (**Fig. 4.3a**, bottom). The latter were kindly provided by Professor Ron Weiss lab (MIT) and had been used by Dr Jeremy Gam for his project aided at modeling miRNA activity [73]. Plasmid design used to measure the activity of a given miRNA is specified in each figure. The control for each plasmid design was mKate without miRNA TS (noTS). For each cell line, we calculated miRNA activity as the fold change between normalized mKate fluorescence over EBFP of the *miRNA sensor* plasmid and mKate fluorescence normalized over EBFP of the control plasmid. Then we did the Log_2 transformation of the fold change (Log_2(FC)) and compared the results for a given miRNA across the cell lines in the same graph (**Fig. 4.3b-c, 4.4, 4.5, 4.6**). The higher the miRNA activity, the lower the expression of mKate, reflected by a negative Log_2(FC). If the miRNA did not show any activity, the bar resulted close to 0, which is the Log_2 of 1, corresponding to the normalized value of the noTS control.

Fig. 4.3b-c, shows the miRNA activity of the miRNAs differentially expressed in H1299 cells. miRNA-31-5p and miR-1973 activity was detected by co-transfection (**Fig. 4.3b**). Their activity reflected miRNA expression, however miR-1973 lead to a Log_2(FC) ≤ 0.5 of mKate expression. Such a small change is usually not considered a significant biological change (see Section 3.4). The TS for miRNA-31-3p, miR-630 and miR-633 TS were encoded on single plasmids (**Fig. 4.3c**). For miRNA-31-3p z-scores refer to the top part of **Fig. 4.3b**; its activity did not represent well the analysis from the data set, conversely to its 5p variant (**Fig. 4.3b**). miR-630 had almost the exact opposite activity as compared to the z-score predictions, same for miR-633. However, as shown in **Fig. 4.2a**, miR-633 level was extremely low in all cell lines and in H1299 it was not even detected.

Fig. 4.4 reports the miRNA activity of the miRNAs differentially expressed in U2OS cells. In **Fig. 4.4a** is displayed the activity of the combination of miR-199a-3p + miR-199b-3p and the activity of miR-520a-5p. The combination of miR-199a-3p + miR-199b-3p indeed showed that they have higher activity in U2OS as compared to the other cell lines, however the impact on mKate expression was little and very similar to the one in H1299. miR-520a-5p activity did not agree with the z-scores. In **Fig. 4.4b** is reported the activity of miRNAs measured by single plasmid transfection. miR-29b-1-5p had opposite activity as compared to the predictions and also miR-522-5p activity poorly



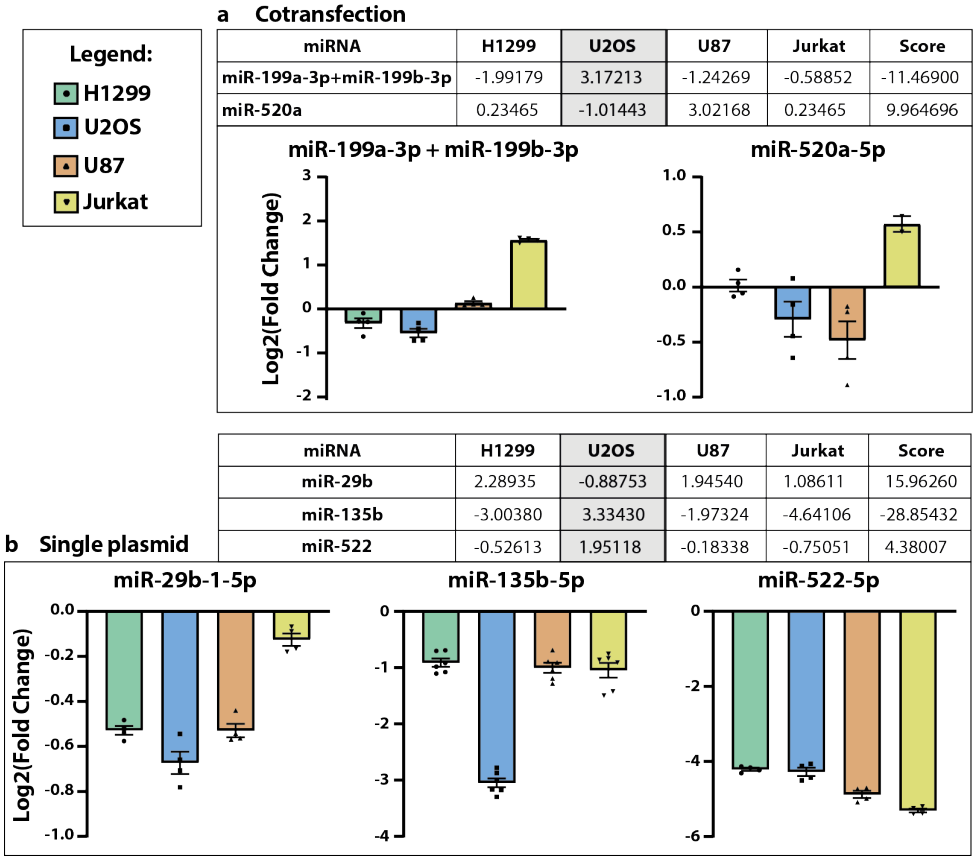
miRNA	H1299	U2OS	U87	Jurkat	Score
miR-31	3.014641	-0.932	-2.80328	-2.71718	-19.3574
miR-1973	3.813797	-1.96193	-0.46028	-0.31701	-8.21765

miRNA	H1299	U2OS	U87	Jurkat	Score
miR-630	2.580926	-0.63248	-1.11832	-0.60111	-7.05575
miR-633	5.743482	-1.98512	-1.60717	-0.20357	-11.3876

Figure 4.3: **Plasmids design and activity of miRNAs differentially expressed in H1299 cells.** **(a)** Plasmids design for the *miRNA sensor* and transfection marker used to measure miRNAs activity. miRNA sensor is a red fluorescent protein mKate with 4x fully complementary miRNA TS; transfection marker is a blue fluorescent protein (EBFP). They were encoded either on two separate plasmids (top), or on one single plasmid (bottom). **(b,c)** Flow cytometry results of co-transfections and single plasmid transfections respectively. Top, z-scores of the miRNAs analyzed; bottom, Log_2(FC) of mKate expression. Data represent the mean +/- SE. SE: standard error. N = 2-6 as indicated by the number of symbols for each bar.

correlated with the z-scores. Instead, miR-135b-5p activity nicely depicted its higher level in U2OS, however the activity in the other 3 cell lines did not reflect the different levels it should have.

In **Fig. 4.5**, is shown the miRNA activity of the miRNAs differentially expressed in U87 cells. miRNA-106b-5p activity was measured by co-transfection (**Fig. 4.5a**). Its activity reflected the z-score values. The activity of the other 5 miRNAs was detected via single plasmid transfection (**Fig. 4.5b**). miR-7b-5p, that should be highly expressed in U87, had instead the lowest activity in that cell line. miR-301-5p only slightly reflected the predictions, especially the high experimental variability in U87 made it difficult to con-



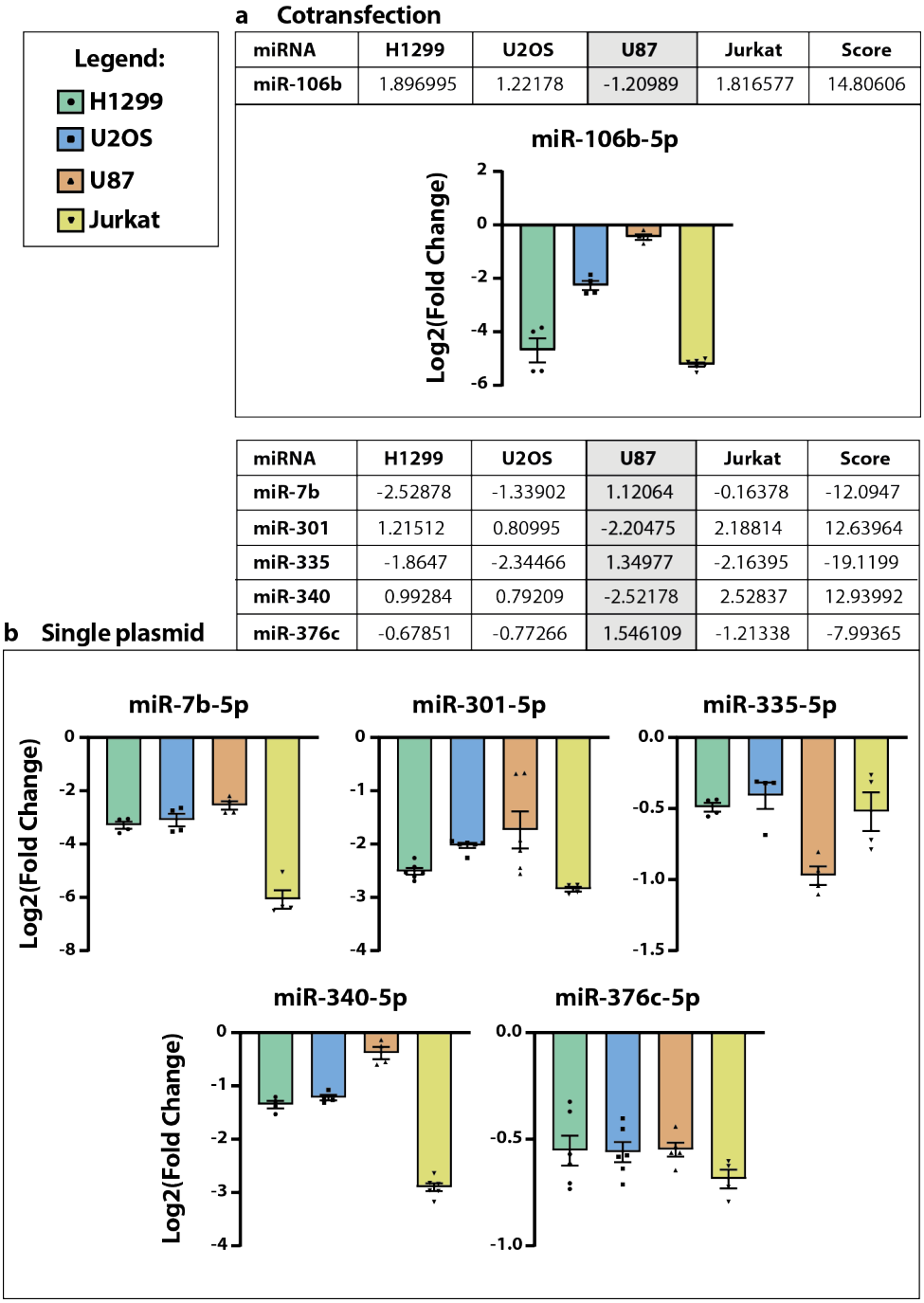
a Cotransfection					
miRNA	H1299	U2OS	U87	Jurkat	Score
miR-199a-3p+miR-199b-3p	-1.99179	3.17213	-1.24269	-0.58852	-11.46900
miR-520a	0.23465	-1.01443	3.02168	0.23465	9.964696

b Single plasmid					
miRNA	H1299	U2OS	U87	Jurkat	Score
miR-29b	2.28935	-0.88753	1.94540	1.08611	15.96260
miR-135b	-3.00380	3.33430	-1.97324	-4.64106	-28.85432
miR-522	-0.52613	1.95118	-0.18338	-0.75051	4.38007

Figure 4.4: **Activity of miRNAs differentially expressed in U2OS cells. (b,c)** Flow cytometry results of co-transfections and single plasmid transfections respectively. Top, z-scores of the miRNAs analyzed; bottom, Log_2(FC) of mKate expression. Data represent the mean +/- SE. SE: standard error. N = 2-6 as indicated by the number of symbols for each bar.

firm. miR-335-5p effectively showed enhanced activity in U87 and similarly miR-340-5p reflected the expression level with its activity, causing the highest mKate downregulation in Jurkat and the lowest in U87. Finally, miR-376c-5p showed no correlation between miRNA expression and activity and by looking at **Fig. 4.2c** this could be explained with its extremely low level in our cell lines of interest.

Fig. 4.6 shows the miRNA activity of the miRNAs differentially expressed in Jurkat cells. The activity of miR-216a-5p was measured by co-transfection (**Fig. 4.6a**) and it reflected the z-score predictions, despite its quantification reported in **Fig. 4.2c** suggested otherwise. Also in this case, the majority of the miRNA activities were measured by the single plasmid transfection. Both miR-100-5p and miR-223-5p behaved accordingly to z-scores. Instead, miR-744-5p did not, exhibiting very low impact on mKate expression.

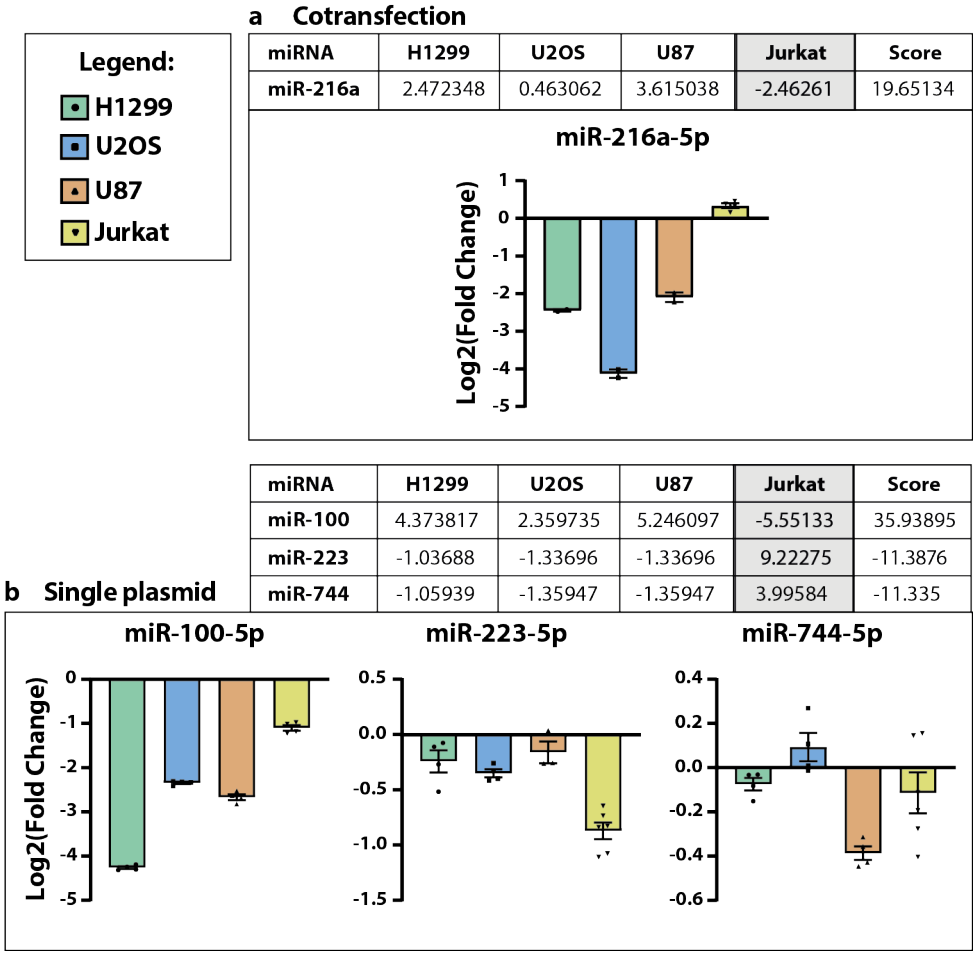


miRNA	H1299	U2OS	U87	Jurkat	Score
miR-106b	1.896995	1.22178	-1.20989	1.816577	14.80606

miRNA	H1299	U2OS	U87	Jurkat	Score
miR-7b	-2.52878	-1.33902	1.12064	-0.16378	-12.0947
miR-301	1.21512	0.80995	-2.20475	2.18814	12.63964
miR-335	-1.8647	-2.34466	1.34977	-2.16395	-19.1199
miR-340	0.99284	0.79209	-2.52178	2.52837	12.93992
miR-376c	-0.67851	-0.77266	1.546109	-1.21338	-7.99365

Figure 4.5: **Activity of miRNAs differentially expressed in U87 cells. (b,c)** Flow cytometry results of co-transfections and single plasmid transfections respectively. Top, z-scores of the miRNAs analyzed; bottom, Log_2(FC) of mKate expression. Data represent the mean +/- SE. SE: standard error. N = 2-6 as indicated by the number of symbols for each bar.

Overall, the miRNAs that qualitatively reflected the z-scores with their activity were 13 out of 20, but the ones that quantitatively showed a close correlation and that caused a significant variation in mKate expression (Log_2 < -0.5) were 8 (miR-31-5p,



miRNA	H1299	U2OS	U87	Jurkat	Score
miR-216a	2.472348	0.463062	3.615038	-2.46261	19.65134

miRNA	H1299	U2OS	U87	Jurkat	Score
miR-100	4.373817	2.359735	5.246097	-5.55133	35.93895
miR-223	-1.03688	-1.33696	-1.33696	9.22275	-11.3876
miR-744	-1.05939	-1.35947	-1.35947	3.99584	-11.335

Figure 4.6: **Activity of miRNAs differentially expressed in Jurkat cells. (b,c)** Flow cytometry results of co-transfections and single plasmid transfections respectively. Top, z-scores of the miRNAs analyzed; bottom, Log_2(FC) of mKate expression. Data represent the mean +/- SE. SE: standard error. N = 2-6 as indicated by the number of symbols for each bar.

miR-135b-5p, miR-106b-5p, miR-335-5p, miR-3405p, miR-216-5p, miR-100-5p and miR-223-5p). Moreover, the correlation with measured miRNA expression in Section 4.2.2 was poor, being confirmed only for miR-31-5p, miR-135b-5p and miR-100-5p.

The poor reliability of the simple calculation of the miRNAs z-scores based on miRNome data, urged us to find a more robust predictive method that would take into account both miRNA abundance and miRNA activity.

4.2.4 A workflow for miRNAs scoring according to expression level and activity

We reasoned to implement a computational approach based on both miRNA expression and activity. We defined miRNA activity as the degree of downregulation a miRNA causes to its targets. The approach is developed in three steps: *(i)* selection of miRNAs upon expression level in the cell lines of interest; *(ii)* calculation of selected miRNAs activity through the expression pattern of their endogenous targets; *(iii)* identification of the final candidates by integrating the output of Step 1 and Step 2 (**Fig. 4.2**).

The first step takes as input the miRNAs expression levels in the cell lines of interest from the CCLE data base. The miRNA amount is measured with NanoString

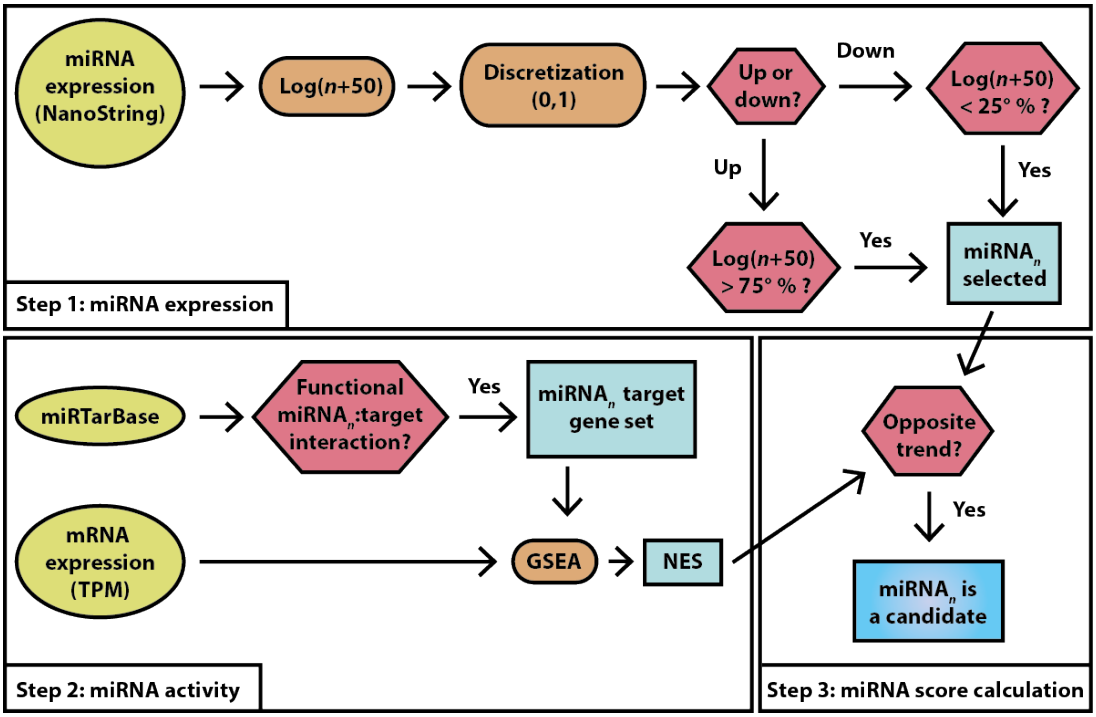


Figure 4.7: **Workflow for the selection of candidate miRNAs.** Step 1: identification of miRNAs differentially expressed in a cell line of interest and belonging to the first ($<$ 25 %) or fourth ($>$ 75 %) quartile of miRNAs expression levels in the given cell line. Step 2: Identification of the validated targets of the miRNAs selected Step 1 via miRTarBase. These targets form the miRNA$_n$ target data set. Gene expression in the cell lines of interest is ranked according to the z-score of each transcript among the cell lines. The ranked list of genes is analyzed with GSEA to assign a normalized enrichment score (NES) to the miRNA$_n$ targets data set as representation of miRNA$_n$ activity. Step 3: if the miRNA expression and NES have opposite trend (e.g. miRNA upregulated and negative NES or *viceversa*), the miRNA is selected as a candidate. TPM: transcripts per million.

technology and is provided as normalized data. The logarithm of the normalized value + 50 is calculated for each miRNA ($miRNA_n$) in the cell lines of interest, to reduce the noise ratio, as described in [210]. Then, miRNA expression level in each cell line is discretized, using the clustering approach developed in [213]. The aim is to reduce the miRNA expression to a binary value (0,1). Only miRNAs with [0,1,1,1] or [1,0,0,0] values pass to the following step. If their binary value is 0 in the cell line in which they are differentially expressed (downregulated), they pass the selection only if they belong to the first quartile of global miRNA expression levels in the given cell line; on the opposite, if the miRNA binary value is 1 in cell line in which it is differentially expressed (upregulated), it is selected only if it belongs to the highest quartile. This filtering step makes sure that the differentially expressed miRNAs selected are effectively poorly or highly present coherently with their downregulation or upregulation in the cell line that exhibits differential expression.

In Step 2, the targets of the selected miRNAs are derived by the miRTarBase database (https://bio.tools/mirtarbase). For each $miRNA_n$ is created a $miRNA_n$ target gene set. The CCLE RNA-seq gene expression data of the cell lines of interest are compared and the z-score for each gene in all cell lines is calculated. Genes are then ranked by z-score. GSEA is performed to test if the $miRNA_n$ target gene set is enriched in the cell line in which $miRNA_n$ is differentially expressed. The result of this step is the NES value of the $miRNA_n$ target gene set. The NES value is a measure of the $miRNA_n$ activity in the considered cell line: if it is positive, the activity is low; if it is negative the activity is high.

In Step 3 the NES value is compared to the expression level of the corresponding miRNA: if they are inversely correlated the miRNA is selected as candidate for the classifier. The inverse correlation between NES and miRNA expression means that if a miRNA is highly expressed the targets should be downregulated compared to the other cell lines (negative NES) and *viceversa*.

We begun the validation of our workflow with the 4 cell lines we previously selected in Section 4.2.1. First we identified the miRNA candidates to discriminate each cell line and we specified whether they are up- or down-modulated in the cell line in which they are differentially expressed (**Table 4.2**).

Table 4.2: **Number of candidate miRNAs at each step according to our workflow.**

Cell Line	miRNAs Step 1	miRNAs Step 3	Up	Down
H1299	61	17	17	0
U20S	40	18	1	17
U87	35	1	1	0
Jurkat	78	2	2	0

We took the miRNAs from the first Step of the workflow and checked against our qPCR quantification results (**Fig. 4.2**). 6 out of 8 miRNAs quantified in Section 4.2.2 were selected from Step 1, miR-216a and miR-106b being the excluded ones. Of the remaining 6, 4 were validated, namely miR-100, miR-31, miR-135b and miR-376c (**Table 4.3**).

Table 4.3: **Validation of Step 1 results with qPCR data from Fig. 4.2.**

miRNAs Step 1	H1299	U2OS	U87	Jurkat	Validated
miR-100				DOWN	Yes
miR-31	UP				Yes
miR-520a-5p	UP				No
miR-633	UP				No
miR-135b		UP			Yes
miR-376c			UP		Yes

Then, we validated the output of the third step according to the miRNAs activity measured in Section 4.2.3. Only 6 out of 20 shortlisted miRNAs in Section 4.2.1 are selected in Step 3 (**Table 4.4**). Of these 6, only 2 showed the expected behavior - miR-31-5p and miR-301a-5p. Of these 2, only miR-31-5p validated experimentally both step 1 and step 3.

Table 4.4: **Validation of Step 3 results with flow cytometry data from Fig. 4.3-4.6.**

miRNAs Step 3	Cell line	miRNA z-score	Targets NES	Validated
miR-31	H1299	1.46	-1.97	Yes
miR-106b	H1299	0.66	-2.11	No
miR-216a	H1299	0.52	-1.54	No
miR-301a	H1299	0.35	-1.14	Yes
miR-29b	U2OS	-1.39	1.17	No
miR-100	U87	0.83	-1.17	No

Clearly our tool still requires consistent implementation to increase prediction ac-

curacy, however it is a promising mean to find robust candidate miRNAs. In fact, we showed that the limiting factor in making reliable predictions is the discrepancy between the reported miRNAs expression in the data sets and their activity. The lack of a proper index to evaluate miRNA activity is the critical step to address to build a reliable and robust predictive tool.

4.3 Future steps to implement a highly reliable tool for classifiers design

Our preliminary results showed that relying on either reported or directly measured miRNA expression level is not enough to predict the ability of a miRNA to discriminate a cell line against another. Critical in the identification of a good miRNA input is the ability to predict its activity more than its abundance in the cellular context. Since, it is not feasible to measure the activity of all the miRNAs in all cell lines every time one needs to implement a classifier, a mean to integrate *omics* data from multiple data-sets is needed.

Here we propose a workflow for the identification of miRNA candidates that accounts both for miRNA expression level and for miRNA activity. miRNA activity is defined as the impact of the miRNAs on their endogenous targets. Our method assumes that expression level of a miRNA and its endogenous targets are negatively correlated.

In the first steps of the tool optimization, we will validate candidate miRNAs activity experimentally. Once we will have optimized the key parameters to generate the miRNA_n target data-set and to rank the genes according to z-scores calculated from RNA-seq data comparison across cell lines, the tool should be able to make predictions with no further need of experimental validations.

Once we will have a set of good miRNA candidates validated individually to discriminate our cell lines of interest among each other, we will implement the actual classifier, exploring different logic gates design to combine miRNAs input to activate cellular response. As described by Mohammadi *et al*, the multiplexing of different logic gates enhances the classifier robustness, as it allows for the detection of multiple inputs, so that cell-to-cell variability does not impair the classification capacity of the circuit.

Cell-type classifiers hold the power to take decisions upon input sensing, showing tremendous potential for biomedical and clinical applications. In 2017, Prochazka *et al* demonstrated that cell classifier can be used to differentiate human pluripotent stem cells (hPSC) with a forward engineering approach, complementing the traditional reverse engineering method [214]. As growing interest is arising for synthetic morphogenesis [215], we envision that cell classifiers will play a crucial role in organoids and synthetic tissues

differentiation and cultures. Furthermore, we believe that the ability to trigger a cellular response only when the input pattern reflects the desired profile, will boost the precision of cell-based immunotherapies that require well localized activation.

Chapter 5

Promises and future challenges of Synthetic Biology

As stated by Professor Guy-Bart Stan at the World Economic Forum Summer Meeting in 2015, synthetic biology is one of the biggest engineering challenges of our times. It holds the potential to offer solutions to many of our present and future needs in medicine, food and energy production [216, 217]. However, the technical challenges of engineering living systems hamper the possibility to fully exploit its potential. In this thesis, I demonstrated the importance of implementing reliable means to control transgene expression and to increase our understanding of the cellular context of host organisms. To this end, I presented our efforts to: *(i)* build a new class of regulatory devices for post-transcriptional and translational control; *(ii)* elucidate the impact of transgene expression on intracellular resource sharing and on cellular physiology due to the so-called gene expression burden; *(iii)* implement a robust predictive tool to build cell-type classifiers starting from gene expression data-sets. The way to achieve perfect control of living systems is still long and encumbered by ethical concerns that should not be ignored. I conclude the thesis with a brief overview on some of the latest groundbreaking applications of synthetic biology tools and on the open technical and ethical challenges that hold back their potential.

5.1 Where are we now?

Synthetic biology has demonstrated tremendous potential to revolutionize both the way we answer unknown biological question and the way we address existing needs in medicine [216]. Rapid progress in tools and circuits development has given birth to new far reaching fields that were unthinkable just a decade ago [218, 219].

The ability to predict circuits behavior and safely deliver genetic parts has boosted the clinical applications [5]. Synthetic immunology applies synthetic biology tools and approaches to reprogram and enhance the function of immune cells, rewiring the overall capabilities of our immune system [220]. It has led to the generation of chimeric antigen receptor T cells (CAR-T) cells that already made it to the clinic, such as the recently approved CAR-T anti-cancer therapies Kymriah (tisagenlecleucel), that Japan has made available under their national health insurance system, and Yescarta (axicabtagene ciloleucel), that the National Health Service (NHS) in the UK has made available to patients [221].

Other therapeutic opportunities are offered by: theranostic cell lines that can sense a disease state and produce an appropriate therapeutic response [222]; blood replacement by engineering hematopoietic stem cells (HCSs) to correct inherited blood disorders [223]; oncolytic DNA viruses to deliver transgenes to the host cell allowing constant expression until the episomal vector is eventually diluted out through cell division, enhancing safety of the treatment [224]; and therapeutic bacteria [225].

Cybergenetics is another emerging field that lays at the interface of control engineering and synthetic biology, developing experimental tools for the computer control of cellular processes at the gene level in real-time [226, 227]. Cybergenetic control can be achieved by interfacing living cells with a digital computer that orchestrates genetic circuits using light (optogenetics) or chemicals [70, 228]. Such systems offer powerful means to both elucidate the dynamics of cellular processes by real-time interference and monitoring and to help to maintain cellular homeostasis by feedback control upon the detection of dysregulation [218, 226]. For example, they could trigger the release of insulin when blood glucose levels rise as detected by a wireless diagnostic tool [229].

5.2 Open challenges

Despite the promises of these cutting-edge applications, several obstacles need to be overcome in order to further apply synthetic biology to biomedicine. As also emerged in this thesis, high variability originates from the use of transient transfection, which is the dominant approach for prototyping genetic circuits [67, 70]. To this end, several studies focused on the implementation of regulatory motifs to achieve system's robustness to plasmid take-up and sample-to-sample variability, by optimizing the amount of cellular resources required to realize their function [59, 60, 67, 69, 70, 72]. Still, circuit's behaviour in transient transfection could largely differ from their behavior in stable integration, due to differences in circuit's copy number and subcellular localization. Generating stable mammalian cell lines is painstakingly slow, however in the last years faster integration methods with highly standardized procedure started to emerge. For example, landing-pad cell lines are immortalized cells edited using CRISPR/Cas9 to express a recombination cassette in a specific genome locus [230]. This method enables the locus-specific integration of large genetic payloads by using a BxB1 recombinase and specific sequences flanking the donor DNA. Unfortunately, immortalized cell lines are a limited model for several applications, as their metabolism is often different from primary cells in our body. For example an accurate model for CD8+ T cells is missing, thus immunotherapies optimization *in vitro* requires stable integration of primary cells, which is time-consuming and requires higher facility biosafety levels for lentiviral particles production and cells transduction. Thus, high-throughput screening of biomedically relevant circuits still suffers of technical limitations.

Additionally, we need to expand the range of molecules that can be recognized by biosensors as inputs and we need better understanding of the cellular factors that regulate gene expression in space and time to engineer better activator systems [218]. The implementation of synthetic circuits with biomedical relevance requires reduced toxicity, minimal basal activity and cross talk with neighboring endogenous pathways to allow better performance and scalability of the system [231]. To this regard our work on the implementation of post-transcriptional and translational orthogonal regulatory devices and on the characterization of gene expression burden is timely. Lastly, optimization of

efficient delivery strategies is required for both *ex vivo* engineered cells and of large-size payloads directly delivered *in vivo* [2]. We envision that studies in the next years will continue to focus on these needs to enhance the safety and the applicability of synthetic biology tools to medical and pharmaceutical applications.

Finally, another open challenge is the perception of synthetic biology from the community. Synthetic biology is an example of a dual-use technology: it promises numerous beneficial applications, but it can also cause harm [232]. For example, our ability to engineer viruses is improving the delivery of gene therapies, at the same time it generates concern for the creation of even more deadly pathogens. As Richard Dawkins said: "DNA neither cares nor knows, DNA just is", and scientists have the responsibility to consider whether the research planned could be misused. The synthetic biology community needs to embrace this challenge by engaging in open dialogue with regulatory bodies and the media [232].

Chapter 6

Materials and methods

In this chapter are listed all the protocols used for the experimental work and the computational analysis described in this thesis.

6.1 Vector cloning

6.1.1 PCR and gel electrophoresis

PCR reactions were performed using AccuPrime™ Pfx SuperMix (ThermoFisher Scientific) according to manufacturer's instructions. For the amplification of backbone or long fragments ($\geq$ 4.0 kbp) Phusion High Fidelity Polymerase with HF buffer (NEB) according to manufacturer's instructions was used instead. Primers annealing temperatures were determined using the NEB Tm Calculator v 1.9.7.

All DNA gels were 0.8% agarose, stained with SYBR Safe (ThermoFisher Scientific), loaded with Orange Loading Dye 10X (ThermoFisher Scientific). The electrophoresis run was performed at 80 V for 45 minutes in TAE Buffer with a Life 1 kb Plus DNA Ladder (ThermoFisher Scientific).

Gel extractions (Clontech Gel Extraction Kit) and PCR purification (Clontech DNA purification kit) were performed according to the manufacturer's instructions.

6.1.2 Plasmids assembly

Restriction digests were carried out at 37°C for 2 hours, except for the ones with DpnI (NEB) that were kept for 1 hour.

All ligations used T4 Ligase (NEB) according to manufacturer's instruction for 1 hour on bench with the DNA ratio of backbone to insert at 1:4. For insertion of miRNA-TS into pL-A1 vector, phosphorylation of oligos was performed using T4 Polynucleotide Kinase (NEB). Reactions were set up (3 µL 100 mM Oligo for a total amount of 300 pmol, 1 µL T4 PNK, 2 µL 10x Ligase Buffer, 14 µL ddH_2O) and incubated at 37 °C for 1h. For oligo annealing, in a new tube 10 µL of each complementary oligo was mixed together and incubated for 4 minutes at 95°C in a dry heating block. After the incubation step, they were annealed by slow cooling turning off the heating block. For ligation, they were finally diluted 1:150 to get a concentration 100 fmol/µL. The digested and purified vectors were dephosphorylated to avoid self-ligation. Reactions were set up (20 µL digested vector, 1 µL Antarctic Phosphatase, 2.2 µL Antarctic Phosphatase Buffer) and were incubated at 37°C for 20 minutes and then inactivated for 2 minutes at 80°C in a dry heating block.

Ligations used 50 ng of vector and 0.8 µL of insert (80 fmol).

In-Fusion® recombinase (In-Fusion® HD Cloning Kit Clontech) was used according to manufacturer instructions, to insert gene blocks (ordered via IDT) or fragments of at least 100 bp that could be amplified by PCR from other other plasmids. For the Infusion protocol a ratio of backbone to insert of 1:2 was used.

6.1.3 Bacterial transformation and culture

Bacterial transformations used Stellar Competent Cells (Clontech). An aliquot of cells was thawed on ice and 25 µL of cells were mixed with roughly 5 µL of each cloned plasmid (obtained by ligation or Infusion as described in 6.1.2) and then were left for 15 minutes on ice. Heat-shock was performed by incubation for exactly 30 seconds at 42°C in a dry heat block and placed for 2 minutes on ice. All transformations underwent an outgrowth step in SOC Medium (Clontech) with incubation in a shaking incubator (37°C, 200 rpm) for 30 minutes. Cells were then pelleted with a short spin centrifugation and 170 µL of medium was discarded. Cells were re-suspended in the remaining SOC medium and plated onto warm LB agar plates with antibiotic resistance 100 $mg*mL^{-1}$.

Inoculated cultures were incubated at 37°C shaking at 200 rpm in LB broth with the right antibiotic resistance at the above concentration and left overnight.

Plasmid purification (VWR Miniprep kit, Qiagen Midiprep Plus) were done according to the manufacturer's instructions. All plasmids were verified by Sanger sequencing.

6.2 *In vivo* characterization

6.2.1 Cell culture

HEK293T, U2OS, HeLa and U87 cells (ATCC) used in this study were maintained in Dulbecco's modified Eagle medium (DMEM, Gibco); H1299 and Jurkat cells were maintained in Roswell Park Memorial Institute medium (RPMI, Gibco); CHO-K1 were maintained in Minimum Essential Medium α(α-MEM, Gibco). All media were supplemented with 10 % FBS (Atlanta BIO), 1 % penicillin/streptomycin/L-Glutamine (Sigma-Aldrich) and 1 % non-essential amino acids (HyClone). The cells were maintained at 37°C and 5 % CO2.

6.2.2 Transfection

Transfections were carried out in 24-well plate for flow cytometry analysis or in a 12-well plate format for flow cytometry and qPCR analysis run on the same biological replicates (see experiments in Section 3.2.1). Transfections for RNA-seq and proteomics analyses (see Section 3.4) were carried out in 6 cm dishes. H1299, HeLa, U2OS, HEK293T, U87 and CHO-K1 cells were transfected with Lipofectamine® 3000 (ThermoFisher Scientific) according to manufacturer's instructions and 300-500 ng total DNA in 24-well plates. miR-31 inhibitor (Invitrogen™ mirVana™ miRNA Inhibitors) was cotransfected using the same method as for DNA (**Fig. 3.10c**).

All transfection were performed on cells in suspension immediately after plating at approximately 130000-200000 cells per well in 24-well plated. Cells were analyzed 48 hours post-transfection.

DNA, transfection reagents and seeded cells were scaled up or down according to the Lipofectamine® 3000 manufacturer's instructions for the desired plate format.

6.2.3 Electroporation

Jurkat cells were electroporated using the NEON electroporation system (Invitrogen). Approximately 300000 cells were electroporated for a single well in 24-well plates, starting from cells that did not exceed 700000 cells*mL^{-1} of confluence. 4 µg of plasmid DNA was electroporated per well in 24-well plates. First, cells were resuspended in R electroporation buffer (Invitrogen) and aliquoted in tubes on ice containing the DNA mixes. Then, the 10 µL electroporation kit (Invitrogen) was used, making sure that the DNA was concentrated enough to not exceed the 10% of the volume in the shock chamber. Electroporated cells were immediately transferred to wells with pre-warmed and pre-equilibrated media without antibiotics.

DNA, electroporation reagents and cells were scaled up or down for the desired plate as described in 6.2.2.

6.2.4 Flow cytometry and data analysis

Cells were analyzed 48 hours post transfection with a BD Facsaria™ cell analyzer (BD Biosciences) or BD Celesta™ cell analyzer (BD Biosciences) using 405, 488 and 561 lasers. For each sample > 20000 events were collected and fluorescence data were acquired with the following cytometer settings: 405 nm laser and 450/450 nm bandpass filter, 488 nm laser and 530/30 nm bandpass filter for EGFP, 561 nm laser and 610/20 nm filter for mKate. Cells transfected in 12-well plates were washed with DPBS, detached with 100 µL of Trypsin-EDTA (0.25 %) and re-suspended in 600 µL of DPBS (Thermo Fisher). 200 µL of cell suspension were used for flow cytometry and 400 µL for RNA extraction.

Fluorescence intensity in arbitrary units (au) was used as a measure of protein expression. For each experiment a compensation matrix was created using unstained (wild type cells), and single-color controls (EBFP only, mKate/mCherry only, EGFP/EYFP only). Live cell population and single cells were selected according to FCS/SSC parameters (**6.1**). Data analysis was performed with Cytoflow. Data fitting was performed using Mathematica's NonlinearModelFit function and the InteriorPoint method.

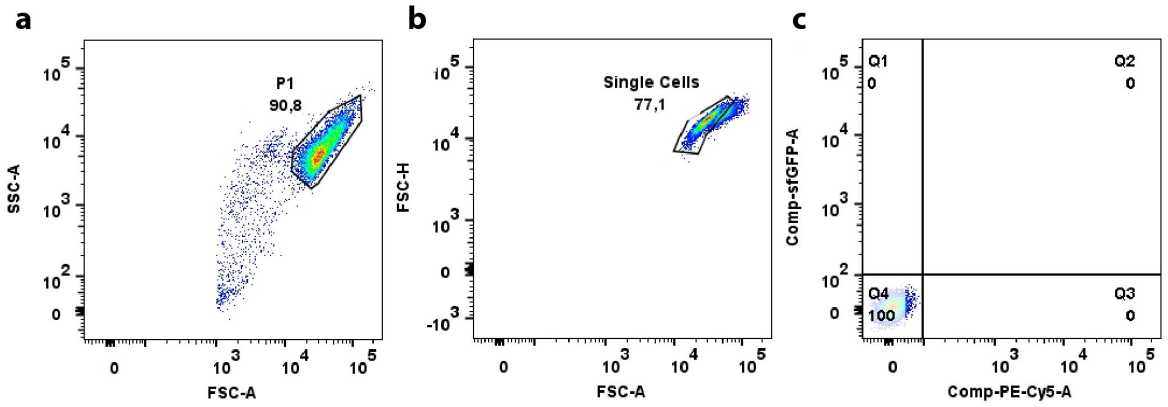


Figure 6.1: **Flow cytometry gating strategy. (a)** The recorded events were gated in the FSC-A *vs* SSC-A channels to select the living cells population (P1). **(b)** The P1 was then gated in the FSC-A vs FSC-H channels to select the single cell population. **(c)** For each experiment a sample of non-transfected cells was used to set the positive threshold for each fluorescence. Cells selected following this pipeline were then analyzed.

6.2.5 Cell sorting

H1299 cells used for the experiment shown in **Fig. 3.4c,3.15** and for RNA-seq and proteomics analysis as described in Section 3.4, were trypsinized from 6 cm dishes and counted. They were then centrifuged at 500 g for 5 min and resuspended at a concentration of 5 mln*mL^{-1} in sorting buffer (PBS 1x + 3 mM EDTA + 0.8 % Trypsin + 1 % FBS). Cells were sorted with a BD FACSMelody™ cell sorter according to their fluorescence levels (**Fig. 3.4b**). 350000-150000 cells per gate were collected.

6.3 RNA level characterization

6.3.1 mRNA extraction and quantification

RNA extraction for downstream mRNA quantification was performed with E.Z.N.A.® Total RNA I (VWR). Cells were washed in PBS and incubated for 2 minutes in lysis buffer before resuspension. Protocol was followed according to manufacturer's instructions and RNA was eluted in 30 µL of RNAse free water to maximize the yield. RNA samples were conserved at -80°C. The protocol was performed exclusively with RNAse free water, in a RNAse free environment.

For reverse transcription to cDNA, PrimeScript RT kit with gDNA eraser (Takara) was used according to manufacturer's instructions. The protocol was performed on ice in a RNAse free environment to avoid RNA degradation. A negative control without Quantiscript Reverse Transcriptase was always prepared to check that samples were not contaminated with genomic DNA.

Fast SYBR Green Master Mix (ThermoFisher Scientific) was used to perform qPCR of cDNAs obtained from 500 ng of RNA and diluted 1:5. Samples were loaded in MicroAmp™ Fast Optical 96-Well Reaction Plate (0.1 mL) and the experiment was carried out with a CFX96 Touch Real-Time PCR Detection System (BioRad) machine. Each well contained 20 µL of final volume (7 µL SYBR Green Master Mix, 10 µL ddHO, 1 µL of each primer, 1 µL of template). Also, a control without template (blank) was set. Primers were designed to amplify a region of 60-200 bp (**Table 6.1**) and with a temperature of annealing between 50 and 65 °C. Data were analyzed using the Comparative Ct

Method according to Applied Biosystems Protocols.

Table 6.1: **List of the primers used for qPCR analyses.**

Primer	Function	Sequence (5'-3')
F7	Forward primer for mKate	GGTGTCTAAGGGCGAAGAGC
F8	Reverse primer for mKate	GCTGGTAGCCAGGATGTCGA
EGFP-F	Forward primer for EGFP	AAGGGCATCGACTTCAAG
EGFP-R	Reverse primer for EGFP	TGCTTGTCGGCCATGATATG
GAPDH-F	Forward primer for GAPDH	GAAGATGGTGATGGGATTTC
GAPDH-R	Reverse primer for GAPDH	GAAGTTGAAGGTCGGAGT
eIF4E Fw	Forward primer for eIf4E	AGAACAGATGGGCACTCTGG
eIF4E Rv	Reverse primer for eIf4E	TGAGTAGTCACAGCCAGGC
CyCA2 Fw	Forward primer for CyCA2	GGGACAAAGCTGGCCTGAATC
CyCA2 Rv	Reverse primer for CyCA2	AGGTAGGTCTGGTGAAGGTCC
18S Fw	Forward primer for 18S	TGTGCCGCTAGAGGTGAAATT
18S Rv	Reverse primer for 18S	TGGCAAATGCTTTCGCTTT

6.3.2 mature miRNA extraction and quantification

Total RNA, including small species, was isolated through the miRNeasy Mini Kit (Qiagen) following the manufacturer's instructions.

For detection of mature miRNA, cDNA was prepared with the miScript II RT Kit (Qiagen) using miScript HiSpec Buffer according to manufacturer's instructions.

qPCR was performed with miScript SYBR Green PCR kit (Qiagen) according to manufacturer's instructions, using cDNA diluted 1:10. miRNAs analyzed in Chapter 4 were detected using the corresponding miScript Primer Assay as listed in **Table 6.2**. The short non-coding RNAs SNORD72 and SNORD95 were used as internal reference (Catalog no. MS00033719 and MS00033726 respectively).

Table 6.2: **List of the miScript primers used to detect mature miRNAs by qPCR.**

miRNA	Catalog number
hsa-miR-31-5p	MS00003290
hsa-miR-106b-5p	MS00003402
hsa-miR-216a-5p	MS00009100
hsa-miR-135b-5p	MS00003472
hsa-miR-520a-5p	MS00007098
hsa-miR-633	MS00010409
hsa-miR-376c-5p	MS00045150
hsa-miR-100-5p	MS00031234

6.4 Polysome sequencing

Polysomal profiling was performed following the protocols described in [168, 169]. To obtain the cytoplasmic lysates, cells were treated with cycloheximide (10 µg*mL^{-1}) for 3-4 minutes and then lysed in 300 µL of cold hypotonic lysis buffer [168]. To remove nuclei, mitochondria and cellular debris, the lysates were centrifuged at 4°C for 5 min at 20000 g. To separate ribosomal subunits, ribosomes and polysomes from other cytoplasmic molecules, the supernatant was loaded on a 10 % – 40 % [w*v^{-1}] sucrose gradient and centrifuged for 1 h and 30 min at 260000 g at 4°C in a Beckman Optima LE-80 Ultracentrifuge. 1 mL fractions were collected and the absorbance at 254 nm was monitored with the UA-6 UV/VIS detector (Teledyne Isco). RNA was purified fraction by fraction using the phenol/chloroform extraction method as in [233]. The retro-transcription reaction was performed using the same volume of RNA for all polysomal fractions. The co-sedimentation profile of mRNAs was obtained by calculating the percentage (or fraction) of mRNAs in each fraction by qPCR as in [234].

6.5 Samples preparation for *omics* analyses

6.5.1 RNA-seq samples preparation

Qiagen RNeasy mini plus Kit (Qiagen) was used for RNA isolation. 10 µL of RNA at the exact concentration of 50 ng*µL^{-1} were provided for library preparation to the NGS facility of Tigem Telethon Institute (Pozzuoli, Italy).

Libraries were prepared using QuantSeq 3' mRNA sequencing for RNA quantification kit (Lexogen). Samples were processed with NovaSeq 6000 System (Illumina).

6.5.2 Proteomics samples preparation

Cell pellets were homogeinzed adding 200 µL of lysis RIPA buffer (ThermoFisher Scientifics) and incubated on ice for 15 minutes. Then, they were centrifuged at 13000 rpm for 30 min at 4 °C and the supernatant was collected. QuantiPro™ BCA Assay Kit (Sigma-Aldrich) was used according to manufacturer's instructions to measure protein concentration.

50 µg of proteins were aliquoted for digestion. First, 10 µl of reducing solution (100 mM DTT + 50 mM NH_4HCO_3 in MilliQ water) were added to each sample. Tubes were incubated at 56 °C for 30 minutes and then spinned. 30 µl of alkylating solution were added to each samples and tubes were incubated for 20 minutes at room temperature in the dark. 10x in volume of cold (-20 °C) acetone were added to each sample and tubes were at -20 °C overnight for protein precipitation.

Samples were centrifuge at 20000 g for 30 minutes at 4 °C. Supernatant was removed and pellet were dried 15 minutes under the hood. 200 µL of digestion buffer (50 mM NH_4HCO_3 in MilliQ water) were added to dissolve the pellet. 2 µL of trypsin solution (20 µg of trypsin dissolved in 40 µL of water + 0.1 % formic acid at final concentration 0.5 $\mu g*\mu L^{-1}$) were added to each sample and tubes were incubated at 37 °C overnight at 600 rpm.

Samples were dried under vacuum with SpeedVac.

Before loading into the instrument, the peptides were dried under vacuum and then dissolved in 100 µl of 3 % Acetonitrile (ACN) + 0.1 % Formic Acid (FA). The volume corresponding to 3 µg of peptides was injected into a NanoAcquity chromatographic system for LC-MS/MS analysis. The peptides were first loaded into a YMC-Triart C18 column (150 x 300 µm) at 5 $\mu L*min^{-1}$ for 4 minutes at 1 % ACN in water (+0.1 % FA) and then moved into a chromatographic column (Chrome XP C18 150 x 0.3 mm,) and eluted at 5 $\mu L*min^{-1}$ with the one-hour gradient from 2 to 35 % of ACN + 0.1 % FA. The column was then washed for 8 minutes with 95 % ACN + 0.1 % FA and re-equilibrated for 22 minutes at 2 % ACN + 0.1 % FA. The temperature of the column was kept at 45 °C.

The MS/MS analysis was conducted on the TripleTOF 5600+ equipped with DuoSpray Turbo V Source. For protein quantification, the mass spectrometer was operated in data-independent acquisition (DIA) mode, following the SWATH protocol for label free proteomics [235]. For SWATH, precursors ions were selected in the 400-1250 $m*z^{-1}$ range, with a variable window width from 7 to 50 Da. After a full range survey scan of 250 ms, 100 consecutive SWATH experiments (100-1500 $m*z^{-1}$) were performed, each lasting 25 ms. The instrument, operating in positive ion mode, was set as follows: ion spray voltage at 5000 V, spray gas 1 at 20, curtain gas at 30 and declustering potential at 80 V.

6.6 Data analysis and statistical method

6.6.1 RNA-Seq Analysis

Sequence reads were trimmed using bbduk software (https://www.lexogen.com/quantseq-data-analysis/) to remove adapter sequences, poly-A tails and low-quality end bases (regions with average quality below 6). Alignment was performed with STAR [236] on the Homo sapiens reference (hg38) provided by UCSC Genome Browser [[237]. The expression level of genes was determined with htseq-count [238] using the Gencode/Ensembl gene model. Differential expression analysis was performed using edgeR [239], a statistical package based on generalized linear models, suitable for multifactorial experiments. Only genes with FDR <0.05, |LogFC| > 0.5 were considered differentially expressed for each comparison.

6.6.2 Proteomics Analysis

For data analysis, the improved version of the PanHuman IL [240] was used to search the MS/MS spectra, using only the non-shared peptides and with no modifications. The parameters were set as follows: minimum peptide confidence 95 %, 50 ppm maximum mass tolerance, 20 minutes maximum RT tolerance, 6 MRM transitions per peptide and modified peptides were not allowed. For statistical analysis, raw data were imported in MarkerView software and normalized using Most Likely Ratio (MLR) method [241].

Differential protein expression among the different samples and conditions was inferred statistically using two-way ANOVA analysis. Significant differences between each condition and the corresponding control (Section 3.4.5) was determined using a Tukey's HSD test. Only differences with FDR < 0.1 and $|\text{LogFC}| > 0.5$ were considered significant.

6.6.3 Functional Pathway Analysis

Gene Sets Enrichment Analysis (GSEA) was performed using clusterProfiler R package [242] based on gene sets annotated in MsigDB [179], the genes list was ranked by statistical significance obtained from differential expression analysis. Only pathways with FDR $<$ 0.05 were considered enriched.

6.6.4 Statistics and reproducibility

Each experiment was repeated independently at least twice with similar results. Statistical significance was calculated using unpaired T-test where specified.

Contributions

Dr. Gabriella Viero *Polysome sequencing.*

Fabiana Tedeschi *Help with mRNA half-life experiments*

Alexander McKinnon (ICL) *Mathematical modelling for miRNA dependent resource reallocation.*

Dr. Antonio Rinaldi *RNA-seq and proteomics data analysis; implementation of the computational workflow for cell classifier.*

Dr. Antonella Iuliano (Tigem) *Initial RNA-seq and proteomics data analysis.*

Dr. Gennaro Gambardella (Tigem) *z-score calculation of the miRNAs in the selected cell lines for cell classifier.*

Rita Tarnai *qPCR data in Section 4.2.2.*

Individual contributions for published work is specified in the corresponding papers: [51] for Chapter 2 and [67] for Section 3.2.

References

1. Lienert, F., Lohmueller, J. J., Garg, A. & Silver, P. A. Synthetic biology in mammalian cells: Next generation research tools and therapeutics. *Nat. Rev. Mol. Cell Biol.* **15,** 95–107 (2014).
2. Black, J. B., Perez-Pinera, P. & Gersbach, C. A. Mammalian Synthetic Biology: Engineering Biological Systems. eng. *Annu. Rev. Biomed. Eng.* **19,** 249–277 (2017).
3. Clarke, L. & Kitney, R. Developing synthetic biology for industrial biotechnology applications. eng. *Biochem. Soc. Trans.* **48,** 113–122 (2020).
4. Schmidt, M. & de Lorenzo, V. Synthetic constructs in/for the environment: Managing the interplay between natural and engineered Biology. *FEBS Lett.* **586,** 2199–2206 (2012).
5. Kitada, T., DiAndreth, B., Teague, B. & Weiss, R. Programming gene and engineered-cell therapies with synthetic biology. *Science (80-.).* **359** (2018).
6. Brinkman, B. M., Zuijdeest, D., Kaijzel, E. L., Breedveld, F. C. & Verweij, C. L. Relevance of the tumor necrosis factor alpha (TNF alpha) -308 promoter polymorphism in TNF alpha gene regulation. eng. *J. Inflamm.* **46,** 32–41.
7. Bamforth, S. D. *et al.* Cardiac malformations, adrenal agenesis, neural crest defects and exencephaly in mice lacking Cited2, a new Tfap2 co-activator. eng. *Nat. Genet.* **29,** 469–474 (2001).
8. Farquhar, K. S. *et al.* Role of network-mediated stochasticity in mammalian drug resistance. eng. *Nat. Commun.* **10,** 2766 (2019).
9. Liu, W. *et al.* Mutations in AXIN2 cause colorectal cancer with defective mismatch repair by activating beta-catenin/TCF signalling. eng. *Nat. Genet.* **26,** 146–147 (2000).
10. Stuible, M. *et al.* Optimization of a high-cell-density polyethylenimine transfection method for rapid protein production in CHO-EBNA1 cells. eng. *J. Biotechnol.* **281,** 39–47 (2018).
11. Xie, M. *et al.* β-cell-mimetic designer cells provide closed-loop glycemic control. eng. *Science* **354,** 1296–1301 (2016).
12. Siciliano, V. *et al.* Engineering modular intracellular protein sensor-Actuator devices. eng. *Nat. Commun.* **9,** 1–7 (2018).
13. Elowitz, M. & Lim, W. A. Build life to understand it. *Nature* **468,** 889–890 (2010).
14. Pleiss, J. The promise of synthetic biology. *Appl. Microbiol. Biotechnol.* **73,** 735–739 (2006).
15. Elani, Y. *et al.* Constructing vesicle-based artificial cells with embedded living cells as organelle-like modules. *Sci. Rep.* **8,** 4564 (2018).

16. Jackson, D. A., Symons, R. H. & Berg, P. Biochemical method for inserting new genetic information into DNA of Simian Virus 40: circular SV40 DNA molecules containing lambda phage genes and the galactose operon of Escherichia coli. eng. *Proc. Natl. Acad. Sci. U. S. A.* **69,** 2904–2909 (1972).

17. Sanger, F., Nicklen, S. & Coulson, A. R. DNA sequencing with chain-terminating inhibitors. eng. *Proc. Natl. Acad. Sci. U. S. A.* **74,** 5463–5467 (1977).

18. Hughes, R. A. & Ellington, A. D. Synthetic DNA Synthesis and Assembly: Putting the Synthetic in Synthetic Biology. eng. *Cold Spring Harb. Perspect. Biol.* **9,** a023812 (2017).

19. Chu, L. & Robinson, D. K. Industrial choices for protein production by large-scale cell culture. eng. *Curr. Opin. Biotechnol.* **12,** 180–187 (2001).

20. Gardner, T. S., Cantor, C. R. & Collins, J. J. Construction of a genetic toggle switch in Escherichia coli. *Nature* **403,** 339–342 (2000).

21. Elowitz, M. B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403,** 335–338 (2000).

22. Khalil, A. S. & Collins, J. J. Synthetic biology: applications come of age. eng. *Nat. Rev. Genet.* **11,** 367–379 (2010).

23. Jinek, M. *et al.* A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity. eng. *Science* **337,** 816–821 (2012).

24. Doudna, J. A. & Charpentier, E. The new frontier of genome engineering with CRISPR-Cas9. *Science (80-.).* **346,** 1258096 (2014).

25. Cella, F. & Siciliano, V. Protein-based parts and devices that respond to intracellular and extracellular signals in mammalian cells. *Curr. Opin. Chem. Biol.* **52,** 47–53 (2019).

26. Bashor, C. J. & Collins, J. J. Understanding Biological Regulation Through Synthetic Biology. *Annu. Rev. Biophys.* **47,** 399–423 (2018).

27. Caliendo, F., Dukhinova, M. & Siciliano, V. Engineered Cell-Based Therapeutics: Synthetic Biology Meets Immunology. eng. *Front. Bioeng. Biotechnol.* **7,** 43 (2019).

28. Guo, Z. S. The 2018 Nobel Prize in medicine goes to cancer immunotherapy (editorial for BMC cancer). eng. *BMC Cancer* **18,** 1086 (2018).

29. Quinn, J. Y. *et al.* SBOL Visual: A Graphical Language for Genetic Designs. *PLOS Biol.* **13,** e1002310 (2015).

30. Mishra, D., Rivera, P. M., Lin, A., Del Vecchio, D. & Weiss, R. A load driver device for engineering modularity in biological networks. *Nat. Biotechnol.* **32,** 1268–1275 (2014).

31. Nielsen, A. A. K. *et al.* Genetic circuit design automation. *Science (80-.).* **352,** aac7341 (2016).

32. Hammer, K., Mijakovic, I. & Jensen, P. R. Synthetic promoter libraries–tuning of gene expression. eng. *Trends Biotechnol.* **24,** 53–55 (2006).

33. Davidson, E. A. & Ellington, A. D. Synthetic RNA circuits. eng. *Nat. Chem. Biol.* **3,** 23–28 (2007).

34. Topp, S. & Gallivan, J. P. Riboswitches in unexpected places–a synthetic riboswitch in a protein coding region. eng. *RNA* **14,** 2498–2503 (2008).

35. Win, M. N. & Smolke, C. D. Higher-order cellular information processing with synthetic RNA devices. eng. *Science* **322,** 456–460 (2008).

36. Beisel, C. L., Bayer, T. S., Hoff, K. G. & Smolke, C. D. Model-guided design of ligand-regulated RNAi for programmable control of gene expression. *Mol. Syst. Biol.* **4,** 224 (2008).

37. Kramer, B. P. *et al.* An engineered epigenetic transgene switch in mammalian cells. *Nat. Biotechnol.* **22,** 867–870 (2004).

38. Atkinson, M. R., Savageau, M. A., Myers, J. T. & Ninfa, A. J. Development of genetic circuitry exhibiting toggle switch or oscillatory behavior in Escherichia coli. eng. *Cell* **113,** 597–607 (2003).

39. Stricker, J. *et al.* A fast, robust and tunable synthetic gene oscillator. *Nature* **456,** 516–519 (2008).

40. Weber, W. *et al.* A synthetic time-delay circuit in mammalian cells and mice. eng. *Proc. Natl. Acad. Sci. U. S. A.* **104,** 2643–2648 (2007).

41. Guido, N. J. *et al.* A bottom-up approach to gene regulation. *Nature* **439,** 856–860 (2006).

42. Del Vecchio, D. Modularity, context-dependence, and insulation in engineered biological circuits. *Trends Biotechnol.* **33,** 111–119 (2015).

43. Brophy, J. A. N. & Voigt, C. A. Principles of genetic circuit design. eng. *Nat. Methods* **11,** 508–520 (2014).

44. Verbič, A., Praznik, A. & Jerala, R. A guide to the design of synthetic gene networks in mammalian cells. *FEBS J.* **n/a** (2020).

45. Singh, V. Recent advances and opportunities in synthetic logic gates engineering in living cells. eng. *Syst. Synth. Biol.* **8,** 271–282 (2014).

46. Saxena, P., Bojar, D. & Fussenegger, M. Design of Synthetic Promoters for Gene Circuits in Mammalian Cells. eng. *Methods Mol. Biol.* **1651,** 263–273 (2017).

47. Khalil, A. S. *et al.* A Synthetic Biology Framework for Programming Eukaryotic Transcription Functions. *Cell* **150,** 647–658 (2012).

48. Mathur, M., Xiang, J. S. & Smolke, C. D. Mammalian synthetic biology for studying the cell. *J. Cell Biol.* **216,** 73–82 (2016).

49. Wroblewska, L. *et al.* Mammalian synthetic circuits with RNA binding proteins for RNA-only delivery. *Nat. Biotechnol.* **33,** 839–841 (2015).

50. Matsuura, S. *et al.* Synthetic RNA-based logic computation in mammalian cells. *Nat. Commun.* **9,** 4847 (2018).

51. Cella, F., Wroblewska, L., Weiss, R. & Siciliano, V. Engineering protein-protein devices for multilayered regulation of mRNA translation using orthogonal proteases in mammalian cells. *Nat. Commun.* **9,** 4392 (2018).

52. Gao, X. J., Chong, L. S., Kim, M. S. & Elowitz, M. B. Programmable protein circuits in living cells. *Science (80-.).* **361,** 1252–1258 (2018).

53. Chen, Z. *et al.* De novo design of protein logic gates. **84,** 78–84 (2020).

54. Senior, A. W. *et al.* Improved protein structure prediction using potentials from deep learning. eng. *Nature* **577,** 706–710 (2020).

55. Chiesa, G., Kiriakov, S. & Khalil, A. S. Protein assembly systems in natural and synthetic biology. *BMC Biol.* **18,** 35 (2020).

56. Kawasaki, S., Ono, H., Hirosawa, M. & Saito, H. RNA and protein-based nanodevices for mammalian post-transcriptional circuits. *Curr. Opin. Biotechnol.* **63,** 99–110 (2020).

57. Székely, T. & Burrage, K. Stochastic simulation in systems biology. *Comput. Struct. Biotechnol. J.* **12,** 14–25 (2014).

58. Prado Casanova, M. Noise and Synthetic Biology: How to Deal with Stochasticity? *Nanoethics* **14,** 113–122 (2020).

59. Aoki, S. K. *et al.* A universal biomolecular integral feedback controller for robust perfect adaptation. *Nature* **570,** 533–537 (2019).

60. Frei, T., Chang, C.-H., Filo, M. & Khammash, M. Genetically Engineered Integral Feedback Controllers for Robust Perfect Adaptation in Mammalian Cells. *bioRxiv,* 2020.12.06.412304 (2020).

61. Karig, D. *et al.* Stochastic Turing patterns in a synthetic bacterial population. *Proc. Natl. Acad. Sci.* **115,** 6572 LP –6577 (2018).

62. Yu, T. C. *et al.* Multiplexed characterization of rationally designed promoter architectures deconstructs combinatorial logic for IPTG-inducible systems. *Nat. Commun.* **12,** 325 (2021).

63. Ceroni, F., Algar, R., Stan, G.-B. & Ellis, T. Quantifying cellular capacity identifies gene expression designs with reduced burden. *Nat. Methods* **12,** 415–418 (2015).

64. Ceroni, F. *et al.* Burden-driven feedback control of gene expression. *Nat. Methods* **15,** 387–393 (2018).

65. Carbonell-Ballestero, M., Garcia-Ramallo, E., Montañez, R., Rodriguez-Caso, C. & Macía, J. Dealing with the genetic load in bacterial synthetic biology circuits: convergences with the Ohm's law. eng. *Nucleic Acids Res.* **44,** 496–507 (2016).

66. Qian, Y., Huang, H.-H., Jiménez, J. I. & Del Vecchio, D. Resource Competition Shapes the Response of Genetic Circuits. *ACS Synth. Biol.* **6,** 1263–1272 (2017).

67. Frei, T. *et al.* Characterization and mitigation of gene expression burden in mammalian cells. *Nat. Commun.* **11,** 1–14 (2020).

68. Jones, R. D. *et al.* An endoribonuclease-based feedforward controller for decoupling resource-limited genetic modules in mammalian cells. *Nat. Commun.* **11,** 5690 (2020).

69. Segall-Shapiro, T. H., Meyer, A. J., Ellington, A. D., Sontag, E. D. & Voigt, C. A. A 'resource allocator' for transcription based on a highly fragmented T7 RNA polymerase. *Mol. Syst. Biol.* **10,** 742 (2014).

70. Lillacci, G., Benenson, Y. & Khammash, M. Synthetic control systems for high performance gene expression in mammalian cells. *Nucleic Acids Res.* **46,** 9855–9863 (2018).

71. Cuccato, G., Della Gatta, G. & di Bernardo, D. Systems and Synthetic biology: tackling genetic networks and complex diseases. eng. *Heredity (Edinb).* **102,** 527–532 (2009).

72. Huang, H.-H., Qian, Y. & Del Vecchio, D. A quasi-integral controller for adaptation of genetic modules to variable ribosome demand. *Nat. Commun.* **9,** 5415 (2018).

73. Gam, J. J., Babb, J. & Weiss, R. A mixed antagonistic/synergistic miRNA repression model enables accurate predictions of multi-input miRNA sensor activity. *Nat. Commun.* **9** (2018).

74. Nowicka, M. & Siebert, H. *Designing Distributed Cell Classifier Circuits Using a Genetic Algorithm BT - Computational Methods in Systems Biology* in (eds Bortolussi, L. & Sanguinetti, G.) (Springer International Publishing, Cham, 2019), 96–119.

75. Mohammadi, P., Beerenwinkel, N. & Benenson, Y. Automated Design of Synthetic Cell Classifier Circuits Using a Two-Step Optimization Strategy. *Cell Syst.* **4,** 207–218.e14 (2017).

76. Xie, Z., Wroblewska, L., Prochazka, L., Weiss, R. & Benenson, Y. Multi-input RNAi-based logic circuit for identification of specific cancer cells. *Science* **333,** 1307–11 (2011).

77. Kallunki, T., Barisic, M., Jäättelä, M. & Liu, B. How to Choose the Right Inducible Gene Expression System for Mammalian Studies? *Cells* **8,** 1–16 (2019).

78. Panayotakis, S. Chapter 19. *"The Story Apollonius, King Tyre"* **1651,** 263–273 (2013).

79. Weber, W. & Fussenegger, M. *Engineering of Synthetic Mammalian Gene Networks* 2009.

80. Tavernier, G. *et al.* MRNA as gene therapeutic: How to control protein expression. *J. Control. Release* **150,** 238–247 (2011).

81. Dixon, N. *et al.* Reengineering orthogonally selective riboswitches. *Proc. Natl. Acad. Sci. U. S. A.* **107,** 2830–2835 (2010).

82. Culler, S. J., Hoff, K. G. & Smolke, C. D. Reprogramming cellular behavior with RNA controllers responsive to endogenous proteins. *Science (80-.).* **330,** 1251–1255 (2010).

83. Chen, X., Denison, L., Levy, M. & Ellington, A. D. Direct selection for ribozyme cleavage activity in cells. *Rna* **15,** 2035–2045 (2009).

84. Berens, C., Groher, F. & Suess, B. RNA aptamers as genetic control devices: The potential of riboswitches as synthetic elements for regulating gene expression. *Biotechnol. J.* **10,** 246–257 (2015).

85. Westmann, C. A., Guazzaroni, M.-E. & Silva-Rocha, R. Engineering Complexity in Bacterial Regulatory Circuits for Biotechnological Applications. *mSystems* **3,** e00151–17 (2018).

86. Nissim, L. *et al.* Synthetic RNA-Based Immunomodulatory Gene Circuits for Cancer Immunotherapy. *Cell* **171,** 1138–1150.e15 (2017).

87. Kallen, K. J. & Theß, A. A development that may evolve into a revolution in medicine: MRNA as the basis for novel, nucleotide-based vaccines and drugs. eng. *Ther. Adv. Vaccines* **2,** 10–31 (2014).

88. Sahin, U., Karikó, K. & Türeci, Ö. MRNA-based therapeutics-developing a new class of drugs. *Nat. Rev. Drug Discov.* **13,** 759–780 (2014).

89. Bloom, K., van den Berg, F. & Arbuthnot, P. Self-amplifying RNA vaccines for infectious diseases. *Gene Ther.* (2020).

90. Yoshioka, N. *et al.* Efficient generation of human iPSCs by a synthetic self-replicative RNA. *Cell Stem Cell* **13,** 246–254 (2013).

91. Beal, J. *et al.* Model-driven engineering of gene expression from RNA replicons. *ACS Synth. Biol.* **4,** 48–56 (2015).

92. Li, Y. *et al.* In vitro evolution of enhanced RNA replicons for immunotherapy. *Sci. Rep.* **9,** 6932 (2019).

93. Fernandez-Rodriguez, J. & Voigt, C. A. Post-translational control of genetic circuits using Potyvirus proteases. *Nucleic Acids Res.* **44,** 6493–6502 (2016).

94. Barnea, G. *et al.* The genetic design of signaling cascades to record receptor activation. *Proc. Natl. Acad. Sci. U. S. A.* **105,** 64–69 (2008).

95. Schwarz, K. A., Daringer, N. M., Dolberg, T. B. & Leonard, J. N. Rewiring human cellular input-output using modular extracellular sensors. *Nat. Chem. Biol.* **13,** 202–209 (2017).

96. López-Otín, C. & Bond, J. S. Proteases: multifunctional enzymes in life and disease. eng. *J. Biol. Chem.* **283,** 30433–30437 (2008).

97. Nuñez, G., Benedict, M. A., Hu, Y. & Inohara, N. Caspases: the proteases of the apoptotic pathway. *Oncogene* **17,** 3237–3245 (1998).

98. Rao, M. B., Tanksale, A. M., Ghatge, M. S. & Deshpande, V. V. Molecular and biotechnological aspects of microbial proteases. eng. *Microbiol. Mol. Biol. Rev.* **62,** 597–635 (1998).

99. Liu, X. & Kokare, C. in (ed Brahmachari, G. B. T. B. o. M. E.) 267–298 (Academic Press, 2017).

100. Phan, J. *et al.* Structural basis for the substrate specificity of tobacco etch virus protease. *J. Biol. Chem.* **277,** 50564–50572 (2002).

101. Wieland, M. *et al.* Engineered UV-A light-responsive gene expression system for measuring sun cream efficacy in mammalian cell culture. eng. *J. Biotechnol.* **189,** 150–153 (2014).

102. Gray, D. C., Mahrus, S. & Wells, J. A. Activation of specific apoptotic caspases with an engineered small-molecule-activated protease. *Cell* **142,** 637–646 (2010).

103. Pauli, A. *et al.* Cell-Type-Specific TEV Protease Cleavage Reveals Cohesin Functions in Drosophila Neurons. *Dev. Cell* **14,** 239–251 (2008).

104. Wehr, M. C. *et al.* Monitoring regulated protein-protein interactions using split TEV. *Nat. Methods* **3,** 985–993 (2006).

105. Uhlmann, F., Wernic, D., Poupart, M. A., Koonin, E. V. & Nasmyth, K. Cleavage of cohesin by the CD clan protease separin triggers anaphase in yeast. *Cell* **103,** 375–386 (2000).

106. Moser, F. *et al.* Dynamic control of endogenous metabolism with combinatorial logic circuits. *Mol. Syst. Biol.* **14,** 1–18 (2018).

107. Nallamsetty, S. *et al.* Efficient site-specific processing of fusion proteins by tobacco vein mottling virus protease in vivo and in vitro. *Protein Expr. Purif.* **38,** 108–115 (2004).

108. Kang, H., Lee, Y. J., Goo, J. H. & Park, W. J. Determination of the substrate specificity of turnip mosaic virus Nla protease using a genetic method. *J. Gen. Virol.* **82,** 3115–3117 (2001).

109. Moore, T., Zhang, Y., Fenley, M. O. & Li, H. Molecular basis of box C/D RNA-protein interactions: Cocrystal structure of archaeal L7Ae and a box C/D RNA. *Structure* **12,** 807–818 (2004).

110. Robertson, B. *et al.* Specificity and functionality of microRNA inhibitors. *Silence* **1,** 1–11 (2010).

111. Lindenbach, B. D. & Rice, C. M. Unravelling hepatitis C virus replication from genome to function. *Nature* **436,** 933–938 (2005).

112. Gal-Tanamy, M. *et al.* HCV NS3 serine protease-neutralizing single-chain antibodies isolated by a novel genetic screen. *J. Mol. Biol.* **347,** 991–1003 (2005).

113. Peabody, D. S. coat. **12,** 595–600 (1993).

114. Yonashiro, R. *et al.* Mutant SOD1 and Attenuates Mutant SOD1-induced Reactive Oxygen Species Generation. *Mol. Biol. Cell* **20,** 4524–4530 (2009).

115. Kapust, R. B., Toözseór, J., Copeland, T. D. & Waugh, D. S. The P1 specificity of tobacco etch virus protease. *Biochem. Biophys. Res. Commun.* **294,** 949–955 (2002).

116. Thyme, S. & Baker, D. Redesigning the specificity of protein-DNA interactions with Rosetta. eng. *Methods Mol. Biol.* **1123,** 265–282 (2014).

117. Bennett, M. R. *et al.* Metabolic gene regulation in a dynamically changing environment. *Nature* **454,** 1119–1122 (2008).

118. Chechik, G. *et al.* Activity motifs reveal principles of timing in transcriptional control of the yeast metabolic network. eng. *Nat. Biotechnol.* **26,** 1251–1259 (2008).

119. Zaslaver, A. *et al.* Just-in-time transcription program in metabolic pathways. eng. *Nat. Genet.* **36,** 486–491 (2004).

120. Wang, L. *et al.* Bistable switches control memory and plasticity in cellular differentiation. *Proc. Natl. Acad. Sci. U. S. A.* **106,** 6638–6643 (2009).

121. Chatterjee, A., Kaznessis, Y. N. & Hu, W. S. Tweaking biological switches through a better understanding of bistability behavior. eng. *Curr. Opin. Biotechnol.* **19,** 475–481 (2008).

122. Rata, S. *et al.* Two Interlinked Bistable Switches Govern Mitotic Control in Mammalian Cells. *Curr. Biol.* **28,** 3824–3832.e6 (2018).

123. Huang, D., Holtz, W. J. & Maharbiz, M. M. A genetic bistable switch utilizing nonlinear protein degradation. *J. Biol. Eng.* **6,** 9 (2012).

124. Morsut, L. *et al.* Engineering Customized Cell Sensing and Response Behaviors Using Synthetic Notch Receptors. *Cell* **164,** 780–791 (2016).

125. Han, X. *et al.* Programmable Synthetic Protein Circuits for the Identification and Suppression of Hepatocellular Carcinoma. *Mol. Ther. - Oncolytics* **17,** 70–82 (2020).

126. Gordley, R. M. *et al.* Engineering dynamical control of cell fate switching using synthetic phospho-regulons. *Proc. Natl. Acad. Sci.* **113,** 13528 LP –13533 (2016).

127. Scheller, L. *et al.* Phosphoregulated orthogonal signal transduction in mammalian cells. *Nat. Commun.* **11,** 3085 (2020).

128. Ross, D. & Huh, J. Patent Application Publication (10) Pub . No .: US 2020/0056189 A1. **2020** (2020).

129. Miki, K. *et al.* Efficient Detection and Purification of Cell Populations Using Synthetic MicroRNA Switches. *Cell Stem Cell* **16,** 699–711 (2015).

130. Ma, C. *et al.* SUMO-specific protease 1 regulates pancreatic cancer cell proliferation and invasion by targeting MMP-9. *Tumor Biol.* **35,** 12729–12735 (2014).

131. Cui, C.-P. *et al.* SENP1 promotes hypoxia-induced cancer stemness by HIF-1α deSUMOylation and SENP1/HIF-1α positive feedback loop. eng. *Gut* **66,** 2149–2159 (2017).

132. Xiang-ming, Y. *et al.* SENP1 regulates cell migration and invasion in neuroblastoma. *Biotechnol. Appl. Biochem.* **63,** 435–440 (2016).

133. Ho, P., Ede, C. & Chen, Y. Y. Modularly Constructed Synthetic Granzyme B Molecule Enables Interrogation of Intracellular Proteases for Targeted Cytotoxicity. *ACS Synth. Biol.* **6,** 1484–1495 (2017).

134. Moriya, H. Quantitative nature of overexpression experiments. eng. *Mol. Biol. Cell* **26,** 3932–3939 (2015).

135. Schmidt, S. F., Larsen, B. D., Loft, A. & Mandrup, S. Cofactor squelching: Artifact or fact? eng. *Bioessays* **38,** 618–626 (2016).

136. Materna, S. C. & Marwan, W. Estimating the number of plasmids taken up by a eukaryotic cell during transfection and evidence that antisense RNA abolishes gene expression in Physarum polycephalum. *FEMS Microbiol. Lett.* **243,** 29–35 (2005).

137. Segall-Shapiro, T. H., Sontag, E. D. & Voigt, C. A. Engineered promoters enable constant gene expression at any copy number in bacteria. eng. *Nat. Biotechnol.* **36,** 352–358 (2018).

138. Ma, W., Trusina, A., El-Samad, H., Lim, W. A. & Tang, C. Defining Network Topologies that Can Achieve Biochemical Adaptation. *Cell* **138,** 760–773 (2009).

139. Briat, C., Gupta, A. & Khammash, M. Antithetic Integral Feedback Ensures Robust Perfect Adaptation in Noisy Biomolecular Networks. *Cell Syst.* **2,** 15–26 (2016).

140. Li, G.-W., Burkhardt, D., Gross, C. & Weissman, J. S. Quantifying absolute protein synthesis rates reveals principles underlying allocation of cellular resources. eng. *Cell* **157,** 624–635 (2014).

141. Scott, M., Gunderson, C. W., Mateescu, E. M., Zhang, Z. & Hwa, T. Interdependence of cell growth and gene expression: origins and consequences. eng. *Science* **330,** 1099–1102 (2010).

142. Weiße, A. Y., Oyarzún, D. A., Danos, V. & Swain, P. S. Mechanistic links between cellular trade-offs, gene expression, and growth. *Proc. Natl. Acad. Sci.* **112,** E1038 LP –E1047 (2015).

143. Gyorgy, A. *et al.* Isocost Lines Describe the Cellular Economy of Genetic Circuits. *Biophys. J.* **109,** 639–646 (2015).

144. Kafri, M., Metzl-Raz, E., Jona, G. & Barkai, N. The Cost of Protein Production. *Cell Rep.* **14,** 22–31 (2016).

145. Kheir Gouda, M., Manhart, M. & Balázsi, G. Evolutionary regain of lost gene circuit function. *Proc. Natl. Acad. Sci.* **116,** 25162 LP –25171 (2019).

146. Day, D. A. & Tuite, M. F. Post-transcriptional gene regulatory mechanisms in eukaryotes: An overview. *J. Endocrinol.* **157,** 361–371 (1998).

147. Ameres, S. L. & Zamore, P. D. Diversifying microRNA sequence and function. *Nat. Rev. Mol. Cell Biol.* **14,** 475–488 (2013).

148. Meleady, P. *et al.* Sustained productivity in recombinant Chinese Hamster Ovary (CHO) cell lines: Proteome analysis of the molecular basis for a process-related phenotype. *BMC Biotechnol.* **11** (2011).

149. Thomson, D. W., Bracken, C. P. & Goodall, G. J. Experimental strategies for microRNA target identification. eng. *Nucleic Acids Res.* **39,** 6845–6853 (2011).

150. Bleris, L. *et al.* Synthetic incoherent feedforward circuits show adaptation to the amount of their genetic template. *Mol. Syst. Biol.* **7** (2011).

151. Strovas, T. J., Rosenberg, A. B., Kuypers, B. E., Muscat, R. A. & Seelig, G. against Perturbations (2014).

152. Zechner, C., Seelig, G., Rullan, M. & Khammash, M. Molecular circuits for dynamic noise filtering. *Proc. Natl. Acad. Sci. U. S. A.* **113,** 4729–4734 (2016).

153. Tian, W. *et al.* High-Throughput Functional MicroRNAs Profiling by Recombinant AAV-Based MicroRNA Sensor Arrays. *PLoS One* **7,** e29551 (2012).

154. Aprelikova, O. *et al.* The role of miR-31 and its target gene SATB2 in cancer-associated fibroblasts. *Cell Cycle* **9,** 4387–4398 (2010).

155. Fabian, M. R., Sonenberg, N. & Filipowicz, W. Regulation of mRNA translation and stability by microRNAs. eng. *Annu. Rev. Biochem.* **79,** 351–379 (2010).

156. Saito, H. *et al.* Synthetic translational regulation by an L7Ae-kink-turn RNP switch. *Nat. Chem. Biol.* **6,** 71–78 (2010).

157. Lee, I. *et al.* New class of microRNA targets containing simultaneous 5'-UTR and 3'-UTR interaction sites. eng. *Genome Res.* **19,** 1175–1183 (2009).

158. O'Brien, J., Hayder, H., Zayed, Y. & Peng, C. Overview of MicroRNA Biogenesis, Mechanisms of Actions, and Circulation. eng. *Front. Endocrinol. (Lausanne).* **9,** 402 (2018).

159. Jonas, S. & Izaurralde, E. Towards a molecular understanding of microRNA-mediated gene silencing. *Nat. Rev. Genet.* **16,** 421–433 (2015).

160. Mullany, L. E., Herrick, J. S., Wolff, R. K. & Slattery, M. L. MicroRNA Seed Region Length Impact on Target Messenger RNA Expression and Survival in Colorectal Cancer. *PLoS One* **11,** e0154177 (2016).

161. Djuranovic, S., Nahvi, A. & Green, R. miRNA-mediated gene silencing by translational repression followed by mRNA deadenylation and decay. eng. *Science* **336,** 237–240 (2012).

162. Venturelli, O. S. *et al.* Programming mRNA decay to modulate synthetic circuit resource allocation. *Nat. Commun.* **8,** 15128 (2017).

163. Khajuria, R. K. *et al.* Ribosome Levels Selectively Regulate Translation and Lineage Commitment in Human Hematopoiesis. *Cell* **173,** 90–103.e19 (2018).

164. Cookson, N. A. *et al.* Queueing up for enzymatic processing: Correlated signaling through coupled degradation. *Mol. Syst. Biol.* **7** (2011).

165. Gambardella, G. *et al.* The impact of microRNAs on transcriptional heterogeneity and gene co-expression across single embryonic stem cells. *Nat. Commun.* **8** (2017).

166. Moretti, F., Thermann, R. & Hentze, M. W. Mechanism of translational regulation by miR-2 from sites in the 5 untranslated region or the open reading frame. *Rna* **16,** 2493–2502 (2010).

167. Stripecke, R., Oliveira, C. C., McCarthy, J. E. & Hentze, M. W. Proteins binding to 5' untranslated region sites: a general mechanism for translational regulation of mRNAs in human and yeast cells. *Mol. Cell. Biol.* **14,** 5898–5909 (1994).

168. Lunelli, L. *et al.* Peering at brain polysomes with atomic force microscopy. *J. Vis. Exp.* **2016,** 1–8 (2016).

169. Bernabò, P. *et al.* In Vivo Translatome Profiling in Spinal Muscular Atrophy Reveals a Role for SMN Protein in Ribosome Biology. *Cell Rep.* **21,** 953–965 (2017).

170. Bensaude, O. Inhibiting eukaryotic transcription: Which compound to choose? How to evaluate its activity? *Transcription* **2,** 103–108 (2011).

171. Jauhiainen, A. *et al.* Distinct Cytoplasmic and Nuclear Functions of the Stress Induced Protein DDIT3/CHOP/GADD153. *PLoS One* **7,** e33208 (2012).

172. Ku, H.-C. & Cheng, C.-F. *Master Regulator Activating Transcription Factor 3 (ATF3) in Metabolic Homeostasis and Cancer* 2020.

173. Memon, A. & Lee, W. K. KLF10 as a Tumor Suppressor Gene and Its TGF-β Signaling. eng. *Cancers (Basel).* **10,** 161 (2018).

174. Brennan, M. S., Matos, M. F., Richter, K. E., Li, B. & Scannevin, R. H. The NRF2 transcriptional target, OSGIN1, contributes to monomethyl fumarate-mediated cytoprotection in human astrocytes. *Sci. Rep.* **7,** 42054 (2017).

175. Bowling, F. Z. *et al.* Crystal structure of human PLD1 provides insight into activation by PI(4,5)P2 and RhoA. *Nat. Chem. Biol.* **16,** 400–407 (2020).

176. Basso, F. *et al.* Comparison of the effects of PRKAR1A and PRKAR2B depletion on signaling pathways, cell growth, and cell cycle control of adrenocortical cells. eng. *Horm. Metab. Res. = Horm. und Stoffwechselforsch. = Horm. Metab.* **46,** 883–888 (2014).

177. Wang, X., Liu, Q. & Zhang, B. Leveraging the complementary nature of RNA-Seq and shotgun proteomics data. eng. *Proteomics* **14,** 2676–2687 (2014).

178. Singh, A., Cao, K.-a. L. & Rohart, F. mixOmics : An R package for ' omics feature selection and multiple data integration, 1–19 (2017).

179. Subramanian, A. *et al.* Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc. Natl. Acad. Sci. U. S. A.* **102,** 15545–15550 (2005).

180. Papoff, G. *et al.* CASP4 gene silencing in epithelial cancer cells leads to impairment of cell migration, cell-matrix adhesion and tissue invasion. *Sci. Rep.* **8,** 17705 (2018).

181. Carignano, A., Mukherjee, S., Singh, A. & Seelig, G. Extrinsic Noise Suppression in Micro RNA mediated Incoherent Feedforward Loops. *bioRxiv,* 422394 (2018).

182. Geisler, A. & Fechner, H. MicroRNA-regulated viral vectors for gene therapy. eng. *World J. Exp. Med.* **6,** 37–54 (2016).

183. Seyhan, A. A. A multiplexed miRNA and transgene expression platform for simultaneous repression and expression of protein coding sequences. eng. *Mol. Biosyst.* **12,** 295–312 (2016).

184. Laganà, A. *et al.* miR-Synth: a computational resource for the design of multi-site multi-target synthetic miRNAs. eng. *Nucleic Acids Res.* **42,** 5416–5425 (2014).

185. Denzler, R. *et al.* Impact of MicroRNA Levels, Target-Site Complementarity, and Cooperativity on Competing Endogenous RNA-Regulated Gene Expression. *Mol. Cell* **64,** 565–579 (2016).

186. Guinn, M. T. & Balázsi, G. Noise-reducing optogenetic negative-feedback gene circuits in human cells. eng. *Nucleic Acids Res.* **47,** 7703–7714 (2019).

187. Barber, R. D., Harmer, D. W., Coleman, R. A. & Clark, B. J. GAPDH as a housekeeping gene: analysis of GAPDH mRNA expression in a panel of 72 human tissues. eng. *Physiol. Genomics* **21,** 389–395 (2005).

188. Nie, X. *et al.* An appropriate loading control for western blot analysis in animal models of myocardial ischemic infarction. eng. *Biochem. Biophys. reports* **12,** 108–113 (2017).

189. Sambandan, S. *et al.* Activity-dependent spatially localized miRNA maturation in neuronal dendrites. *Science (80-.).* **355,** 634 LP –637 (2017).

190. Vicidomini, G., Bianchini, P. & Diaspro, A. STED super-resolved microscopy. *Nat. Methods* **15,** 173–182 (2018).

191. Zanacchi, F. C. & Diaspro, A. in (ed Roberts, G. C. K.) 2475–2476 (Springer Berlin Heidelberg, Berlin, Heidelberg, 2013).

192. Cox, S. Super-resolution imaging in live cells. eng. *Dev. Biol.* **401,** 175–181 (2015).

193. Rosa-Mercado, N. A. *et al.* Hyperosmotic stress alters the RNA polymerase II interactome and induces readthrough transcription despite widespread transcriptional repression. *Mol. Cell* (2021).

194. Schenk, S. *et al.* Combined transcriptome and proteome profiling reveals specific molecular brain signatures for sex, maturation and circalunar clock phase. eng. *Elife* **8** (2019).

195. Irmisch, A. *et al.* The Tumor Profiler Study: integrated, multi-omic, functional tumor profiling for clinical decision support. *Cancer Cell* (2021).

196. Origel Marmolejo, C. A., Bachhav, B., Patibandla, S. D., Yang, A. L. & Segatori, L. A gene signal amplifier platform for monitoring the unfolded protein response. *Nat. Chem. Biol.* **16,** 520–528 (2020).

197. Espah Borujeni, A., Zhang, J., Doosthosseini, H., Nielsen, A. A. K. & Voigt, C. A. Genetic circuit characterization by inferring RNA polymerase movement and ribosome usage. *Nat. Commun.* **11,** 5001 (2020).

198. Xiang, J. S. *et al.* Massively parallel RNA device engineering in mammalian cells with RNA-Seq. *Nat. Commun.* **10,** 4327 (2019).

199. Harimoto, T. Reprogramming the proteome for cell-free protein expression. *Synth. Biol.* **5** (2020).

200. Contreras-Llano, L. E. *et al.* Holistic engineering of cell-free systems through proteome-reprogramming synthetic circuits. *Nat. Commun.* **11,** 3138 (2020).

201. Gorochowski, T. E. *et al.* Genetic circuit characterization and debugging using RNA-seq. eng. *Mol. Syst. Biol.* **13,** 952 (2017).

202. Marx, V. Method of the Year: spatially resolved transcriptomics. *Nat. Methods* **18,** 9–14 (2021).

203. Angelici, B. *et al.* Synthetic Biology Platform for Sensing and Integrating Endogenous Transcriptional Inputs in Mammalian Cells. *Cell Rep.,* 2525–2537 (2016).

204. Wu, H. *et al.* miRNA Profiling of Naïve, Effector and Memory CD8 T Cells. *PLoS One* **2,** e1020 (2007).

205. Becker, K., Klarner, H., Nowicka, M. & Siebert, H. *Designing miRNA-Based Synthetic Cell Classifier Circuits Using Answer Set Programming* 2018.

206. Brewka, G., Eiter, T. & Truszczyński, M. Answer Set Programming at a Glance. *Commun. ACM* **54,** 92–103 (2011).

207. Didovyk, A. *et al.* Distributed Classifier Based on Genetically Engineered Bacterial Cell Cultures. *ACS Synth. Biol.* **4,** 72–82 (2015).

208. Brancati, G. & Großhans, H. An interplay of miRNA abundance and target site architecture determines miRNA activity and specificity. eng. *Nucleic Acids Res.* **46,** 3259–3269 (2018).

209. Barretina, J. *et al.* The Cancer Cell Line Encyclopedia enables predictive modelling of anticancer drug sensitivity. eng. *Nature* **483,** 603–607 (2012).

210. Ghandi, M. *et al.* Next-generation characterization of the Cancer Cell Line Encyclopedia. *Nature* **569,** 503–508 (2019).

211. Bignotti, E. *et al.* Identification of stably expressed reference small non-coding RNAs for microRNA quantification in high-grade serous ovarian carcinoma tissues. eng. *J. Cell. Mol. Med.* **20,** 2341–2348 (2016).

212. Ross, D. T. *et al.* Systematic variation in gene expression patterns in human cancer cell lines. *Nat. Genet.* **24,** 227–235 (2000).

213. Gallo, C. A., Carballido, J. A. & Ponzoni, I. Discovering time-lagged rules from microarray data using gene profile classifiers. eng. *BMC Bioinformatics* **12,** 123 (2011).

214. Prochazka, L., Benenson, Y. & Zandstra, P. W. Synthetic gene circuits and cellular decision-making in human pluripotent stem cells. *Curr. Opin. Syst. Biol.* **5,** 93–103 (2017).

215. Gramelsberger, G. Synthetic Morphology: A Vision of Engineering Biological Form. *J. Hist. Biol.* **53,** 295–309 (2020).

216. Cameron, D. E., Bashor, C. J. & Collins, J. J. A brief history of synthetic biology. *Nat. Rev. Microbiol.* **12,** 381–390 (2014).

217. French, K. E. Harnessing synthetic biology for sustainable development. *Nat. Sustain.* **2,** 250–252 (2019).

218. El Karoui, M., Hoyos-Flight, M. & Fletcher, L. *Future Trends in Synthetic Biology—A Report* 2019.

219. Duportet, X. *et al.* A platform for rapid prototyping of synthetic gene networks in mammalian cells. *Nucleic Acids Res.* **42,** 13440–13451 (2014).

220. Roybal, K. T. & Lim, W. A. Synthetic Immunology: Hacking Immune Cells to Expand Their Therapeutic Capabilities. eng. *Annu. Rev. Immunol.* **35,** 229–253 (2017).

221. Ilic, D. & Liovic, M. Industry updates from the field of stem cell research and regenerative medicine in June 2020. *Regen. Med.* **15,** 2145–2152 (2020).

222. Kojima, R., Aubel, D. & Fussenegger, M. Building sophisticated sensors of extracellular cues that enable mammalian cells to work as "doctors" in the body. *Cell. Mol. Life Sci.* **77,** 3567–3581 (2020).

223. Negre, O. *et al.* Gene Therapy of the β-Hemoglobinopathies by Lentiviral Transfer of the β(A(T87Q))-Globin Gene. eng. *Hum. Gene Ther.* **27,** 148–165 (2016).

224. Guo, W. & Song, H. Development of Gene Therapeutics for Head and Neck Cancer in China: From Bench to Bedside. eng. *Hum. Gene Ther.* **29,** 180–187 (2018).

225. Xie, M., Viviani, M. & Fussenegger, M. Engineering Precision Therapies: Lessons and Motivations from the Clinic. *Synth. Biol.* (2020).

226. Khammash, M., Di Bernardo, M. & Di Bernardo, D. Cybergenetics: Theory and Methods for Genetic Control System. *Proc. IEEE Conf. Decis. Control* **2019-Decem,** 916–926 (2019).

227. Lugagne, J.-B. *et al.* Balancing a genetic toggle switch by real-time feedback control and periodic forcing. *Nat. Commun.* **8,** 1671 (2017).

228. Benzinger, D. & Khammash, M. Pulsatile inputs achieve tunable attenuation of gene expression variability and graded multi-gene regulation. *Nat. Commun.* **9,** 3521 (2018).

229. Ye, H., Daoud-El Baba, M., Peng, R.-W. & Fussenegger, M. A synthetic optogenetic transcription device enhances blood-glucose homeostasis in mice. eng. *Science* **332,** 1565–1568 (2011).

230. Duportet, X. A platform for rapid prototyping of synthetic gene networks in mammalian cells. en. *Nucleic Acids Res.* **42,** 13440–13451 (2014).

231. DeNies, M. S., Liu, A. P. & Schnell, S. Are the biomedical sciences ready for synthetic biology? English. *Biomol. Concepts* **11,** 23–31.

232. Wimmer, E. *Synthetic Biology, Dual Use Research, and Possibilities for Control BT - Defence Against Bioterrorism* in (eds Radosavljevic, V., Banjari, I. & Belojevic, G.) (Springer Netherlands, Dordrecht, 2018), 7–11.

233. Tebaldi, T. *et al.* Widespread uncoupling between transcriptome and translatome variations after a stimulus in mammalian cells. *BMC Genomics* **13** (2012).

234. Darnell, J. C. *et al.* FMRP stalls ribosomal translocation on mRNAs linked to synaptic function and autism. *Cell* **146,** 247–261 (2011).

235. Huang, Q. *et al.* SWATH enables precise label-free quantification on proteome scale. eng. *Proteomics* **15,** 1215–1223 (2015).

236. Dobin, A. *et al.* STAR: Ultrafast universal RNA-seq aligner. *Bioinformatics* **29,** 15–21 (2013).

237. Kent, W. J. *et al.* The Human Genome Browser at UCSC. eng. *Genome Res.* **12,** 996–1006 (2002).

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

239. Robinson, M. D., McCarthy, D. J. & Smyth, G. K. edgeR: A Bioconductor package for differential expression analysis of digital gene expression data. eng. *Bioinformatics* **26,** 139–140 (2009).

240. Rosenberger, G. *et al.* A repository of assays to quantify 10,000 human proteins by SWATH-MS. *Sci. Data* **1,** 140031 (2014).

241. Wu, J. X. *et al.* SWATH mass spectrometry performance using extended peptide MS/MS assay libraries. *Mol. Cell. Proteomics* **15,** 2501–2514 (2016).

242. Yu, G., Wang, L. G., Han, Y. & He, Q. Y. ClusterProfiler: An R package for comparing biological themes among gene clusters. eng. *Omi. A J. Integr. Biol.* **16,** 284–287 (2012).

Appendix A

Modelling framework for gene expression in resource-limited environment

A.1 Derivation of the effective production rate constant $k_i^{\mathbf{eff}}(A_1, \ldots, A_n)$

Here is show the derivation of the effective production rate constant as the quasi-steady-state solution for the complexes formed between substrate species A_i and their corresponding resource pool R. The reactions shown in **Fig. 3.12a** are:

$$A_i + R \xrightarrow{k_i^+} C_i \tag{A.1}$$

$$C_i \xrightarrow{k_i^-} A_i + R \tag{A.2}$$

$$C_i \xrightarrow{k_i^{\text{cat}}} A_i + R + B_i \tag{A.3}$$

In these reactions, B_i is the product formed. The law of mass-action is applied to get the following ordinary differential equation for the complex species C_i:

$$\dot{C}_i(t) = k_i^+ R(t) A_i(t) - \left(k_i^- + k_i^{\text{cat}}\right) C_i(t) \tag{A.4}$$

Assuming a limited amount of available resources, the free resources can b expressed as:

$$R(t) = R^{\text{total}} - \sum_{j=1}^{n} C_j(t) \tag{A.5}$$

Since we would like to obtain a quasi-steady-state expression for any n, we cast the equations in matrix from:

$$\dot{\mathbf{C}} = \mathrm{diag}(\mathbf{k}^{+})(R^{\mathrm{total}} - \mathbf{1}^{T}\mathbf{C})\mathbf{A} - \mathrm{diag}(\mathbf{k}^{-} + \mathbf{k}^{\mathrm{cat}})\mathbf{C} \tag{A.6}$$

Here bold font is used to denote vectors and diag($\mathbf{x}$) to represent the matrix with the vector x as its diagonal. When we set the left-hand side to zero, lump the binding, unbinding and production rates into $\mathbf{k_m}$ such that $k_{m_i} := \frac{k_i^- + k_i^{\mathrm{cat}}}{k_i^+}$ and solve for $\mathbf{C}^*$ we get the expression:

$$\begin{aligned}
\mathbf{C}^* &= R^{\mathrm{total}}(I + \mathrm{diag}(\mathbf{k_m^{-1}})\mathbf{A}^*\mathbf{1}^{\mathrm{T}})^{-1}\mathrm{diag}(\mathbf{k_m^{-1}})\mathbf{A}^* \\
&= R^{\mathrm{total}}(I - \frac{\mathrm{diag}(\mathbf{k_m^{-1}})\mathbf{A}^*\mathbf{1}^{\mathrm{T}}}{1 + \mathbf{1}^{\mathrm{T}}\mathrm{diag}(\mathbf{k_m^{-1}})\mathbf{A}^*})\mathrm{diag}(\mathbf{k_m})\mathbf{A}^* \quad \text{(Sherman-Morrison formula)} \\
&= \frac{R^{\mathrm{total}}\mathrm{diag}(\mathbf{k_m^{-1}})\mathbf{A}^*}{1 + \mathbf{1}^{\mathrm{T}}\mathrm{diag}(\mathbf{k_m^{-1}})\mathbf{A}^*}
\end{aligned} \tag{A.7}$$

The effective production rate is obtained by multiplying this last expression with diag($\mathbf{k^{cat}}$). For an individual element this gives:

$$k_i^{\mathrm{eff}}(A_1, \ldots, A_n) := k_i^{\mathrm{cat}} R^{\mathrm{total}} \frac{k_{m_i}^{-1} A_i}{1 + \sum_{j=1}^{n} k_{m_j}^{-1} A_j} \tag{A.8}$$

A.2 Models and parameters for RNA-binding proteins and miR-31 characterization

A.2.1 RNA-binding proteins

L7Ae

The system of ordinary differential equations used to obtain the steady state expressions is:

$$\dot{M}_1 = k_{M_1}^{\mathrm{eff}}(G_1, G_2, G_3)G_1 - (\delta_{M_1} + \eta P_3)M_1 + \nu C_{M_1,P_3} \tag{A.9}$$

$$\dot{M}_2 = k_{M_2}^{\mathrm{eff}}(G_1, G_2, G_3)G_2 - \delta_{M_2}M_2 \tag{A.10}$$

$$\dot{M}_3 = k_{M_3}^{\mathrm{eff}}(G_1, G_2, G_3)G_3 - \delta_{M_3}M_3 \tag{A.11}$$

$$\dot{P}_1 = k_{P_1}^{\mathrm{eff}}(M_1, M_2, M_3)M_1 - \delta_{P_1}P_1 \tag{A.12}$$

$$\dot{P}_2 = k_{P_2}^{\mathrm{eff}}(M_1, M_2, M_3)M_2 - \delta_{P_2}P_2 \tag{A.13}$$

$$\dot{P}_3 = k_{P_3}^{\mathrm{eff}}(M_1, M_2, M_3)M_3 - \delta_{P_3}P_3 \tag{A.14}$$

$$\dot{C}_{M_1,P_3} = \eta M_1 P_3 - (\nu + \delta_{C_{M_1,P_3}}) C_{M_1,P_3} \tag{A.15}$$

In this set of equations, M_1 denotes the mRNA species of the *X-tra* protein P_1. The *capacity monitor* mRNA is represented by M_2 and its protein by P_2. The RBP L7Ae is denote by P_3 and its mRNA species is given by M_3.

$$P_1^* = \frac{\alpha_{P_1} \frac{\beta_{M_1}\gamma_{M_1}(1+\nu)}{(1+\gamma_{M_1}+\gamma_{M_2}+\gamma_{M_3})(1+\nu+\rho_{M_1}P_3^*)}}{1 + \frac{\beta_{M_1}\gamma_{M_1}(1+\nu)}{(1+\gamma_{M_1}+\gamma_{M_2}+\gamma_{M_3})(1+\nu+\rho_{M_1}P_3^*)} + \frac{\beta_{M_2}\gamma_{M_2}}{1+\gamma_{M_1}+\gamma_{M_2}+\gamma_{M_3}} + \frac{\beta_{M_3}\gamma_{M_3}}{1+\gamma_{M_1}+\gamma_{M_2}+\gamma_{M_3}}} \tag{A.16}$$

$$P_2^* = \frac{\alpha_{P_2} \frac{\beta_{M_2}\gamma_{M_2}}{1+\gamma_{M_1}+\gamma_{M_2}+\gamma_{M_3}}}{1 + \frac{\beta_{M_1}\gamma_{M_1}(1+\nu)}{(1+\gamma_{M_1}+\gamma_{M_2}+\gamma_{M_3})(1+\nu+\rho_{M_1}P_3^*)} + \frac{\beta_{M_2}\gamma_{M_2}}{1+\gamma_{M_1}+\gamma_{M_2}+\gamma_{M_3}} + \frac{\beta_{M_3}\gamma_{M_3}}{1+\gamma_{M_1}+\gamma_{M_2}+\gamma_{M_3}}} \tag{A.17}$$

$$P_3^* = \frac{\alpha_{P_3} \frac{\beta_{M_3}\gamma_{M_3}}{1+\gamma_{M_1}+\gamma_{M_2}+\gamma_{M_3}}}{1 + \frac{\beta_{M_1}\gamma_{M_1}(1+\nu)}{(1+\gamma_{M_1}+\gamma_{M_2}+\gamma_{M_3})(1+\nu+\rho_{M_1}P_3^*)} + \frac{\beta_{M_2}\gamma_{M_2}}{1+\gamma_{M_1}+\gamma_{M_2}+\gamma_{M_3}} + \frac{\beta_{M_3}\gamma_{M_3}}{1+\gamma_{M_1}+\gamma_{M_2}+\gamma_{M_3}}} \tag{A.18}$$

Here, is introduced a new lumped parameter $\rho_{M_1} := \frac{\eta}{\delta_{M_1}}$. Due to the complexity of the expressions, the system has not been fully solved for its steady state. To obtain the expressions used for fitting to the data, the steady state expression for P_3^* above was solved for P_3^* and plugged into the steady state expressions of P_1^* and P_2^*. Parameters fit is shown in Table A.1.

Table A.1: **Parameters fit for L7Ae experimental results (Fig. 3.12b).**

Parameter	**Unit**	**Value**
α_{P1}	$Arbitrary fluorecence$	41764.7
α_{P2}	$Arbitrary fluorecence$	2604.06
α_{M3}	$Arbitrary fluorecence$	16208
β_{M1}	$Unitless$	657.816
β_{M2}	$Unitless$	3508.74
β_{M3}	$Unitless$	386.641
γ_{M1}	$Unitless$	142.346
γ_{M2}	$Unitless$	23.082
γ_{M3}	$Unitless$	10.0254
ν	$Time^{-1}$	36932
ρ_{M1}	$Arbitrary fluorecence^{-1}$	1093.94

MS2-cNOT7

The system of ordinary differential equations used to obtain the steady state expressions is:

$$\dot{M}_1 = k^{\text{eff}}_{M_1}(G_1, G_2, G_3)G_1 - (\delta_{M_1} + \eta P_3)M_1 \tag{A.19}$$

$$\dot{M}_2 = k^{\text{eff}}_{M_2}(G_1, G_2, G_3)G_2 - \delta_{M_2}M_2 \tag{A.20}$$

$$\dot{M}_3 = k^{\text{eff}}_{M_3}(G_1, G_2, G_3)G_3 - \delta_{M_3}M_3 \tag{A.21}$$

$$\dot{P}_1 = k^{\text{eff}}_{P_1}(M_1, M_2, M_3)M_1 - \delta_{P_1}P_1 \tag{A.22}$$

$$\dot{P}_2 = k^{\text{eff}}_{P_2}(M_1, M_2, M_3)M_2 - \delta_{P_2}P_2 \tag{A.23}$$

$$\dot{P}_3 = k^{\text{eff}}_{P_3}(M_1, M_2, M_3)M_3 - \delta_{P_3}P_3 \tag{A.24}$$

Again, M_1 denotes the mRNA species of the *X-tra* protein P_1. The *capacity monitor* mRNA is represented by M_2 and its protein by P_2. The RBP MS2-cNOT7 is is captured in P_3 and its mRNA species is given by M_3. Solving the above system for steady state yields:

$$P_1^* = \frac{\alpha_{P_1}\frac{\beta_{M_1}\gamma_{M_1}}{(1+\gamma_{M_1}+\gamma_{M_2}+\gamma_{M_3})(1+\rho_{M_1}P_3^*)}}{1 + \frac{\beta_{M_1}\gamma_{M_1}}{(1+\gamma_{M_1}+\gamma_{M_2}+\gamma_{M_3})(1+\rho_{M_1}P_3^*)} + \frac{\beta_{M_2}\gamma_{M_2}}{1+\gamma_{M_1}+\gamma_{M_2}+\gamma_{M_3}} + \frac{\beta_{M_3}\gamma_{M_3}}{1+\gamma_{M_1}+\gamma_{M_2}+\gamma_{M_3}}} \tag{A.25}$$

$$P_2^* = \frac{\alpha_{P_2}\frac{\beta_{M_2}\gamma_{M_2}}{1+\gamma_{M_1}+\gamma_{M_2}+\gamma_{M_3}}}{1 + \frac{\beta_{M_1}\gamma_{M_1}}{(1+\gamma_{M_1}+\gamma_{M_2}+\gamma_{M_3})(1+\rho_{M_1}P_3^*)} + \frac{\beta_{M_2}\gamma_{M_2}}{1+\gamma_{M_1}+\gamma_{M_2}+\gamma_{M_3}} + \frac{\beta_{M_3}\gamma_{M_3}}{1+\gamma_{M_1}+\gamma_{M_2}+\gamma_{M_3}}} \tag{A.26}$$

$$P_3^* = \frac{\alpha_{P_3}\frac{\beta_{M_3}\gamma_{M_3}}{1+\gamma_{M_1}+\gamma_{M_2}+\gamma_{M_3}}}{1 + \frac{\beta_{M_1}\gamma_{M_1}}{(1+\gamma_{M_1}+\gamma_{M_2}+\gamma_{M_3})(1+\rho_{M_1}P_3^*)} + \frac{\beta_{M_2}\gamma_{M_2}}{1+\gamma_{M_1}+\gamma_{M_2}+\gamma_{M_3}} + \frac{\beta_{M_3}\gamma_{M_3}}{1+\gamma_{M_1}+\gamma_{M_2}+\gamma_{M_3}}} \tag{A.27}$$

As before, the expressions fit to the data are obtained by solving the equation for P_3^* for P_3^* and plugin the result into the other two expressions. Parameters fit is shown in Table A.2.

A.2.2 miR-31 characterization

The system of ordinary differential equations used to obtain the steady state expressions is:

$$\dot{M}_1 = k^{\text{eff}}_{M_1}(G_1, G_2, G_3)G_1 - (\delta_{M_1} + \eta m)M_1 \tag{A.28}$$

$$\dot{M}_2 = k^{\text{eff}}_{M_2}(G_1, G_2, G_3)G_2 - \delta_{M_2}M_2 \tag{A.29}$$

$$\dot{m} = k^{\text{eff}}_{m}(G_1, G_1, G_m)G_m - \delta_m m \tag{A.30}$$

Table A.2: **Parameters fit for MS2-cNOT7 experimental results (Fig. 3.12b).**

Parameter	**Unit**	**Value**
α_{P1}	$Arbitrary fluorecence$	19444.9
α_{P2}	$Arbitrary fluorecence$	14480
α_{M3}	$Arbitrary fluorecence$	1.77696
β_{M1}	$Unitless$	2115.61
β_{M2}	$Unitless$	858.958
β_{M3}	$Unitless$	74.2967
γ_{M1}	$Unitless$	41.1893
γ_{M2}	$Unitless$	104.793
γ_{M3}	$Unitless$	31.3533
ρ_{M1}	$Arbitrary fluorecence^{-1}$	213.586

$$\dot{P}_1 = k_{P_1}^{\text{eff}}(M_1, M_2, M_3)M_1 - \delta_{P_1}P_1 \tag{A.31}$$

$$\dot{P}_2 = k_{P_2}^{\text{eff}}(M_1, M_2, M_3)M_2 - \delta_{P_2}P_2 \tag{A.32}$$

In this system of equations, we denote the *miRNA sensor* mRNA by M_1 and the respective fluorescent protein output as P_1. The capacity monitor is expressed through the mRNA M_2 and the protein P_2. The miRNA itself is represented by m. Solving for steady state yields:

$$P_1^* = \frac{\alpha_{P_1}\frac{\beta_{M_1}\gamma_{M_1}}{1+\gamma_{M_1}+\gamma_{M_2}+\gamma_m(1+\theta_{M_1}\lambda\tau)}}{1+\frac{\beta_{M_1}\gamma_{M_1}}{1+\gamma_{M_1}+\gamma_{M_2}+\gamma_m}+\frac{\beta_{M_2}\gamma_{M_2}}{1+\gamma_{M_1}+\gamma_{M_2}+\gamma_m(1+\theta_{M_1}\lambda\tau)}} \tag{A.33}$$

$$P_2^* = \frac{\alpha_{P_2}\frac{\beta_{M_2}\gamma_{M_2}}{1+\gamma_{M_1}+\gamma_{M_2}+\gamma_m}}{1+\frac{\beta_{M_1}\gamma_{M_1}}{1+\gamma_{M_1}+\gamma_{M_2}+\gamma_m}+\frac{\beta_{M_2}\gamma_{M_2}}{1+\gamma_{M_1}+\gamma_{M_2}+\gamma_m(1+\theta_{M_1}\lambda\tau)}} \tag{A.34}$$

The additional lumping parameter was used: $\theta_{M_i} := \frac{\eta_{M_i}}{\delta_{M_i}}\frac{k_{M_2}^{\text{cat}}R_M^{\text{total}}}{\delta_m}$. To account for the placement of the miRNA targets in different UTRs and different numbers of targets we introduce the two parameters λ (UTR) and τ (target number). Specifically, $\lambda \in \{\lambda_{3'}, \lambda_{5'}\}$ and $\tau \in \{\tau_{1x}, \tau_{3x}\}$. Parameters fit is shown in Table A.3.

Table A.3: **Parameters fit for miR-31 characterization in H1299 (Fig. 3.12c).**

Parameter	Unit	Value
α_{P1}	$Arbitrary fluorecence$	433.658
α_{P2}	$Arbitrary fluorecence$	4855.31
β_{M1}	$Unitless$	992.56
β_{M2}	$Unitless$	895.094
γ_{M1}	$Unitless$	2232.66
γ_{M2}	$Unitless$	518.349
γ_{M2}	$Unitless$	2.56322
$\lambda_{3'}$	$Unitless$	155.969
$\lambda_{5'}$	$Unitless$	926.904
τ_{1x}	$Unitless$	3.0671
τ_{3x}	$Unitless$	70.1826

A.3 Model and parameters for endogenous microRNA-based iFFL circuit

The system of equations used to derive the steady state expressions is given by:

$$\dot{M}_1 = k^{\text{eff}}_{M_1}(G_1, G_2, G_3)G_1 - \delta_{M_1} M_1 \tag{A.35}$$

$$\dot{M}_2 = k^{\text{eff}}_{M_2}(G_1, G_2, G_m)G_2 - (\delta_{M_2} + \eta_{M_2} m)\, M_2 \tag{A.36}$$

$$\dot{m} = k^{\text{eff}}_{m}(G_1, G_2, G_m)G_m - \delta_m m \tag{A.37}$$

$$\dot{P}_1 = k^{\text{eff}}_{P_1}(M_1, M_2)M_1 - \delta_{P_1} P_1 \tag{A.38}$$

$$\dot{P}_2 = k^{\text{eff}}_{P_2}(M_1, M_2)M_2 - \delta_{P_2} P_2 \tag{A.39}$$

Here, M_1 and M_2 represent the mRNA species for the fluorescent proteins EGFP (X-tra) and mKate (GOI) respectively. P_1 and P_2 correspond to the proteins themselves and m denotes the microRNA miR-31. The degradation rates are shown as δ subscripted with the species they correspond to and the repression rates are shown as η subscripted with the respective species. G_1 and G_2 are the plasmid copy number of the EGFP and mKate plasmids respectively and G_m is the copy number of the microRNA on the genome. The steady states for the protein species can be written as:

$$P_1^* = \frac{\alpha_{P_1} \frac{\beta_{M_1} G_1 k_{m_{M_1}}^{-1}}{1+G_1 k_{m_{M_1}}^{-1} + \gamma_{M_2} + \gamma_m}}{1 + \frac{\beta_{M_1} G_1 k_{m_{M_1}}^{-1}}{1+G_1 k_{m_{M_1}}^{-1} + \gamma_{M_2} + \gamma_m} + \frac{\beta_{M_2}\gamma_{M_2}}{1+G_1 k_{m_{M_1}}^{-1} + \gamma_{M_2} + \gamma_m\left(1+\theta_{M_2}\right)}} \tag{A.40}$$

$$P_2^* = \frac{\alpha_{P_2} \frac{\beta_{M_2}\gamma_{M_2}}{1+G_1 k_{m_{M_1}}^{-1}+\gamma_{M_2}+\gamma_m\left(1+\theta_{M_2}\right)}}{1+\frac{\beta_{M_1}G_1 k_{m_{M_1}}^{-1}}{1+G_1 k_{m_{M_1}}^{-1}+\gamma_{M_2}+\gamma_m}+\frac{\beta_{M_2}\gamma_{M_2}}{1+G_1 k_{m_{M_1}}^{-1}+\gamma_{M_2}+\gamma_m\left(1+\theta_{M_2}\right)}} \tag{A.41}$$

For fitting, the expressions were normalized the same way as introduced in Appendix A.1, which yields the expressions:

$$\hat{P}_1^* = \frac{G_1\left(1+G_{1_0}k_{m_{M_1}}^{-1}+\gamma_{M_2}+\gamma_m\right)}{G_{1_0}\left(1+G_1 k_{m_{M_1}}^{-1}+\gamma_{M_2}+\gamma_m\right)} \frac{1+\frac{\beta_{M_1}G_{1_0}k_{m_{M_1}}^{-1}}{1+G_{1_0}k_{m_{M_1}}^{-1}+\gamma_{M_2}+\gamma_m}+\frac{\beta_{M_2}\gamma_{M_2}}{1+G_{1_0}k_{m_{M_1}}^{-1}+\gamma_{M_2}+\gamma_m\left(1+\theta_{M_2}\right)}}{1+\frac{\beta_{M_1}G_1 k_{m_{M_1}}^{-1}}{1+G_1 k_{m_{M_1}}^{-1}+\gamma_{M_2}+\gamma_m}+\frac{\beta_{M_2}\gamma_{M_2}}{1+G_1 k_{m_{M_1}}^{-1}+\gamma_{M_2}+\gamma_m\left(1+\theta_{M_2}\right)}} \tag{A.42}$$

$$\hat{P}_2^* = \frac{1+G_{1_0}k_{m_{M_1}}^{-1}+\gamma_{M_2}+\gamma_m\left(1+\theta_{M_2}\right)}{1+G_1 k_{m_{M_1}}^{-1}+\gamma_{M_2}+\gamma_m\left(1+\theta_{M_2}\right)} \frac{1+\frac{\beta_{M_1}G_{1_0}k_{m_{M_1}}^{-1}}{1+G_{1_0}k_{m_{M_1}}^{-1}+\gamma_{M_2}+\gamma_m}+\frac{\beta_{M_2}\gamma_{M_2}}{1+G_{1_0}k_{m_{M_1}}^{-1}+\gamma_{M_2}+\gamma_m\left(1+\theta_{M_2}\right)}}{1+\frac{\beta_{M_1}G_1 k_{m_{M_1}}^{-1}}{1+G_1 k_{m_{M_1}}^{-1}+\gamma_{M_2}+\gamma_m}+\frac{\beta_{M_2}\gamma_{M_2}}{1+G_1 k_{m_{M_1}}^{-1}+\gamma_{M_2}+\gamma_m\left(1+\theta_{M_2}\right)}} \tag{A.43}$$

The parameters used to fit the experimental data of the iFFL circuits in H1299, U2OS and HEK are listed in Tables A.4,A.5.

Table A.4: **Parameters fit for miR-31 iFFL circuit in H1299 (Fig. 3.13c).**

Parameter	**Unit**	**Value**
β_{M1}	$Unitless$	0.740656
β_{M2}	$Unitless$	32.429
$k_{m_{M_1}}$	$Unitless$	59.104
γ_{M1}	$Unitless$	0.480552
γ_m	$Unitless$	33.2046
θ_{M_2}	$Unitless$	227.144

Table A.5: **Parameters fit for miR-221 iFFL circuit in U2OS, left, and HEK, right (Fig. 3.14).**

Parameter	**Unit**	**Value**
β_{M1}	$Unitless$	1.01232e-9
β_{M2}	$Unitless$	0.94343
$k_{m_{M_1}}$	$Unitless$	3.57119
γ_{M1}	$Unitless$	2.85075e-11
γ_m	$Unitless$	0.173874
θ_{M_2}	$Unitless$	10.9904

Parameter	**Unit**	**Value**
β_{M1}	$Unitless$	2.81166
β_{M2}	$Unitless$	1.95439e-5
$k_{m_{M_1}}$	$Unitless$	2.32439
γ_{M1}	$Unitless$	1.1604e-13
γ_m	$Unitless$	6.65805
θ_{M_2}	$Unitless$	6.38791

Appendix B

Model of miRNA-dependent resource re-allocation

B.1 Model variables

B.1.1 Species

Free Pol II binding site (miRNA production pathway)	d_q
Free Pol II binding site (Protein production pathway)	d_j
Free ribosome binding site (RBS)	m_j
Protein species j	P_j
Free RNA Polymerase	Pol
Free Ribosomes	R
Free mRNA-degradation complex (XRN1/Exosome)	X
Transcription initiation complex (miRNA production pathway)	$C_{d,q}$
Transcription elongation complex (miRNA production pathway)	$N_{d,q}$
Transcription initiation complex (protein production pathway)	$C_{d,j}$
Transcription elongation complex (protein production pathway)	$N_{d,j}$
Translation initiation complex	$C_{m,j}$
Translation elongation complex	$N_{m,j}$
mRNA bound by mRNA-degradation complex	$Y_{m,j}$
miRNA 5'UTR	mi_5
miRNA 3'UTR	mi_3
mRNA blocked by miRNA 5'UTR	$L_{5,m,j}$
Ribosomes elongating on $L_{5,m,j}$	$K_{5,m,j}$
mRNA bound by miRNA 5'UTR and mRNA-degradation complex	$Y_{5,m,j}$
mRNA bound by miRNA 3'UTR	$L_{3,m,j}$
mRNA translation initiation complex bound by miRNA 3'UTR	$L_{3,C,j}$
Ribosomes elongating on $L_{3,m,j}$	$K_{3,m,j}$
Ribosomes upon $L_{3,C,j}$	$K_{3,C,j}$
mRNA bound by miRNA 3'UTR and mRNA-degradation complex	$Y_{3,m,j}$
mRNA translation initiation complex bound by miRNA 3'UTR and mRNA-degradation complex	$Y_{3,C,j}$

B.1.2 Rates

Pol II binding rate to d_q	$k^+_{d,q}$
Pol II unbinding rate from d_q	$k^-_{d,q}$
Pol II binding rate to d_i	$k^+_{d,j}$
Pol II unbinding rate from d_j	$k^-_{d,j}$
Transcription Initiation rate (miRNA production pathway)	$k_{d,q}$
Transcription Termination rate (miRNA production pathway)	$k^c_{d,q}$
Transcription Initiation rate (protein production pathway)	$k_{d,j}$
Transcription Termination rate (protein production pathway)	$k^c_{d,j}$
RBS:Ribosome Binding rate	$k^+_{m,j}$
RBS:Ribosome Unbinding rate	$k^-_{m,j}$
Translation Initiation rate	$k_{m,j}$
Translation Termination rate	$k^c_{m,j}$
mRNA-degradation complex:mRNA binding rate	$k_{X,m}$
Protein degradation rate	λ_P
miRNA degradation rate	λ_{mi}
miRNA binding rate	k_b
mRNA-degradation complex recruitment rate by miRNA 5'UTR	k_{X5}
mRNA-degradation complex recruitment rate by miRNA 3'UTR	k_{X3}
mRNA degradation assisted by mRNA-degradation complex rate	k_d
mRNA-degradation complex recycle rate	β

B.2 Model reactions

All the reactions follow mass-action kinetics, however species with a hat above them - eg. $\hat{x}$ - take part to the reactions without contributing to the rate dynamics. For example:

$$\mathrm{A} + \mathrm{B} + \widehat{\mathrm{C}} \xrightarrow{\mathrm{k}} \mathrm{D}, \quad \Longrightarrow \quad \frac{dC}{dt} = -kAB, \ \frac{dD}{dt} = kAB \tag{B.1}$$

If the species is scaled, it receives a scaled rate. For example:

$$\mathrm{A} + \mathrm{B} + \widehat{3\mathrm{C}} \xrightarrow{\mathrm{k}} \mathrm{D}, \quad \Longrightarrow \quad \frac{dC}{dt} = -3kAB, \ \frac{dD}{dt} = kAB \tag{B.2}$$

B.2.1 mRNA and protein production

For the species in this subsection $j \in \{r, g\}$, corresponding to mKate and EGFP respectively.

Transcription

Transcription is modelled in three steps:

- First, RNA Polymerases II (Pol) reversibly binds to a Pol II binding site (d_j) generating the transcription initiation complex ($C_{d,j}$).
- Second, transcription begins and the initiation complex ($C_{d,j}$) becomes a transcription elongation complex ($N_{d,j}$); the Pol II binding site (d_j) is released in the process.
- Third, the transcription elongation terminates and the RNA Polymerases II (Pol) and the mRNA exemplified by m_j are released.

$$\mathrm{d}_j + \mathrm{Pol} \underset{\mathrm{k}^-_{d,j}}{\overset{\mathrm{k}^+_{d,j}}{\rightleftharpoons}} \mathrm{C}_{d,j} \xrightarrow{\mathrm{k}_{d,j}} \mathrm{N}_{d,j} + \mathrm{d}_j \tag{B.3}$$

$$\mathrm{N}_{d,j} \xrightarrow{\mathrm{k}^c_{d,j}} \mathrm{Pol} + \mathrm{m}_j \tag{B.4}$$

Translation

Translation is being modelled in three steps:

- First, ribosomes (R) reversibly binds to a RBS (m_j) generating the translation initiation complex ($C_{m,j}$).
- Second, translation begins and the initiation complex ($C_{m,j}$) becomes a translation elongation complex ($N_{m,j}$); the RBS (m_j) is released in the process.
- Third, the translation elongation terminates and the ribosomes (R) and the proteins exemplified by p_j are released.

$$\mathrm{m}_j + \mathrm{R} \underset{\mathrm{k}^-_{m,j}}{\overset{\mathrm{k}^+_{m,j}}{\rightleftharpoons}} \mathrm{C}_{m,j} \xrightarrow{\mathrm{k}_{m,j}} \mathrm{N}_{m,j} + \mathrm{m}_j \tag{B.5}$$

$$\mathrm{N}_{m,j} \xrightarrow{\mathrm{k}^c_{m,j}} \mathrm{R} + \mathrm{P}_j \tag{B.6}$$

γ_j is the number of ribosomes in elongation phase:

$$\gamma_j = \frac{|N_j|}{|m_j + C_{m,j}|} \tag{B.7}$$

Degradation

mRNA unspecific degradation is catalyzed by the mRNA-degradation complex X:

$$\mathrm{m}_j + \mathrm{X} + \widehat{\gamma_\mathrm{j} \mathrm{N_j}} \xrightarrow{\mathrm{k_{Xm}}} \mathrm{Y_{mj}} \tag{B.8}$$

$$\mathrm{Y_{mj}} \xrightarrow{\mathrm{k_{dm}}} \mathrm{X} + \gamma_j \mathrm{R} \tag{B.9}$$

Protein unspecific degradation:

$$\mathrm{P}_j \xrightarrow{\lambda_{p,j}} \varnothing \tag{B.10}$$

B.2.2 miRNA production

For the species in this subsection $q \in \{3, 5\}$, corresponding to miRNA 3'UTR and miRNA 5'UTR respectively. miRNAs are produced in the same way as mRNA:

$$\mathrm{d}_q + \mathrm{Pol} \underset{\mathrm{k_d^-}}{\overset{\mathrm{k_d^+}}{\rightleftharpoons}} \mathrm{C}_{d,q} \xrightarrow{\mathrm{k}_d} \mathrm{N}_{d,q} + \mathrm{d}_q \tag{B.11}$$

$$\mathrm{N}_{d,q} \xrightarrow{\mathrm{k_d^c}} \mathrm{Pol} + \mathrm{mi}_q \tag{B.12}$$

miRNA unspecific degradation:

$$\mathrm{mi}_q \xrightarrow{\lambda_5} \varnothing \tag{B.13}$$

B.2.3 miRNA action

Both species of miRNA have the same mRNA binding rate (k_b), but different mRNA-degradation complex recruitment rate ($k_{X3},\ k_{X5}$). The mRNA-degradation complex X then exploits mRNA degradation always at k_d rate.

miRNA 5'UTR

When a miRNA binds to the 5'UTR of the target mRNA, it blocks its translation preventing further ribosomes from binding and producing a blocked mRNA ($L_{5,j}$). However, the ribosomes already bound to the mRNA can still complete translation. This applies to ribosomes in elongation phase $K_{5,m,j}$:

$$\mathrm{m}_j + \mathrm{mi}_5 + \widehat{\gamma_\mathrm{j} \mathrm{N_j}} \xrightarrow{\mathrm{k}_b} \mathrm{L}_{5,j} + \gamma_j \, \mathrm{K}_{5,m,j} \tag{B.14}$$

$$\mathrm{K}_{5,m,j} \xrightarrow{\mathrm{k}^c_{m,j}} \mathrm{R} + \mathrm{P}_j \tag{B.15}$$

It should be noted that we are using the rate $k^c_{m,j}$, which is the rate for a ribosomes to terminate translation (complete elongation), from start to finish. However, when the miRNA 5'UTR binds to the mRNA, ribosomes would be evenly distributed across the mRNA. Therefore, our model 'resets' all the currently bound ribosomes to the start of

elongation. Even if this trade-off of our current reaction schema should only minimally impact the qualitative trend of the species' dynamics, we aim at fixing it in the next future.

The mRNA-degradation complex X is then recruited to bind the blocked mRNA complex $L_{5,j}$ to form a degradation elongation complex $Y_{5,m,j}$. When X binds, translation no longer occurs:

$$\mathrm{L}_{5,j} + \mathrm{X} + \widehat{\zeta_{5,\mathrm{j}}\mathrm{K}_{5,\mathrm{m,j}}} \xrightarrow{\mathrm{k_{X5}}} \mathrm{Y}_{5,m,j} \tag{B.16}$$

Similar to the definition of γ_j, $\zeta_{q,j}$ is the number of ribosomes in elongation phase on j transcript bound by mi_q:

$$\zeta_{q,j} = \frac{|K_{q,j}|}{|L_{q,j}|} \tag{B.17}$$

Finally, the degradation complex $Y_{5,m,j}$ undergoes degradation. In this process mRNA-degradation resources X and ribosomes R are released, while mi_5 is generally degraded with little recycling (captured by $\frac{1}{\beta}$ scaling factor):

$$\mathrm{Y}_{5,m,j} \xrightarrow{\mathrm{k}_d} \mathrm{X} + \zeta_{5,j}\,\mathrm{R} \tag{B.18}$$

$$\mathrm{Y}_{5,m,j} \xrightarrow{\mathrm{k}_d/\beta} \mathrm{mi}_5 + \mathrm{X} + \zeta_{5,j}\,\mathrm{R} \tag{B.19}$$

miRNA 3'UTR

When a miRNA binds to the 3'UTR of its target, it triggers assisted degradation of the mRNA, regardless of whether the mRNA is beginning translation (C_i), or is free (m_i).

$$\mathrm{m}_j + \mathrm{mi}_3 + \widehat{\gamma_{\mathrm{j}}\mathrm{N}_{\mathrm{j}}} \xrightarrow{\mathrm{k}_b} \mathrm{L}_{3,m,j} + \gamma_j\,\mathrm{K}_{3,m,j} \tag{B.20}$$

$$\mathrm{C}_{m,j} + \mathrm{mi}_3 + \widehat{\gamma_{\mathrm{j}}\mathrm{N}_{\mathrm{j}}} \xrightarrow{\mathrm{k}_b} \mathrm{L}_{3,\mathrm{C,j}} + \gamma_j\,\mathrm{K}_{3,m,j} + \mathrm{K}_{3,\mathrm{C,j}} \tag{B.21}$$

When mi_3 binds to m_j, the mRNA is still able to undergo all the translation steps: initiation, elongation and termination. Translation initiation:

$$\mathrm{L}_{3,m,j} + \mathrm{R} \underset{\mathrm{k}^-_{m,j}}{\overset{\mathrm{k}^+_{m,j}}{\rightleftharpoons}} \mathrm{L}_{3,\mathrm{C,j}} + \mathrm{K}_{3,\mathrm{C,j}} \tag{B.22}$$

Translation elongation:

$$\mathrm{L}_{3,\mathrm{C,j}} + \widehat{\mathrm{K}_{3,\mathrm{C,j}}} \xrightarrow{\mathrm{k}_{m,j}} \mathrm{L}_{3,m,j} + \mathrm{K}_{3,m,j} \tag{B.23}$$

Translation termination:

$$\mathrm{K}_{3,m,j} \xrightarrow{\mathrm{k}^{c}_{m,j}} \mathrm{R} + \mathrm{P}_{j} \tag{B.24}$$

Once mi_3 has bound to m_j or C_j, it recruits the mRNA-degradation complex, forming a complex that cannot undergo translation anymore:

$$\mathrm{L}_{3,m,j} + \mathrm{X} + \widehat{\zeta_{3,\mathrm{j}}\mathrm{K}_{3,\mathrm{m,j}}} \xrightarrow{\mathrm{k_{X3}}} \mathrm{Y}_{3,m,j} \tag{B.25}$$

$$\mathrm{L}_{3,\mathrm{C,j}} + \mathrm{X} + \widehat{\zeta_{3,\mathrm{j}}\mathrm{K}_{3,\mathrm{m,j}}} + \widehat{\mathrm{K}_{3,\mathrm{C,j}}} \xrightarrow{\mathrm{k_{X3}}} \mathrm{Y}_{3,\mathrm{C,j}} \tag{B.26}$$

Finally, miRNA-dependent mRNA degradation begins. In this process mRNA-degradation resources X and ribosomes R are released, while mi_5 is generally degraded with little recycling (captured by $\frac{1}{\beta}$ scaling factor):

$$\mathrm{Y}_{3,m,j} \xrightarrow{\mathrm{k}_d} \zeta_{3,j}\,\mathrm{R} + \mathrm{X} \tag{B.27}$$

$$\mathrm{Y}_{3,m,j} \xrightarrow{\mathrm{k}_d/\beta} \zeta_{3,j}\,\mathrm{R} + \mathrm{mi}_3 + \mathrm{X} \tag{B.28}$$

$$\mathrm{Y}_{3,\mathrm{C,j}} \xrightarrow{\mathrm{k}_d} (\zeta_{3,j} + 1)\,\mathrm{R} + \mathrm{X} \tag{B.29}$$

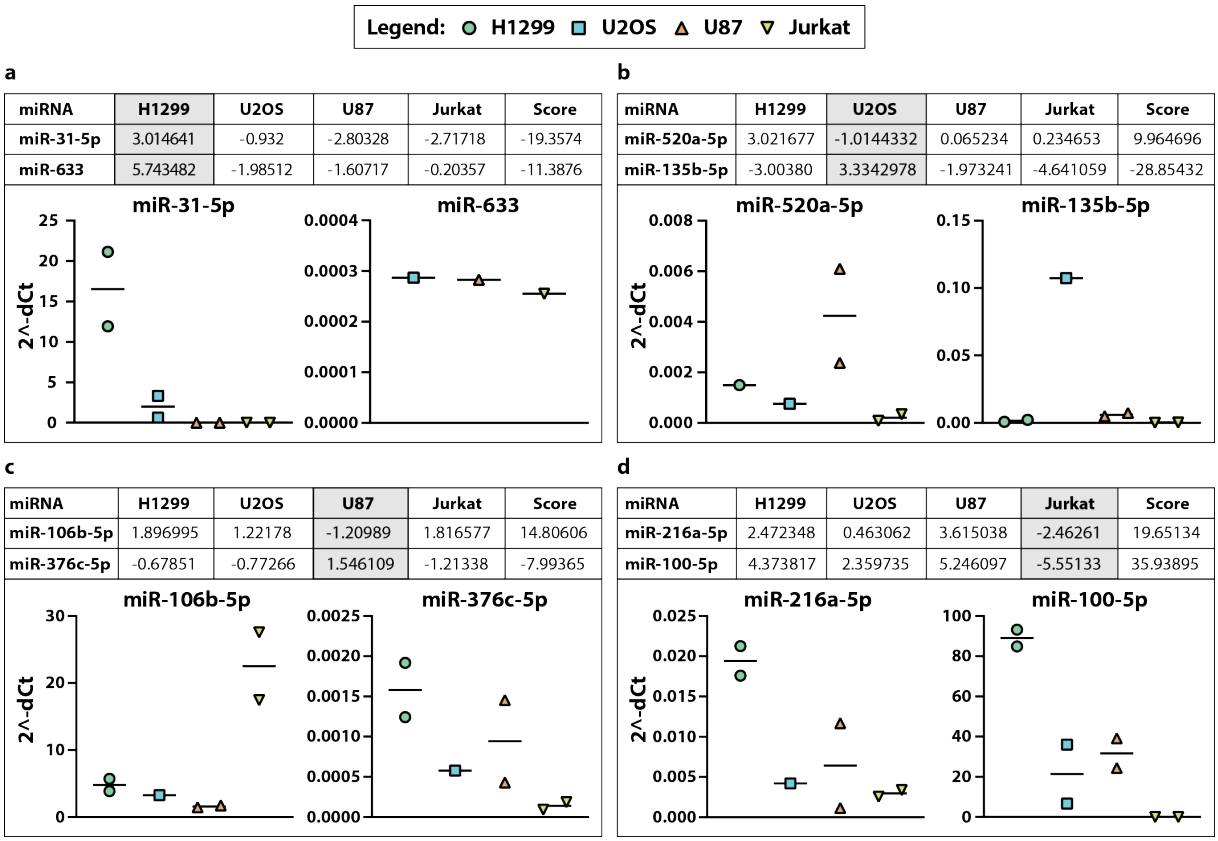
$$\mathrm{Y}_{3,\mathrm{C,j}} \xrightarrow{\mathrm{k}_d/\beta} (\zeta_{3,j} + 1)\,\mathrm{R} + \mathrm{mi}_3 + \mathrm{X} \tag{B.30}$$

B.2.4 Reactions summary

mRNA	Production	$\mathrm{d}_j + \mathrm{Pol} \underset{\mathrm{k}^-_{d,j}}{\overset{\mathrm{k}^+_{d,j}}{\rightleftharpoons}} \mathrm{C}_{d,j} \xrightarrow{\mathrm{k}_{d,j}} \mathrm{N}_{d,j} + \mathrm{d}_j$ $\mathrm{N}_{d,j} \xrightarrow{\mathrm{k}^c_{d,j}} \mathrm{Pol} + \mathrm{m}_j$
	Degradation	$\mathrm{m}_j + \mathrm{X} + \widehat{\gamma_\mathrm{j} \mathrm{N_j}} \xrightarrow{\mathrm{k_{Xm}}} \mathrm{Y_{mj}} \xrightarrow{\mathrm{k_{dm}}} \mathrm{X} + \gamma_j \, \mathrm{R}$
Protein	Production	$\mathrm{m}_j + \mathrm{R} \underset{\mathrm{k}^-_{m,j}}{\overset{\mathrm{k}^+_{m,j}}{\rightleftharpoons}} \mathrm{C}_{m,j} \xrightarrow{\mathrm{k}_{m,j}} \mathrm{N}_{m,j} + \mathrm{m}_j$ $\mathrm{N}_{m,j} \xrightarrow{\mathrm{k}^c_{m,j}} \mathrm{R} + \mathrm{P}_j$
	Number of ribosomes bound to mRNA	$\gamma_j = \frac{\lvert N_j \rvert}{\lvert m_j + C_{m,j} \rvert}$
	Degradation	$\mathrm{P}_j \xrightarrow{\lambda_\mathrm{P}} \varnothing$
miRNA	Production	$\mathrm{d}_q + \mathrm{Pol} \underset{\mathrm{k}^-_{d}}{\overset{\mathrm{k}^+_{d}}{\rightleftharpoons}} \mathrm{C}_{d} \xrightarrow{\mathrm{k}_{d}} \mathrm{N}_{d,q} + \mathrm{d}_q$ $\mathrm{N}_{d,q} \xrightarrow{\mathrm{k}^c_{d}} \mathrm{Pol} + \mathrm{mi}_q$
	Degradation	$\mathrm{mi}_q \xrightarrow{\lambda_\mathrm{mi}} \varnothing$
	Number of ribosomes bound to mRNA bound by mi_q	$\zeta_{q,j} = \frac{\lvert K_{q,j} \rvert}{\lvert L_{q,j} \rvert}$
miRNA 5'	Binding to mRNA	$\mathrm{m}_j + \mathrm{mi}_5 + \widehat{\gamma_\mathrm{j} \mathrm{N_j}} \xrightarrow{\mathrm{k}_b} \mathrm{L}_{5,m,j} + \gamma_j \, \mathrm{K}_{5,m,j}$
	Blocked mRNA translation completion	$\mathrm{K}_{5,m,j} \xrightarrow{\mathrm{k}^c_{m,j}} \mathrm{R} + \mathrm{P}_j$
	mRNA-degradation complex recruitment	$\mathrm{L}_{5,m,j} + \mathrm{X} + \zeta_{5,\mathrm{j}} \widehat{\mathrm{K}_{5,\mathrm{m,j}}} \xrightarrow{\mathrm{k_{X5}}} \mathrm{Y}_{5,m,j}$
	Blocked mRNA degradation (non-recycle)	$\mathrm{Y}_{5,m,j} \xrightarrow{\mathrm{k}_d} \mathrm{X} + \zeta_{5,j}$
	Blocked mRNA degradation (recycle)	$\mathrm{Y}_{5,m,j} \xrightarrow{\mathrm{k}_d/\beta} \mathrm{mi}_5 + \mathrm{X} + \zeta_{5,j}$
miRNA 3'	Binding to free mRNA	$\mathrm{m}_j + \mathrm{mi}_3 + \widehat{\gamma_\mathrm{j} \mathrm{N_j}} \xrightarrow{\mathrm{k}_b} \mathrm{L}_{3,m,j} + \gamma_j \, \mathrm{K}_{3,m,j}$
	Binding to mRNA initiating translation	$\mathrm{C}_{m,j} + \mathrm{mi}_3 + \widehat{\gamma_\mathrm{j} \mathrm{N_j}} \xrightarrow{\mathrm{k}_b} \mathrm{L}_{3,\mathrm{C,j}}{}^{+} \, \gamma_j \, \mathrm{K}_{3,m,j} + \mathrm{K}_{3,\mathrm{C,j}}$
	Bound mRNA translation initiation	$\mathrm{L}_{3,m,j} + \mathrm{R} \underset{\mathrm{k}^-_{m,j}}{\overset{\mathrm{k}^+_{m,j}}{\rightleftharpoons}} \mathrm{L}_{3,\mathrm{C,j}} + \mathrm{K}_{3,\mathrm{C,j}}$
	Bound mRNA translation elongation	$\mathrm{L}_{3,\mathrm{C,j}} + \widehat{\mathrm{K}_{3,\mathrm{C,j}}} \xrightarrow{\mathrm{k}_{m,j}} \mathrm{L}_{3,m,j} + \mathrm{K}_{3,m,j}$
	Bound mRNA translation termination	$\mathrm{K}_{3,m,j} \xrightarrow{\mathrm{k}^c_{m,j}} \mathrm{R} + \mathrm{P}_j$
	mRNA-degradation complex recruitment on bound mRNA in translation elongation	$\mathrm{L}_{3,m,j} + \mathrm{X} + \zeta_{3,\mathrm{j}} \widehat{\mathrm{K}_{3,\mathrm{m,j}}} \xrightarrow{\mathrm{k_{X3}}} \mathrm{Y}_{3,m,j}$
	Bound mRNA exoribonuclease recruitment on bound mRNA in translation initiation	$\mathrm{L}_{3,\mathrm{C,j}} + \mathrm{X} + \zeta_{3,\mathrm{j}} \widehat{\mathrm{K}_{3,\mathrm{m,j}}} + \widehat{\mathrm{K}_{3,\mathrm{C,j}}} \xrightarrow{\mathrm{k_{X3}}} \mathrm{Y}_{3,\mathrm{C,j}}$
	Blocked mRNA degradation (non-recycle)	$\mathrm{Y}_{3,m,j} \xrightarrow{\mathrm{k}_d} \zeta_{3,j} \, \mathrm{R} + \mathrm{X}$
	Blocked mRNA degradation (recycle)	$\mathrm{Y}_{3,m,j} \xrightarrow{\mathrm{k}_d/\beta} \zeta_{3,j} \, \mathrm{R} + \mathrm{mi}_3 + \mathrm{X}$
	Degradation of blocked mRNA initiating translation (non-recycle)	$\mathrm{Y}_{3,\mathrm{C,j}} \xrightarrow{\mathrm{k}_d} (1 + \zeta_{3,j}) \, \mathrm{R} + \mathrm{X}$
	BDegradation of blocked mRNA initiating translation (recycle)	$\mathrm{Y}_{3,\mathrm{C,j}} \xrightarrow{\mathrm{k}_d/\beta} (1 + \zeta_{3,j}) \mathrm{R} + \mathrm{mi}_3 + \mathrm{X}$

Appendix C

miRNA quantification using SNORD72 as internal reference



a					
miRNA	H1299	U2OS	U87	Jurkat	Score
miR-31-5p	3.014641	-0.932	-2.80328	-2.71718	-19.3574
miR-633	5.743482	-1.98512	-1.60717	-0.20357	-11.3876

b					
miRNA	H1299	U2OS	U87	Jurkat	Score
miR-520a-5p	3.021677	-1.0144332	0.065234	0.234653	9.964696
miR-135b-5p	-3.00380	3.3342978	-1.973241	-4.641059	-28.85432

c					
miRNA	H1299	U2OS	U87	Jurkat	Score
miR-106b-5p	1.896995	1.22178	-1.20989	1.816577	14.80606
miR-376c-5p	-0.67851	-0.77266	1.546109	-1.21338	-7.99365

d					
miRNA	H1299	U2OS	U87	Jurkat	Score
miR-216a-5p	2.472348	0.463062	3.615038	-2.46261	19.65134
miR-100-5p	4.373817	2.359735	5.246097	-5.55133	35.93895

Figure C.1: **miRNA quantification using SNORD72 as internal reference.** Each subfigure shows the selected miRNAs and their z-score with respect to the mean and the variance among the 4 cell lines 4.2.1. In grey is highlighted the cell line in which the displayed miRNAs are differentially expressed (top). Bottom right and left, the corresponding quantification by qPCR relative to the internal reference SNORD72, round green symbols correspond to H1299 data, blue squares to U2OS, upwards orange triangles to U87 and downwards yellow triangles to Jurkat. To compare with the quantification relative to SNORD72 see Appendix C. **(a,b,c,d)** selected miRNAs differentially expressed in H1299, U2OS, U87 and Jurkat respectively. Data represent the mean and single values for N = 2 biological replicates. Missing data points correspond to undetected values due to too little abundance of miRNA in the sample.

Appendix D

Sequence of miRNAs and Target Sites used in Chapter 4

Table D.1: **List of the miRNAs and the TS sequences used in the study.**

miRNA	**miRNA Sequence**	**TS sequence**
miR-100-5p	aacccguagauccgaacuugug	cacaaguucggaucuacgggguu
miR-106b-5p	uaaagugcugacagugcagau	aucugcacugucagcacuuua
miR-135b-5p	uauggcuuuucauuccuauguga	ucacauaggaaugaaaagccaua
miR-1973	accgugcaaagguagcaua	uaugcuaccuuugcacggu
miR-199a-3p + miR-199b-3p	acaguagucugcacauugguua + acaguagucugcacauugguua	uaaccaaugugcagacuacugu + uaaccaaugugcagacuacugu
miR-216a-5p	ucacaguggucucugggauuau	auaaucccagagaccacuguga
miR-223-5p	cguguauuugacaagcugaguu	aacucagcuugucaaauacacg
miR-29b-1-5p	gcugguuucauaugguggguuuaga	gcugguuucauaugguggguuuaga
miR-301a-5p	aguagugcaauaaagucagagc	aguagugcaauaaagucagagc
miR-31-3p	auggcaauauguuggcauagca	auggcaauauguuggcauagca
miR-31-5p	aggcaagaugcuggcauagcu	agcuaugccagcaucuugccu
miR-335-5p	ucaagagcaauaacgaaaaaugu	acauuuuucguuauugcucuuga
miR-340-5p	uuauaaagcaaugagacugauu	aaucagucucauugcuuuauaa
miR-376c-5p	gguggauauuccuucuauguu	aacauagaaggaauauccacc
miR-520a-5p	cuccagagggaaguacuuucu	agaaaguacuucccucuggag
miR-522-5p	cucuagagggaagcgcuuucug	cagaaagcgcuucccucuagag
miR-630	aguauucuguaccagggaaggu	accuucccugguacagaauacu
miR-633	cuaauaguaucuaccacaauaaa	uuuauugugguagauacuauuag
miR-744-5p	ugcggggcuagggcuaacagca	ugcuguuagcccuagccccgca
miR-7b-5p	cuauacaaccuacugccuuccc	gggaaggcaguagguuguauag

Appendix E

Publications

2021

Bonfá G., **Cella F.**, Siciliano V. (2021). Engineering Protein-Based Parts for Genetic Devices in Mammalian Cells. In: Menolascina F. (eds) Synthetic Gene Circuits. Methods in Molecular Biology, vol 2229. Humana, New York, NY.

2020

Frei, T.*, **Cella F.***, Tedeschi, F., Gutierrez, J., Stan, GB., Khammash, M., Siciliano, V. (2020). Characterization and mitigation of gene expression burden in mammalian cells. Nature Communications, 11.

∗ These authors contributed equally to this work

2019

Cella F., Siciliano, V. (2019). Protein-based parts and devices that respond to intracellular and extracellular signals in mammalian cells. Current Opinion in Chemical Biology, (52).

2018

Cella F., Wroblewska, L., Weiss, R., Siciliano, V. (2018). Engineering protein-protein devices for multilayered regulation of mRNA translation using orthogonal proteases in mammalian cells. Nature Communications, 9(1).