Resource-aware mammalian cell engineering

Roberto Di Blasi

Thesis submitted for the degree of Doctor of Philosophy

Supervised by Dr Francesca Ceroni

Imperial College London

Department of Chemical Engineering

The Imperial College Centre for Synthetic Biology

July 2023

Abstract

Resource competition within engineered cells is a significant source of unpredictability in cell engineering, causing altered cellular physiology and unexpected behaviour of genetic circuits. While we now have a clear picture of how resource competition impacts microbial physiology, together with a plethora of mitigation strategies to address these issues, our understanding of how resource competition comes into play in engineered mammalian systems is currently limited.

Here we present our contribution to advancing the current understanding of resource competition in mammalian systems. Using a synthetic biology approach, we investigated resource allocation profiles in engineered mammalian cells. We identified resource bottlenecks in key cellular pathways, developed tools for appropriate quantification of cellular stress in resource-depleted engineered cells, and proposed resource-informed construct design as a new mitigation strategy. Whenever possible, we showed how our work can be useful to support cell engineering applications and correct interpretation of foundational biology experiments.

Statement of originality

I hereby confirm that everything included in this thesis was the outcome of my own work. Any results obtained in collaboration have been highlighted and acknowledged.

Copyright statement

Acknowledgements

I want to take this space to thank everyone who supported me during this PhD. It has not been an easy journey, and I am grateful for your help in achieving this huge milestone.

First and foremost, I would like to thank my parents, who always pushed me to follow my aspirations. I appreciate all you have done to get me where I am today, even more so knowing you had limited means. Your words of encouragement warmed my heart, and your advice kept me grounded.

I want to thank Francesca for believing in me after our first-ever video call. I am grateful for all the efforts you put into helping me develop as a researcher, the time you spent discussing science with me, and the support you gave me when every experiment went wrong. You have been approachable, kind, understanding, and a great motivator. You taught me a lot, and I am forever grateful for this.

Thank you to all the lab people for making failed experiments less grim. To my dear friends Anna and Carolina, a special thank you. I cherish the countless hours we three spent together. You broke the routine of lab work and made each day very enjoyable. Sharing successes and failures with you has made all the difference.

Thanks to all the researchers and collaborators that helped me through this journey. A special thanks to Jacopo, Masue, and Simone for their help with some of the experiments presented in this thesis. Thanks to Karen for her advice and invaluable support in my next career steps. A heartfelt thank you also to all the students I had the pleasure of supervising: Owen, Sophia, Ahmed, and Henry. I appreciate the effort you put into your projects; working with you was truly one of the most enjoyable experiences of this PhD.

To my friends in London, thank you for enriching my time outside the lab. You kept me balanced and made London my home. A special thank you to Sabrina for being such a great friend, always open to chatting about our PhD experiences and sharing words of support.

Finally, and not for importance, thanks to my partner Giovanni. Your positivity has lightened up every day and put failures in perspective. I am extremely proud of what we achieved together during our time in London: sticking together during a pandemic, parenting two kitties, starting up your new business, and now this PhD sound like a lot to me! I have not felt alone a single day with you by my side, and I cannot wait to see what we will achieve in the coming years. You made this journey incredibly easier, and I am immensely grateful to you for that.

Table of contents

List of figures

List of tables

List of abbreviations

mRNA: messenger RNA; **dox**: doxycycline; **TU**: transcriptional unit; **RFU**: relative fluorescent units; **MFE**: minimum free energy. **std:** standard deviation. **ER**: endoplasmic reticulum. **UPS**: ubiquitin-proteasome system. **UPR**: untranslated protein response.

Chapter 1 - Introduction

1.1. Cell engineering

Gene expression is a fundamental process by which cells read the information stored in their genome and use it to perform a cellular function. During this process, information in the form of DNA is transcribed into messenger RNA (mRNA) and translated into proteins, the key effectors of cellular behaviour. Cell engineering exploits this physiological process and, by delivering heterologous synthetic information intracellularly, programs cells to fulfil useful-to-human functions and behaviours. The type of information varies greatly, from simple genes to genetic circuits, to entire metabolic pathways. In complex applications, cells can be engineered to sense multiple cues, compute logical operations, and produce appropriate responses.

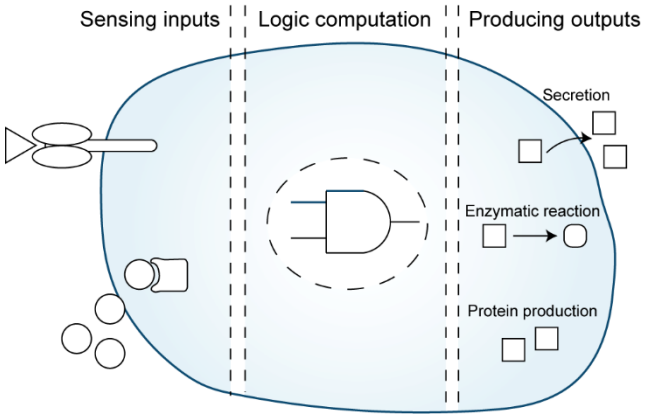

Figure 1.1 Engineering complex cellular behaviours. Cells can be engineered to detect specific inputs, process them through logical operations, and produce regulated outputs. Image adapted from Kitada *et al.*[1]

In the last two decades, cell engineering has pioneered the development of breakthrough technologies that transformed society, delivering on many of the 17 UN global sustainable development goals[2]. Some major achievements are outlined below, including practical applications and paradigm-shifting frameworks.

1.1.1. Engineering cells as therapeutics

Cells can be engineered to detect and treat disease, often providing effective therapeutic options[3–5]. For instance, much attention has been paid to the engineering of T-cells towards increased recognition and clearance of cancer cells (CAR-T)[6]. With approved therapies available to treat aggressive blood tumours and numerous clinical trials underway[7–9], CAR-T therapies represent one of the most successful achievements of cell engineering. Many other ongoing efforts need mentioning, such as the treatment of autoimmune diseases[10,11] and the first FDA-approved cell-based gene therapy to treat Beta-thalassemia[12].

Cell engineers are also working to deliver on diseases that are not yet treatable. One promising area is the engineering of induced pluripotent stem cells (iPSCs)[13]. iPSCs are patient-derived cells that are de-differentiated to a pluripotent state by expressing specific pluripotency factors. Once pluripotent, these

cells can be differentiated to obtain virtually any cell type of the body, providing a great asset to replace diseased tissues. Clinical trials are underway to use iPSCs to treat many forms of disease, including diabetes, Parkinson's disease, heart failure, and osteoarthritis, amongst others[14].

In addition to treatment, engineered cells can also function as theranostics; that is, they can be programmed to detect the onset of a disease-causing condition and respond with appropriate mitigating measures[15]. For instance, Choi *et al*. engineered cells to detect inflammatory cytokines and trigger an anti-inflammatory therapeutic response[16]. When implanted in a murine model of inflammatory arthritis, they observed reduced severity of inflammatory symptoms. Similar cell-based technologies to treat inflammation, immune response, and diabetes have been developed[17–22].

Cell engineering to detect and treat disease is an active area of research, and with the development of better delivery methods and implementation of increasingly complex genetic circuits, it promises to deliver technologies that will drastically increase our space for therapeutic intervention.

1.1.2. Engineering cells as production factories

Cells can be engineered as production factories for high-value molecules. Today, a plethora of molecules is produced by engineered cells, resulting in increased availability and lowered market costs[23]. Microorganism-based production platforms are used to produce bulk and fine chemicals, fuels, and natural products that involve inefficient chemical synthesis or are difficult to extract from natural sources. Refactoring in microbial hosts biosynthetic pathways typical of other organisms provides a cheaper, scalable, and more sustainable alternative for the mass production of chemicals. By employing this approach, many natural products of therapeutic value have been bio-synthesised, including cannabinoids and drugs to treat malaria, cancer, and neurodegenerative diseases[24–29].

Although microorganisms are also employed for recombinant production of commercially-valuable proteins, they are not amenable to producing animal and human proteins, as they cannot reproduce correct folding and post-translational modifications for which mammalian hosts are preferred. Mammalian-based production processes mainly focus on synthesising therapeutic proteins, including antibodies, growth hormones, clotting factors, and cytokines[30,31]. The demand for these high-value proteins is increasing, with the monoclonal antibodies alone accounting for a projected market size of USD 300 billion by 2025[32] and production pipelines widespread worldwide.

1.1.3. Engineering cells as true engines

Following initial landmark achievements in cell engineering, in the early 2000s it became apparent that complex cellular behaviour could be engineered using bottom-up approaches in which molecular components are rationally assembled into higher-order functional modules. This realisation effectively initiated synthetic biology as a new field of study focusing on advancing our engineering capabilities[33]. The

starting point is to treat biology as an engineering discipline. This means applying key engineering frameworks to biology, leading to a deeper comprehension of biological problems and improving the outcome of biological experiments[34]. Typical engineering concepts such as standardisation, modularity, orthogonality, predictability, reliability, and abstraction have permeated synthetic biology to the point that they started to modify how we design, conduct, and describe biological experiments.

When pinpointing when the field originated, two seminal works need to be mentioned: the development of the first biological oscillator[35] – the repressilator - and the engineering of the first bistable gene network[36] – the toggle switch. Both these works introduced rational design using basic biological components as the first step in construct assembly. This typical engineering approach has been at the core of synthetic biology: understanding complex systems by breaking them down into minimal systems with defined outputs.

Over the last two decades, the synthetic biology community has witnessed groundbreaking research and advanced both our understanding of biological systems and our ability to equip them with increasingly complex functions[37]. Engagement with the public and policymakers has led to synthetic biology impacting society for the better. Meaningful scientific interactions among professionals of different fields brought synthetic biology to maturity and laid the foundations for today's multibillion-dollar synthetic biology industry[37]. Despite the tremendous successes, the field has also uncovered a great deal of biological complexity that we still struggle to overcome. Parts are defined only under specific conditions; functional modules do not behave as expected when implemented in bigger circuits; variability dooms most of our experiments to failure[38]. Until we master this complexity, a scenario where synthetic biology-inspired cell engineering scales to deliver transformative technologies at pace may be unfeasible.

However, hitting roadblocks is an essential part of the learning curve that will one day allow us to engineer cells predictably and reliably, finally departing from the many iterations of a trial-and-error process. Untangling biological complexity is an active area of research within synthetic biology. Below is a detailed discussion of one of the key sources of unpredictability in cell engineering that has been the focus of my PhD studies.

1.2. Challenges in cell engineering

Although synthetic biology delivered many landmark achievements[37], work is still needed to tame the many sources of unpredictability of biological systems. The major challenges we face when engineering biology are poor standardisation of biological parts[38,39], noise[40,41], and unwanted interactions between the engineered cellular host and the genetic circuits[42]. While all of these are important, in this thesis I will discuss only host-construct interactions to present the background of my work clearly.

1.2.1. Host-construct interactions as a cause of unpredictability

Unwanted interactions between cellular hosts and genetic constructs have been identified as a major source of unpredictability, often leading to extensive preliminary testing before achieving a positive experimental outcome[42]. These interactions are bidirectional, including interactions that the cellular host imposes on genetic constructs and interactions that the genetic constructs impose on the cellular host.

1.2.1.1. Host-to-construct interactions

The cellular host significantly influences the functionality of genetic constructs, representing one unresolved cause of variability. The performance of biological parts, genetic constructs, and functional circuits greatly varies upon changes in the cellular host background. For instance, it is well known that even for standardised biological parts, that is, for those parts whose qualitative and quantitative behaviour has been assessed and reported on, performance varies in different microbial strains or culture conditions[43,44]. Similar reports exist in mammalian cells, where the functionality of promoters, transcriptional termination sequences, and regulatory proteins have been shown to be dependent on the cell type[45–48]. Initial attempts to manage this complexity proposed the development of improved standardisation frameworks. In a seminal paper, Kelly *et. al* measured the performance of a panel of promoters in different strains and growth conditions and normalised the expression levels of each promoter to a reference promoter. This resulted in a relative promoter activity measurement with decreased variance compared to raw promoter activity[43]. In a later work, Keren *et al.* screened a library of ~900 *Saccharomyces cerevisiae* promoters and ~1800 *Escherichia coli* promoters, showing that the activity of 60-90% of the promoters was scaled by a constant factor depending on the growth conditions[49]. However, these methods have been focused mostly on the analysis of promoter sequences in microbes and, more importantly, it was argued that they fail to capture global variations[50]. Context dependencies within the same cellular host under controlled growth conditions add complexity to this scenario. For instance, it is well known that the performance of integrative constructs changes dramatically based on the integration site in the genome[39]. This has been observed in bacteria[51,52] and has even bigger implications for eukaryotes, where examples of epigenetic silencing of integrated constructs have been extensively reported[53–55]. Finally, when biological parts are assembled to form functional genetic circuits, suboptimal behaviours can arise, leading, in some cases, to complete circuit failure. This was already evidenced in an early synthetic biology study, showing failure of a genetic logic gate when transferred to different strains of *E. coli*[56], and was later supported by more recent studies highlighting similar conclusions[57–59].

Understanding the causes behind this variability is crucial to enable predictive and reliable cell engineering. A seminal study in *E. coli* reported on how growth dependencies of key gene expression parameters - such as transcription rate, gene copy number, cell volume and protein dilution rate - can lead to qualitative and

quantitative changes in genetic circuits' behaviour[60]. This highlights the impact of biological complexity on the key synthetic biology's core principles of standardisation, modularity, predictability, and reliability.

1.2.1.2. Construct-to-host interactions

Complementarily, genetic constructs also exert an impact on the cellular host. Heterologous expression places an unnecessary load on the cells, causing the sequestration of essential cellular resources by genetic constructs. As a result of heterologous expression, the pool of cellular resources is depleted, and fewer resources are available to sustain physiological cellular functions. This condition of competition for shared cellular resources ultimately leads to unanticipated effects, which will be thoroughly discussed in the following section.

1.3. Resource competition in engineered cells

Resource competition arises from the simultaneous use of cellular resources by different expression cassettes. Any cellular resources can be the subject of competition, including DNA replication resources, transcriptional and translational resources, post-translational resources, RNA and protein degradation resources, and secretion resources. In addition, resource competition can arise between the cellular host and genetic constructs or amongst different genetic constructs introduced in the same cellular host.

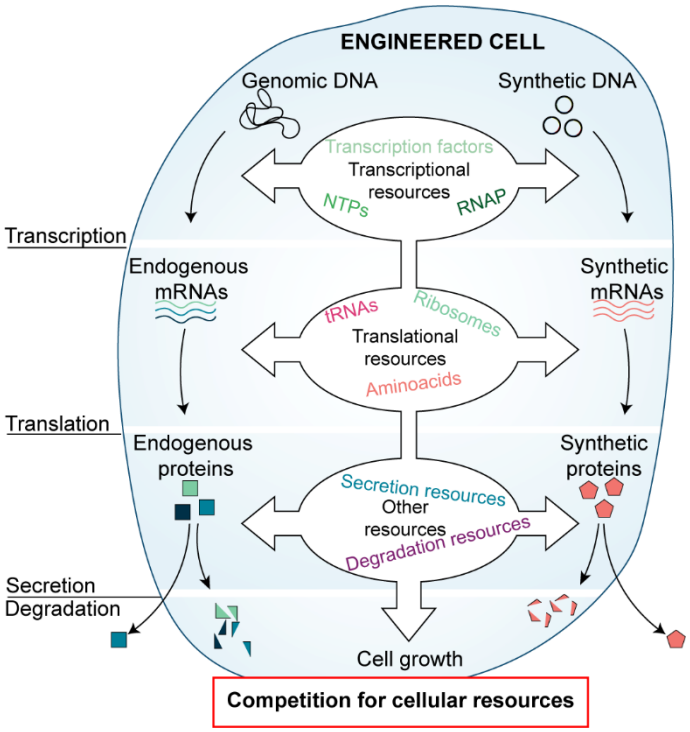

Figure 1.2 Resource competition in engineered cells. Resource competition arises from the simultaneous use of cellular resources by different expression cassettes. Any cellular resources can be the object of competition, including gene expression and post-translational resources. Heterologous expression of synthetic DNA depletes the limited pool of cellular resources, causing alteration of endogenous gene expression profiles and decreased cellular growth.

Bottlenecks in essential resource pathways often give rise to a state known as *cellular burden*, characterised by slowed cellular growth and altered endogenous gene expression profiles. While cellular burden has been extensively described in microbial hosts, where the impact of heterologous expression on cellular growth has been repeatedly proven[42,61–64], initial evidence has also been published for mammalian hosts, with studies showing increased growth capabilities following the shutdown of unnecessary heterologous genes[65,66]. Cellular burden can manifest at multiple levels. It can be caused by the overloading of gene expression resources, such as DNA and RNA polymerases, tRNAs, and ribosomes. It can also result from the generation of toxic intermediates, the depletion of critical metabolic resources (including ATP, cofactors, metabolites, and carbon building blocks), or the strain imposed on post-translational resources[42,67]. The burden caused by limitations in gene expression resources was shown to be one of the primary sources of cellular burden in *E. coli*[61]. Following heterologous expression, gene expression resources become limiting, reducing the available transcriptional and translational capacity for endogenous expression and cellular growth. In particular, RNA polymerases and ribosomes have been identified as the key cellular resources undergoing limitations following heterologous expression[61,68]. Further evidence showed how bottlenecks in resource pathways other than gene expression can also lead to burdened cells. For instance, Cookson *et al.* showed how the saturation of the *E. coli* ClpX protease alters the relative intracellular concentration of highly-degraded proteins[69]. Sosa-Carrillo *et al.* used a real-time quantification approach to assess the secretory stress, showing that different secreted proteins had unique stress profiles on the yeast secretory pathway[70]. Finally, Cella *et al.* developed a mathematical model to investigate the role of microRNAs on resource redistribution in mammalian cells, showing a concerted action of microRNAs on the allocation of ribosomes and RNA degradation resources in engineered mammalian cells[71].

1.3.1. Resource competition as a source of emergent properties

Resource competition dominates construct-to-host interactions, leading to emergent properties that hamper our capacity to engineer at scale predictably and reliably. These include coupling of the expression of independent genes, coupling of circuit performance to cellular states, and emergence of escape mutants in engineered cellular populations.

1.3.1.1. Coupling of the expression of independent genes

The first indication that two genes are competing for limiting resources is the coupling of their expression levels (Figure 1.3a). This was clearly exemplified by Gyorgy *et al.,* who showed that the expression of an inducible cassette led to decreased expression levels from a constitutive cassette in *E. coli*[68]. Although the two expression cassettes were not functionally linked, their output levels were coupled due to their reliance on the same pool of finite resources. As more resources were required for the expression of the inducible cassette, fewer were available for the expression of the competing constitutive cassette, which ultimately suffered from decreased expression output. More recently, similar coupling effects have been

observed between co-transfected genetic constructs in mammalian cells[48,72]. For example, Frei *et al.* showed that upon co-transfection of a constitutively-expressed gene and a doxycycline-repressible gene, doxycycline repression of the latter resulted in higher expression levels of the former[72]. More importantly, coupling effects can also arise between genetic constructs and endogenous cellular genes, as shown by Ceroni *et al.* in *E. coli*[61]. This has important implications as by affecting the expression of endogenous genes, resource competition will inevitably impact cellular physiology and growth, as mentioned above.

1.3.1.2. Coupling of circuit performance to cellular states

The impact of resource competition on cellular growth can alter the behaviour of entire circuits in unpredictable ways. For example, Tan *et al.* noticed that bistability can arise from a simple positive feedback circuit due to the growth delay of burdened engineered bacteria, resulting in a nonlinear dilution of the activating protein[62]. These findings uncovered an intrinsic coupling mechanism of engineered cells – referred to as growth feedback - whereby genetic circuits cause suppressed cellular growth, which in turns alters the circuits' behaviour (Figure 1.3b). Many examples have shown that growth feedback modifies the behaviour of genetic circuits in difficult-to-predict ways[73–76]. Among these, recent literature has computationally characterised the effect of growth feedback on 435 circuit topologies, uncovering failure modes and identifying circuit architectures with increased robustness characteristics[76]. Finally, resource competition can endow genetic circuits with emergent properties that are not necessarily due to growth feedback. Notable examples are winner-takes-all behaviours in cascading bistable switches[77] and coupling of otherwise independent genetic oscillators[78].

1.3.1.3. Emergence of escape mutants

Growth-impaired resource-depleted engineered cells display a fitness disadvantage compared to non-engineered cells[61]. This plays an important role in the emergence of escape mutants within engineered cellular populations[42,67]. Escape mutants carry loss-of-function mutations that impair the expression of genetic constructs, circuits, or pathways, ultimately releasing the burden of heterologous expression and increasing cellular fitness and proliferation (Figure 1.3c). Escape mutants have important implications for microbial-based bioproduction processes, which will be discussed in section 1.3.2.2.

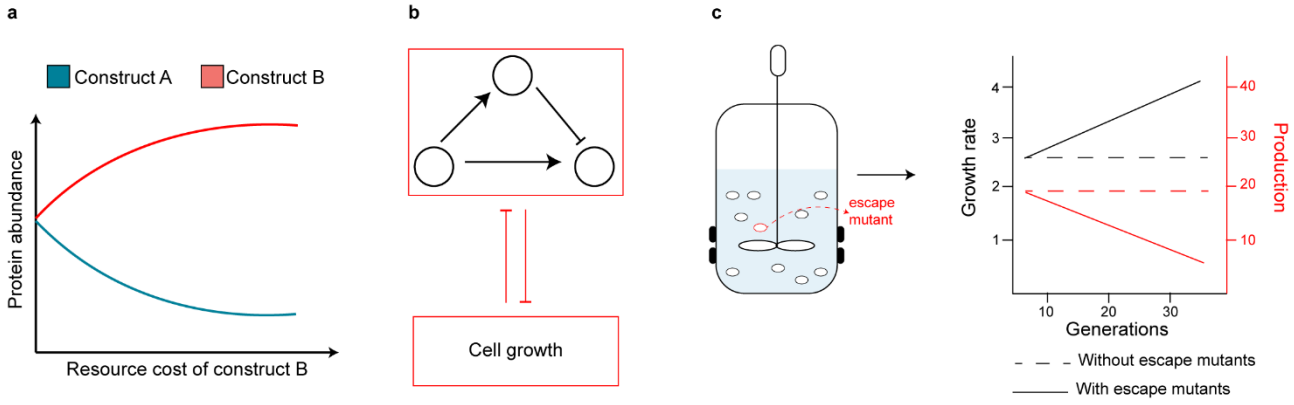

Figure 1.3 Emergent properties of resource competition. (a) Resource competition arises from the simultaneous use

of cellular resources by different expression cassettes. Co-expressed synthetic cassettes (Construct A and Construct B) rely on the same pool of cellular resources. As the resource cost of one cassette is increased (Construct B), fewer resources will be available for the expression of the other cassette (Construct A), ultimately affecting the abundance of protein A. **(b)** Heterologous expression of genetic circuits leads to decreased cellular growth, which in turn impacts the genetic circuits' behaviour. **(c)** Engineered cells are at a fitness disadvantage compared to non-engineered cells, carrying profound implications for bioprocesses. Faster-growing non-producing escape mutants are doomed to emerge over prolonged cultivation and eventually outcompete the slower-growing producing population, causing a total production shutdown.

1.3.2. Resource competition as a systemic problem for life science

Despite increasing evidence on the impact of resource competition on cell engineering, the topic remains largely unexplored outside the synthetic biology field. Acknowledging resource competition as a systemic problem for life science will be the first pivotal step towards improving our ability to understand and reliably engineer living systems. Below we discuss how resource competition affects foundational biology, bioproduction, and biotherapy.

1.3.2.1. Implications of resource competition for foundational biology

While attention to resource competition has mainly originated from synthetic biology, implications for foundational biology are severe and potentially leading to data misinterpretation when unaccounted for. One prime example is the adoption of transfection controls for transient transfections experiments in mammalian cells[79]. Transfection controls are genetic constructs encoding for a readout gene, such as a fluorescence or luminescence gene. Researchers use transfection controls to avoid experimental biases introduced by the intra- or inter-experimental variability in transfection efficiencies[80]. To do so, transfection controls are co-transfected with the test genetic construct of interest, serving as a marker for plasmid uptake. In some cases, co-transfection is followed by a normalisation procedure in which the output of the test genetic construct of interest is divided by the output of the transfection control readout gene. This facilitates data interpretation by accounting for variability in transfection efficiency. However, this practice can dangerously lead to data misinterpretation by failing to acknowledge that transfection controls and test genetic constructs compete for cellular resources and will inevitably face coupling of their outputs as soon as resources become limiting. A very early warning comes from the characterisation of promoter activity in mammalian cells. Farr *et al*. showed that expression of their transfection control was significantly suppressed when co-transfected with their highly expressed positive control plasmid in HeLa cells[81]. Although they did not point to resource competition, we now know that this is probably a result of coupling induced by resource shortages. Following this observation, a call for caution must be issued for using transfection controls to normalise raw experimental data, as this would result in the overestimation of the output of genetic constructs with high resource demands compared to constructs with lower demands. Given the widespread use of transfection controls in biology experiments, awareness must be raised to

avoid experimental artefacts, and, if possible, alternative ways to assess transfection efficiency should be used[79].

While transfection controls represent a clear case study, resource competition has the potential to confound biological results in many ways, especially when it comes to quantitative measurements. Evidence of downregulation of endogenous genes in resource-overloaded bacteria and mammalian cells is present in the literature[61,72]. Importantly, Frei *et al.* provided initial evidence of reduced transcript levels of a housekeeping gene often used as a reference in qPCR applications (i.e. GAPDH) in H1299 cells transfected with a resource-depleting genetic construct[72]. This has the potential to bias the outcome of qPCR experiments in scenarios of resource limitations. Similar considerations can be drawn for quantitative measurements at the protein levels, such as for western blot experiments.

Besides these obvious examples, resource competition might confound foundational biological experiments in ways that we cannot fully predict or comprehend due to the limited information we have today. What should be clear is that the impact of resource competition on foundational biology extends way beyond the misinterpretation of specific experiments, flagging a systemic problem where reproducibility and correct interpretation of biological phenomena may be compromised.

1.3.2.2. Implication of resource competition for bioproduction

The implications of resource competition for bioproduction have been extensively investigated. The issues are twofold: coupling of the expression of production genes and the emergence of escape mutants in long-term bioprocesses. Eventually, both result in suboptimal titers of the product of interest.

First, for bioproduction applications, cells are often engineered with more than one expression cassette. For instance, mammalian-based production of monoclonal antibodies requires the expression of more than one gene, and for the microbial-based production of natural products, entire metabolic pathways might be needed. Whether these expression cassettes are genomically integrated or transiently delivered, they rely on the same pool of resources and face competition, resulting in suboptimal co-expression of the production genes of interest. It is likely that resource competition causes a more meaningful interference for those production processes with strict stoichiometry requirements for optimal production[82,83].

In addition to this, resource-depleted burdened cells are at a fitness disadvantage. As mentioned before, escape mutants with normal growth rates will eventually emerge within the engineered cellular population, impacting the performance of large-scale bioprocesses. While at the start of the bioprocess there will only be producer cells in the bioreactor, in few generations genetic drifting of the engineered population will lead to phenotypic heterogeneity, with some cells losing production due to loss-of-function mutations. After phenotypic heterogeneity arises inside the bioreactor, the non-producing faster-growing escape mutant population will quickly outcompete the slower-growing producing population. This ultimately shortens the lifetime of bioprocesses, hampering yield and increasing production costs (Figure 1.3c). While

for mammalian-based bioprocesses this is not a major problem, as stability studies are performed beforehand and only very stable cell clones are selected for production processes, the opposite is true for microbial-based bioprocesses, where phenotypic heterogeneity is doomed to arise at some point during the bioprocess, possibly due to the faster growth of microbes compared to mammalian cells. Overall, there is space to improve bioproduction processes once these issues are addressed.

1.3.2.3. Implications of resource competition for biotherapy

How resource competition comes to play in biotherapy remains largely unexplored. As current gene therapies often include the expression of two or more expression cassettes, competition for resources is likely to affect these scenarios. Nevertheless, the problem is rarely acknowledged, and many questions remain unanswered.

First, we do not know how resource competition affects the titers of viral vectors for gene therapies. Adeno-associated viruses and retroviruses are often used as platforms to deliver information inside human cells. The production processes of these viral vectors entail mammalian-based transient expression and assembly of appropriate viral components in definite stoichiometry. From what we know today, during these production processes, resource competition is likely to cause coupling of the expression of viral genes, preventing optimal assembly of viral vectors. While suboptimal production of viral vectors represents a problem[84], resource competition has never been considered a contributing factor to our knowledge. Studies highlighting the implications of resource competition for viral vector production processes are awaited.

Another interesting question is how resource competition comes into play in *in vivo* therapies. So far, all evidence of resource competition has involved microbes or fast-dividing mammalian cell lines in a controlled culture environment. Understanding how resource competition affects human cells inside a living organism might open to considerations for biotherapy. How resource competition could influence *in vivo* therapies is twofold. First, it could cause coupling and suboptimal expression of therapeutic genes. Second, it could cause cellular stress that limits physiological functions. Although this is very much an unexplored area, initial proof-of-concept investigations have the potential to spark a new research field focused on developing resource-aware biotherapies that perform better than currently available approaches.

1.4. Quantification and mitigation of resource competition

In the last decade, synthetic biologists developed tools, frameworks, and clever strategies to visualise and mitigate resource competition *in vivo*. This section will start by discussing available quantification methods and will continue with an in-depth analysis of current mitigation strategies. I grouped the available mitigation strategies into three macro-categories: engineering of the cellular host chassis, implementation of regulatory circuits, and informed genetic construct design. Each of these strategies has its benefits and

drawbacks, with some being more suited for specific applications than others. Notably, most of these strategies were tailored for microorganisms, and only recent research opened to mammalian cells.

1.4.1. Quantification of resource competition

Traditional methods to quantify the burden imposed on engineered microbes relied on measuring growth rates[85–87]. While these methods allowed us to visualise cellular stress, they did not provide information on its causes. The development of fluorescence-based tools allowed real-time quantification of resource competition *in vivo* and played a crucial role in studying resource competition in living cells. One example comes from Ceroni *et al.,* who integrated a constitutively-expressed fluorescence-based cassette – termed *capacity monitor* – into the genome of two *E. coli* strains[61]. The authors used the capacity monitor to report on how much gene expression resources were sequestered by different genetic constructs. They reasoned that since the capacity monitor and the genetic constructs relied on the same pool of cellular resources, coupling effects arising from resource shortages could be visualised by looking at the fluorescence produced by the capacity monitor. Indeed, when the resource demand from genetic constructs increased, the authors observed decreased capacity monitor production rate, indicating decreased available capacity for gene expression resources. Importantly, the authors found a positive linear correlation between the capacity monitor production rate – which they term *capacity* - and the cellular growth rate, with capacity decreasing earlier than the growth rate. This suggests that the decrease in capacity stands as a cause for reduced growth and that the capacity monitor fluorescence can be used as a proxy measure of the burden imposed on the engineered cells. Capacity monitors are great assets to compare the resource footprint of libraries of genetic constructs, allowing the identification of crucial design parameters impacting the resource demand of genetic constructs and ultimately enabling the informed design of genetic constructs with minimised resource footprint. For instance, using their capacity monitor, Ceroni *et al.* identified ribosome-binding site efficiency as the key design parameter determining the load of genetic constructs on gene expression resources, highlighting ribosomes as essential limiting resources in engineered *E. coli*. This was in accordance with previous models of bacterial growth[88].

Recently, fluorescence-based sensors allowed the visualisation of resource competition in mammalian cells[48,72]. In these cases, the sensors were transiently co-delivered with the test genetic construct of interest and were not integrated into the genome of mammalian cells. In this thesis similar transfected designs will be referred to as *transient capacity monitors,* as opposed to the *integrated capacity monitors* developed by Ceroni *et al.*[61,89]*.* The distinction is important as the delivery method of capacity monitors eventually dictates the type of information that can be inferred using these tools. One of the main advantages of integrated capacity monitors is that they are placed in the same context and have equal copy number to endogenous genes. On the other hand, for transient capacity monitors, context and copy number greatly differ from endogenous genes. This makes integrated capacity monitor the best option to assess the effect

of heterologous expression on endogenous genes – and ultimately on cell physiology -, while relegating transient capacity monitors to effective tools to quantify resource competition amongst co-transfected genetic constructs. In my work, we adopted transient capacity monitors to shed light on the resource demand imposed by different transiently-expressed genetic designs in mammalian cells. We also generated integrated capacity monitors to inform on the cell response to heterologous expression in mammalian systems.

As eukaryotes possess compartmentalisation and higher functional complexity than prokaryotes, fluorescence-based sensors to monitor more specific cellular stress conditions have also been developed. For instance, several studies proposed new tools to visualise the unfolded protein response (UPR) in yeast and mammalian cells[70,90–93]. The UPR is a complex signalling pathway that is activated in response to stress caused by the overproduction and misfolding of proteins in the endoplasmic reticulum[94]. Activation of the UPR results in increased secretory capacity and augmented degradation of accumulated proteins[95]. Given its implication for the bioproduction of secreted proteins, tools for the real-time visualisation of UPR stress can be invaluable to increase our understanding of the topic. For instance, Du *et al.* implemented a UPR sensor by placing a fluorescence gene under the control of a UPR-inducible promoter and used this tool to monitor the UPR response during mammalian fed-batch fermentation processes[90]. Similarly, UPR-responsive promoters have been used to implement fluorescence-based sensors to monitor UPR stress in yeast[92]. While these and similar studies represent foundational advances in developing tools to quantify ER capacity in eukaryotes[70,91], a comprehensive assessment of the impact of genetic constructs on ER resources – and, more in general, on resources involved in the secretory pathway - is still lacking.

1.4.2. Mitigation of resource competition by engineering the cellular host

Cellular hosts can be engineered towards increased genetic stability, reduced resource demand, and division of labour.

1.4.2.1. Engineering cellular hosts with increased genetic stability

Biological populations, both engineered and natural, are constantly evolving, with rates from 10^{-2} to 10^{-10} per base pair per generation[67]. Evolution is enabled by various mechanisms, including mutation, recombination, and transposition, all contributing to genetic instability. For instance, mutations in antibody-producing genes have been shown to hamper production in CHO cells[96]. Recombination-mediated disruption of production genes has been observed in microbial- and mammalian-based production chassis[97–99]. Transposable bacterial insertion sequence elements terminated expression of a mevalonate production pathway in 70 generations[100]. As previously stated, the disruption of production genes typical of escape mutants rescues cellular fitness by releasing cellular resources. This ultimately causes the escape mutants to overcome the slower-growing producing population, completely shutting down production. One approach to overcome this issue is to increase the genetic stability of production chassis (Figure 1.4a).

For instance, deleting transposable insertion sequences in bacterial production strains achieved increased genome stability and more stable production[101–103]. Similar efforts are being carried out in yeast, with the Sc2.0 project aiming to remove all the Ty transposable elements from the genome of *S. cerevisiae*[104]. Complementarily, bacterial strains with deleted recombinases and error-prone polymerase genes have also been engineered[105,106]. Although these strategies can effectively prolong the lifetime of production processes, they do not mitigate resource shortages or reduce the cellular burden caused by heterologous expression.

1.4.2.2. Engineering cellular hosts with reduced resource demand

Deleting non-essential genes from the genome of the cellular host was proved to be an effective strategy to reallocate cellular resources (Figure 1.4b)[107,108]. This is supported by instances of genome-reduced bacterial strains with increased production capabilities than their wild-type counterparts[109–111]. Arguably, the most elegant contribution to this area of study so far came from Lastiri-Pancardo *et al.*[112]. The authors integrated transcription factor interaction networks and proteomics data to rationally identify the minimal set of genetic deletions to maximise resource reallocation. By generating combinatorial mutant *E. coli* strains, they released proteome budget and increased heterologous production by 18%[112]. Similar strategies have been proposed for mammalian cells. Kallehauge *et al*. demonstrated increased antibody production and cellular growth following the silencing of non-essential antibiotic resistance genes in an antibody-producing CHO cell line[65]. This laid the foundation for more mature work by Kol *et. al.*, where the authors identified 14 genes coding for non-essential secreted host cell proteins that they successfully removed in multiplexed deletion experiments[66]. While deletion of the 14 genes resulted in decreased cellular growth, by knocking-out only 11 out of the 14 selected genes the authors obtained increased viable cell density and antibody titer[66]. Overall, genome minimisation approaches proved effective in microorganisms and mammalian cells, with benefits over cellular growth and heterologous production from genetic constructs. The development of genome-scale metabolic models promises to streamline the identification of suitable genomic targets for these experiments, resulting in the optimal release of cellular resources and reallocation[113,114]. One drawback of these methods is the necessity of new engineering experiments once a new cellular host is adopted.

1.4.2.3. Engineering cellular hosts towards division of labour

The engineering of microbial cells for bioproduction purposes often involves the refactoring of complex metabolic pathways into the cellular host of choice. This leads to a significant burden on the host as a considerable portion of its cellular resources will have to be allocated to express all the genes of the metabolic pathway, ultimately hampering productivity. One potential solution to this limitation is to divide the burden of the metabolic pathway amongst different cellular hosts by programming different cellular populations to perform different steps of the pathway (Figure 1.4c)[115]. The final product can be synthesised at reduced burden costs for cellular hosts by co-culturing the production strains together. Division of labour

proved effective in enhancing the bio-based production of commodity and high-value chemicals[116–121], but limitations remain. Intermediates molecules often must be shared amongst various production strains, introducing considerations on the transport efficiency of these intermediates across cellular membranes. In addition, co-cultured production strains often have different growth rates, thus complicating the optimisation of co-culture processes with extended production lifetime[122,123].

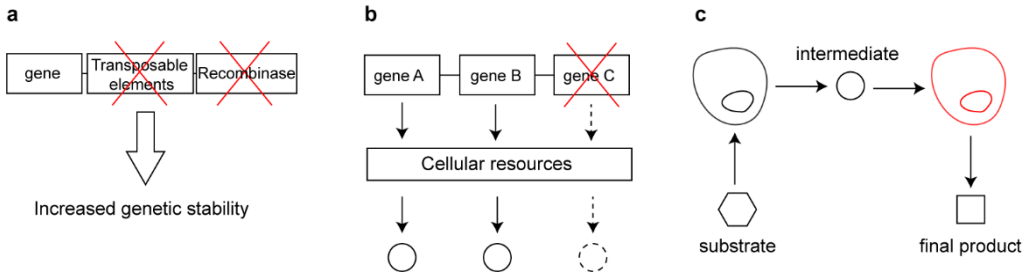

Figure 1.4 Mitigation of resource competition by engineering of the cellular host. (a) Cellular hosts can be engineered towards increased genetic stability by deletion of recombinase genes and transposable elements. **(b)** Cellular hosts can be engineered towards decreased resource demand by deleting non-essential resource-depleting genes. **(c)** Cellular hosts can be engineered towards division of labour, whereby resource-consuming production pathways are split into different hosts that are then co-cultured together.

1.4.3. Mitigation of resource competition by implementing regulatory circuits

Other approaches to mitigate resource competition involve the implementation of regulatory genetic circuits, including adaptive controllers, circuits with orthogonal components, and circuits for population control.

1.4.3.1. Engineering adaptive genetic controllers

The engineering of adaptive controllers based on negative feedback, antithetic integral feedback, or incoherent feed-forward topologies showed great promise to improve robustness to disturbances, including fluctuation in the pool of cellular resources[124].

Negative feedback regulation is widespread in biological systems[125,126]. By implementing negative feedback loops in engineered circuits, it is possible to decouple the expression of competing cassettes (Figure 1.5a – left panel)[107,124]. For instance, Shopera *et al.* demonstrated reduced coupling between the output of two competing genetic circuits after implementing negative feedback regulation compared to the open topology[127]. Examples of feedback loops where cellular stress or burden are used as input are also present in the literature. Ceroni *et al.* developed a burden-responsive feedback controller to alleviate the burden of *E. coli* engineered to express a burdensome gene[89]. The authors identified a burden-responsive promoter and used it to guide the expression of a sgRNA that, together with a constitutively-expressed dCas9, repressed the expression of the burdensome gene[89]. In this way, the expression of the burdensome gene

can be adjusted to minimise the stress on the cellular host, increasing cellular growth and ultimately maximising product yields.

More recently, antithetic integral feedback architectures have been engineered to enable robust perfect adaptation (Figure 1.5a – right panel)[128–132]. Similar topologies are implemented by two species that annihilate each other (Z1 and Z2), where Z1 is constitutively produced and works by affecting processes leading to increased expression of Z2. Huang *et al.* engineered a post-transcriptional quasi-integral feedback controller to overcome fluctuation in the free ribosome pool in *E. coli,* demonstrating increased robustness to resource loading from a competing genetic circuit compared to the open loop configuration[131]. Frei *et al.* engineered a proportional-integral feedback controller to mitigate expression coupling in mammalian cells due to resource sharing[132].

Finally, incoherent feed-forward loops have also been used by synthetic biologists to buffer expression outputs to perturbations (Figure 1.5a – middle panel)[133]. In these architectures, an input Z1 controls the expression of two species, Z2 and Z3, with Z2 repressing the expression of the regulated Z3. Two seminal works implemented incoherent feed-forward loops to mitigate resource competition in mammalian cells. Jones *et al.* engineered an incoherent feed-forward controller where cellular resources are the input Z1 driving expression of the Cas6 protein (Z2) and a regulated fluorescence gene (Z3), with Cas6 repressing the expression of the fluorescence gene[48]. Following shortages in cellular resource availability, the expression of Cas6 is suppressed, thus releasing the repression of the fluorescence gene. This way, the decrease in the expression of the fluorescence gene due to resource shortages is counteracted by the release of the repression from Cas6, thereby achieving robustness of the expression of the fluorescence gene to fluctuations in the pool of cellular resources[48]. Similarly, Frei *et al*. used cellular resources as input for their incoherent feed-forward loops but used endogenous or synthetic microRNAs (Z2) to repress the regulated output gene (Z3)[72].

Although adaptive genetic circuits can effectively mitigate resource competition, they usually come at the cost of suppressed expression of the regulated output gene. This is a consequence of all these topologies relying on a negative regulation step, making these strategies not ideal when maximisation of the output gene is desired.

1.4.3.2. Engineering genetic circuits using orthogonal components

Thought-provoking ideas of a completely orthogonal central dogma are gaining traction within the synthetic biology community[134]. Orthogonal components can decouple heterologous expression processes from the native expression machinery (Figure 1.5b). To this date, orthogonal DNA replication[135], transcription[136], translation[137–141], and protein degradation[142] have been engineered in synthetic biology applications. A prime example of how orthogonal resources can be useful in mitigating resource competition comes from Darlington *et al.,* where the authors used an orthogonal 16S rRNA – RBS RNA pair

to partition translation between a host ribosome pool and an orthogonal ribosome pool[143]. While the host ribosome pool was allocated to the translation of endogenous proteins, the orthogonal ribosome pool would uniquely engage in the translation of heterologous proteins. Using this approach, the authors managed to increase the yield of heterologously-expressed violacein[143]. In addition, they designed a feedback-based resource allocator that dynamically adjusted the expression of orthogonal 16S rRNA based on the translational demand placed by heterologous production, reducing coupling effects between co-expressed genes by 50%[143]. However, this resource allocator still relies on the host 32S rRNAs subunits and better alternatives are awaited to achieve perfect orthogonality. Entirely orthogonal ribosome alternatives have been engineered to completely decouple endogenous and heterologous translation[138]. Examples of fully orthogonal transcriptional-translational networks have also been explored[144,145].

1.4.3.3. Engineering genetic circuits for population control

Engineering genetic circuits that couple cellular growth to heterologous production represents an effective strategy to combat the evolutionary drift of engineered cellular populations and is particularly valuable in extending the lifetime of bioprocesses (Figure 1.5c)[67]. Rugbjerg *et al.* coupled the expression of non-conditional essential genes to heterologous production in *E. coli* cells engineered with a pathway for mevalonic acid production[146]. In particular, the authors placed genes necessary for bacterial growth under the control of a promoter responsive to the end product of the metabolic pathway. In this way, escape mutants are at a fitness disadvantage due to the shutdown of the expression of essential genes. This strategy – that the authors termed *synthetic addiction* - extended the lifetime of the fermentation process up to 95 generations, stabilising both population growth and product titer over prolonged cultivation[146]. Similar approaches have been published in yeasts[147,148]. Synthetic addiction is probably one of the most effective strategies to overcome population heterogeneity, although it does not directly mitigate resource competition and burden. An advantage over the engineering of adaptive circuits is that, in this case, no suppression of the product of interest is observed, making synthetic addiction more suited for bioproduction applications. One of the main drawbacks is the requirement to finely tune the selective pressure of the addiction platform so that it is not too weak to allow the propagation of the escape mutant population and not too strong to hamper the growth of the producing population[148]. Another disadvantage is their lack of portability due to the reliance on biosensors responding to specific end products.

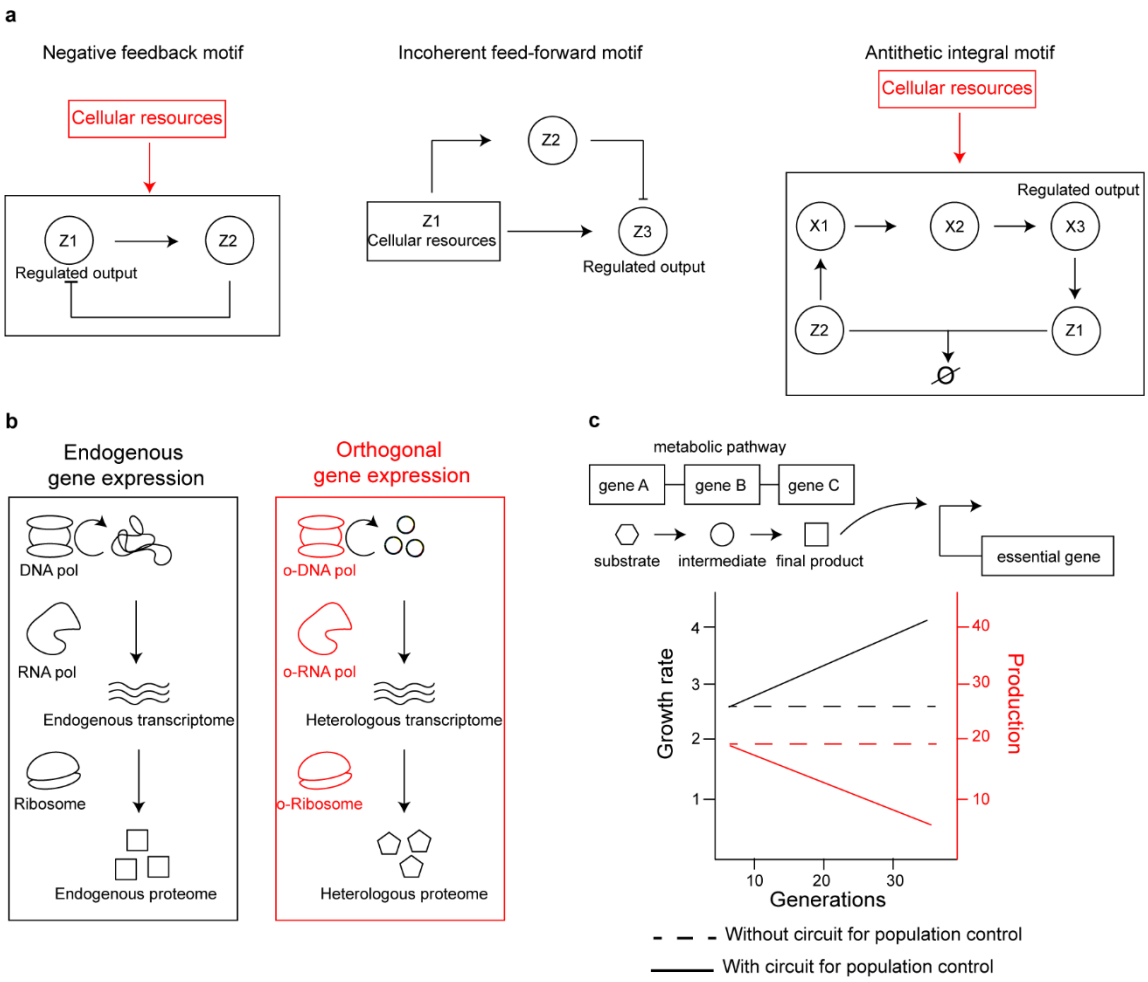

Figure 1.5 Mitigation of resource competition by the implementation of regulatory circuits. (a) Engineering adaptive genetic controllers based on negative feedback (left panel), incoherent feed-forward (middle panel), or antithetic integral feedback (right panel) architectures can buffer the expression of a regulated gene to perturbations in the pool of cellular resources. **(b)** Engineering genetic circuits using orthogonal components can reduce the resource demand that heterologous expression places on the native cellular machinery. **(c)** Engineering genetic circuits for population control can curb the emergence of escape mutants, ensuring stable growth rates and production during long-term fermentation processes.

1.4.4. Mitigation of resource competition by informed genetic construct design

Lastly, implementing informed genetic construct design has also been used to mitigate resource competition. This was pioneered by Ceroni *et al.*, where the authors used an integrated capacity monitor to infer design rules resulting in genetic constructs with decreased resource footprints in *E. coli*[61]. More importantly, the authors observed that higher gene expression outputs from genetic constructs did not always lead to increased resource footprints and that minimised resource demand of genetic constructs could be achieved by implementing resource-aware design considerations. For instance, the authors found that increasing transcriptional efficiency (through stronger promoters) and decreasing translational efficiency (through weaker ribosome-binding sites) led to reduced resource footprint on similar expression

outputs[61]. These considerations are extremely powerful for many reasons. First, they represent general rules that can be applied to design any genetic constructs of interest. Second, they can be used to generate genetic designs with optimised resource footprints without compromising on their expression outputs. More recently, Jones *et al.* used transient capacity monitors bearing different promoters to screen the resource footprint of a library of activator domains in a panel of mammalian cell lines[48]. The authors showed that coupling effects were exacerbated for specific promoter-activator combinations and noticed variability of their results across cell lines. This introduces variability as one of the key shortcomings of this approach. Strains and cell lines likely differ in resource allocation profiles, necessitating an *ad hoc* prior characterisation before intelligent design can be implemented.

Computational-aided design is likely to benefit this area of study. Already in 2016, Cello made headlines for enabling the computer-aided design of genetic logic circuits[149]. With the advances in high-throughput biology, it is now possible to generate vast biological datasets that can be analysed with machine learning approaches to achieve predictive and reliable biological design[150].

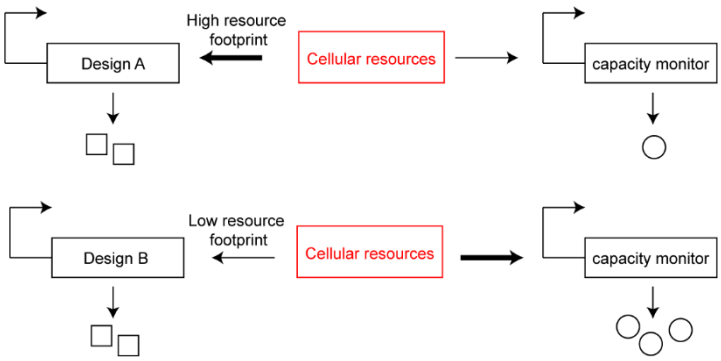

Figure 1.6 Mitigation of resource competition by informed genetic design. Fluorescence-based capacity monitors can report on the resource cost of libraries of genetic constructs, allowing the individuation of efficient designs with high gene expression performance and minimal resource cost.

1.5. State of the art of resource competition research in mammalian cells

From the previous discussion, it should be evident that most research in resource competition has focused on microorganisms and that research in mammalian cells has lagged behind (Figure 1.7). This section will summarise the current state of the art of resource competition research in mammalian cells, highlighting key outstanding questions and proposing an agenda for future research efforts. This will clearly outline the research background of my PhD studies.

1.5.1. Resource competition in mammalian cells: state of the art

Studies of resource competition in mammalian cells originated within the last five years. Since then, proof of concept works have used transient capacity monitors to visualise resource competition amongst co-transfected systems in mammalian cells[48,72], providing initial insights on the locations of resource bottlenecks in the gene expression pathway. While there seems to be consensus on the limitations of

transcriptional resources in engineered cells, contrasting evidence is present on translational resources[48,71,72]. The reason for these discrepancies might be attributed to the early stage of these studies, indicating that a more thorough investigation is needed to correctly inform on the location of resource bottlenecks in engineered mammalian cells.

Besides gene expression, other resource pathways have been suggested to undergo limitations. For instance, Cella *et al.* reported shortages in the RNA degradation pathway[71]. Reports of resource bottlenecks in the secretion pathway are also present, although – based on current data – it is not possible to disentangle the contribution of resource shortages in the gene expression and secretion pathways[66,113].

Initial mitigation strategies have been developed. Adaptive circuits based on antithetic integral feedback or incoherent feed-forward architectures have been engineered[48,72,132]. As stated before, while these strategies effectively buffer the expression of the regulated gene to fluctuations in the pool of cellular resources, they do so at the cost of dramatic output suppression. Other mitigation strategies fall into the genome minimisation category, with a seminal study proving that deletion of unnecessary genes resulted in increased heterologous production of secreted proteins[66].

Overall, our understanding of resource competition in mammalian cells is way more scattered than in microbes, with less evidence and limited mitigation strategies available. This places resource competition in mammalian cells as an area of study with disruptive potential once key outstanding questions are addressed.

1.5.2. Resource competition in mammalian cells: four calls to action

Four are the most pressing outstanding questions that need action to fully understand the implications of resource competition in mammalian cells.

First, while competition for gene expression resources has been shown to affect co-transfected systems in mammalian cells, a clear indication as to what types of resources are limiting has yet to be provided. Understanding which resources are involved can help us design improved genetic constructs with a reduced resource footprint. It is also still unclear to what extent other types of resources come into play to shape the overall competition landscape. Initial works suggest that RNA degradation and protein secretion resources might be limiting in mammalian cells[66,71], but clear evidence is still missing.

The second outstanding question is whether resource competition is limited to co-transfected systems or affects other scenarios. Can coupling effects arise between transfected and integrated genes or amongst integrated genes? This key question will help us define the extent of the resource competition problem in mammalian cells. As genomic integration is one of the most sought-after methods to deliver genetic payloads for bioproduction and biotherapy applications, understanding whether coupling effects can disrupt the functionality of genetic constructs in these scenarios is crucial.

Third, it still needs to be determined what cellular effects mammalian cells experience as a consequence of resource limitations. While resource depletion is typically associated with altered endogenous expression profiles and reduced cellular growth for microbes, comprehensive characterisation and quantification of the problem are still missing in mammalian cells. Mammalian cells are more complex than microbes and might reasonably have evolved natural mechanisms to decouple cellular growth from resource availability. They also possess much longer doubling times, and any effects of resource depletion on cellular growth might not be as relevant for slow-growing mammalian cells as it is for fast-growing microbes. While early evidence shows the rescuing of cellular growth after the release of cellular resources in mammalian cells[65,66], a more coherent discussion would benefit the community.

Finally, mitigation strategies have also been identified for mammalian cells. As previously stated, engineered mammalian hosts with minimised genome and adaptive circuits have been implemented. These strategies come with their perks and limitations, and new complementary approaches are awaited.

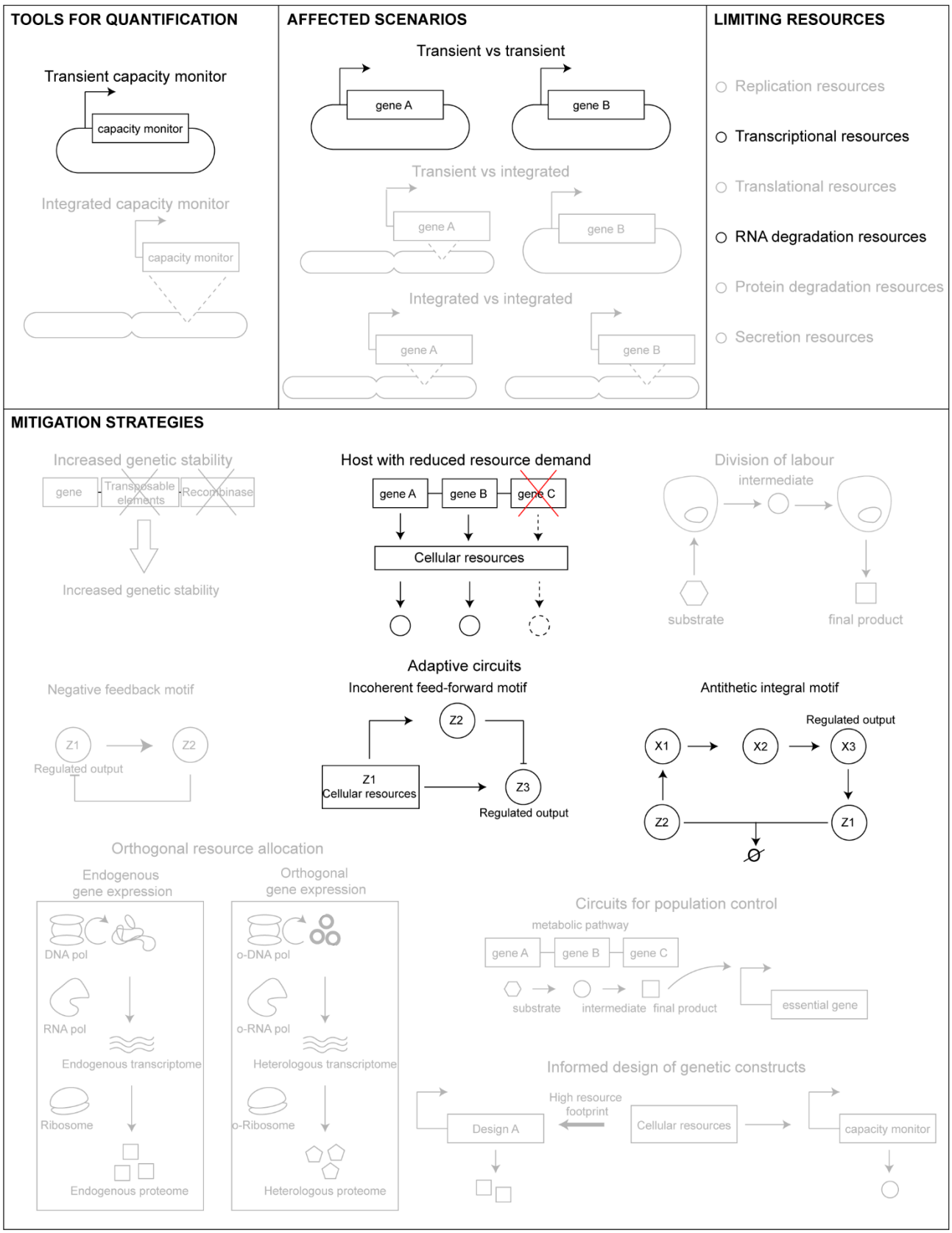

Figure 1.7 State of the art of resource competition research in mammalian cells. The figure displays published evidence of resource competition in mammalian cells in black. Awaited advancements are shown in grey.

1.6. Aims and objectives

The aim of this PhD project was to characterise and quantify resource competition in engineered mammalian cells and develop novel tools for burden analysis in mammalian systems, extending current considerations to integrated systems, alternative resource pools, and resource-informed construct design.

Firstly, in Chapter 2, we introduce a resource-aware framework to characterise the resource footprint of libraries of genetic constructs. By screening libraries of biological parts, we uncover important insights on the location of resource bottlenecks in the gene expression pathway. We also outline design rules to streamline the generation of genetic constructs with maximised performance and reduced resource footprint, opening to resource-informed design and showing usefulness for bioproduction applications.

In Chapter 3, we use the same resource-aware framework to characterise the robustness to resource shortages of libraries of genetic constructs. We identify design parameters influencing the robustness of genetic constructs to resource fluctuation. Additionally, we demonstrate the implications of our findings for co-transfection experiments, highlighting sources of bias and proposing potential solutions.

In Chapter 4, we develop the first version of a fluorescence-based integrated capacity monitor in mammalian cells. Using the integrated capacity monitor, we uncover the effects of resource competition on the physiology of engineered mammalian cells. Additionally, during the process, we highlight design insights to generate dynamic and responsive integrated biosensing systems in mammalian cells.

Finally, in Chapter 5, we adapt the resource-aware framework to investigate resource bottlenecks in other cellular pathways, such as protein degradation and secretion. We prove that resource shortages affect these two pathways in engineered mammalian cells, laying the foundation for further research.

1.7. Note to the reader: contributions and publications

1.7.1. Contributions made by other researchers

Part of this work has been done in collaboration with other researchers. Here are outlined relevant contributions made by other researchers.

In Chapter 2, Dr Simone Furini developed a custom script for the automatic gating of flow cytometry data. Dr Masue Marbiah performed all western blot experiments. Other researchers provided feedback during the project.

In Chapter 3, Dr Simone Furini developed a custom script for the automatic gating of flow cytometry data.

In Chapter 4, no external contribution was made.

In Chapter 5, PhD student Jacopo Gabrielli and I co-led the work regarding competition for protein degradation resources. Specifically, I conceived the initial idea of the work and independently generated proof of concept, performed most of the cloning, generated the stable cell line, and performed data analysis. Jacopo took care of part of the cloning, performed transient transfection experiments, and supported the project's advancement. MRes student Henry Chippindale performed transfection experiments to assess competition for secretion resources.

1.7.2. Publications resulting from the PhD work

All the work published during the PhD is listed below. Asterisks indicate first authors.

R Di Blasi*, O Blyuss, JF Timms, D Conole, F Ceroni, HJ Whitwell. Non-histone protein methylation: biological significance and bioengineering potential. *ACS chemical biology* 16 (2), 238-250

R Di Blasi*, A Zouein*, T Ellis, F Ceroni. Genetic toolkits to design and build mammalian synthetic systems. *Trends in Biotechnology* 39 (10), 1004-1018

A Grob*, **R Di Blasi***, F Ceroni. Experimental tools to reduce the burden of bacterial synthetic biology. Current *Opinion in Systems Biology* 28, 100393

R Di Blasi*, MM Marbiah*, V Siciliano, K Polizzi, F Ceroni. A call for caution in analysing mammalian co-transfection experiments and implications of resource competition in data misinterpretation. *Nature Communications* 12 (1), 2545

R Di Blasi*, M Pisani, F Tedeschi, MM Marbiah, K Polizzi, S Furini, V Siciliano, F Ceroni. Resource-aware construct design in mammalian cells. *Nature Communications* 14 (1), 3576

A Grob*, CE Bena*, **R Di Blasi**, D Pessina, M Sood, Z Yunyue, C Bosia, M Isalan, F Ceroni. Mammalian cell characterisation by non-invasive plate reader assay. *bioRxiv* 2023.03. 22.533848

1.7.2.1. Works in preparation for publication

R Di Blasi*, S Furini, F Ceroni. Construct design shapes robustness to resource load in mammalian cells. *In preparation using results in Chapter 3.*

R Di Blasi*, T Ellis, F Ceroni. A burden shared is a burden halved: understanding resource competition to achieve predictable gene expression. *In preparation.*

R Di Blasi*, J Gabrielli*, F Ceroni. Resource competition for protein degradation resources in mammalian cells. *In preparation using results in Chapter 5.*

Chapter 2 - Resource-aware construct design in mammalian cells

2.1. Introduction

Two seminal works proved that competition for gene expression resources causes coupling of the expression of co-transfected genes in mammalian cells and proposed the implementation of adaptive regulatory circuits as a mitigation strategy[48,72]. However, these works led to different conclusions on what types of gene expression resources undergo limitation, and they did not provide information on how knowledge of resource limitations can aid the design of improved genetic constructs. This greatly limits our ability to rationally design genetic constructs with reduced resource footprints, with transversal implications for mammalian cell engineering. Given the wide use of co-transfected systems in foundational biology, bioproduction, and biotherapy, a coherent and in-depth investigation of resource competition in mammalian cells is awaited. This inspired the work presented in this Chapter.

This Chapter reports a consistent body of work to address key outstanding questions. First, we used transient capacity monitors to characterise the resource footprint of a library of genetic constructs designs for mammalian expression. The library included constructs bearing different promoters, translation initiation sites (Kozaks), and transcriptional termination sequences (polyA signals), leading to the identification of the resource bottlenecks in the gene expression pathway of engineered mammalian cells. We applied our findings to streamline the design of genetic constructs with minimised resource footprint and maximised performance, showing importance to bioproduction and biotherapy applications. Finally, we expanded our analysis to the design of multi-gene constructs and optimised the resource footprint of a genetic switch for use in mammalian cells.

Overall, this Chapter presents a coherent discussion of resource competition in co-transfected systems, showing how resource-aware frameworks provide new insights on resource allocation in mammalian cells. We also showed the feasibility of resource-informed genetic constructs design as a new mitigation strategy for mammalian cells, establishing an important milestone in the field. Data and figures presented in this Chapter have been published in Nature Communications under a Creative Commons Attribution 4.0 International Licence (https://creativecommons.org/licenses/by/4.0/)[151].

2.2. Results and discussion

2.2.1. Resource-aware framework to screen libraries of genetic constructs

Before beginning our investigation, we needed to develop a real-time method for easy and robust characterisation of resource competition in mammalian cells. Inspired by previous research[48,61,72], we used a fluorescence-based transient capacity monitor to quantify the demand imposed on cellular resources by a competing test plasmid – here referred to as *resource footprint*. When co-expressed in mammalian cells, the capacity monitor and the test plasmid compete for the same pool of gene expression resources. As the test plasmid consumes resources for gene expression, fewer resources are available for expressing the transient capacity monitor, leading to a decrease in intracellular capacity monitor fluorescence (Figure 2.1a). This way, the capacity monitor fluorescence informs on the resource footprint of a competing plasmid. Since we envisioned the testing of many test plasmid designs, we built a modular test plasmid architecture to easily generate a library of EGFP-expressing test plasmids with different promoters, Kozaks, and polyA signals variations. Each test plasmid design was individually co-transfected with the capacity monitor in HEK293T and CHO-K1, and intracellular fluorescence was recorded by flow cytometry two days after transfection. While the mKATE fluorescence from the capacity monitor reported on the resource footprint of the tested design, the EGFP fluorescence informed on its gene expression performance. Combining these two measures allowed us to identify test plasmid designs with optimal efficiency, defined as the simultaneous maximisation of the test plasmid expression – indicating high gene expression performance - and capacity monitor expression - indicating low resource footprint of the test design (Figure 2.1b).

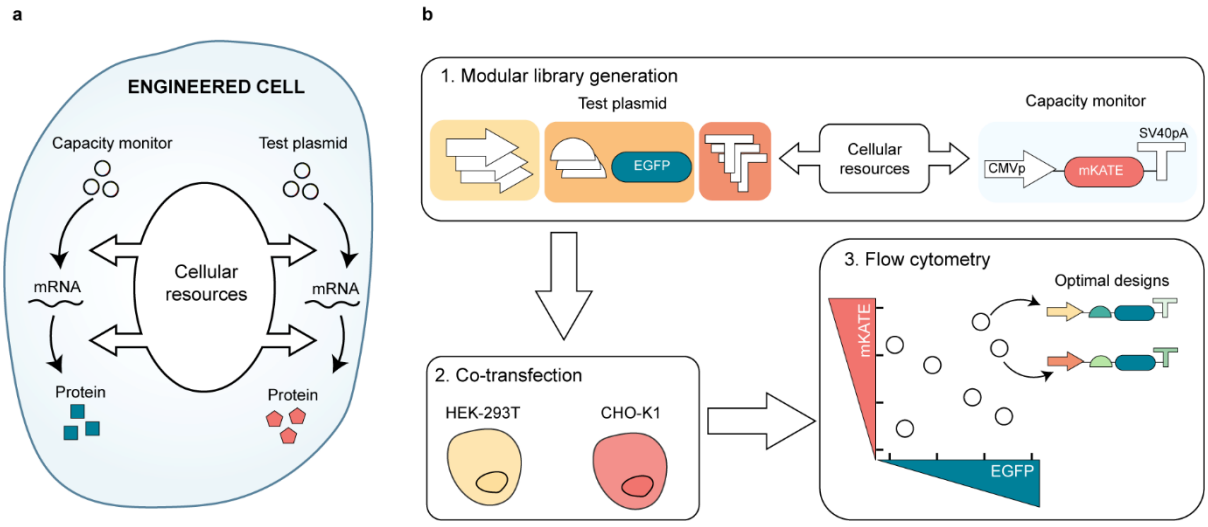

Figure 2.1 Resource-aware framework to screen libraries of genetic constructs. (a) A fluorescence-based transient capacity monitor can measure the resource footprint of a co-transfected test plasmid in mammalian cells. **(b)** Modular plasmid architecture used to generate a library of EGFP-expressing test plasmid designs bearing different promoters, Kozaks, and polyA signals. Each test plasmid design has been individually co-transfected with the capacity monitor in

HEK293T and CHO-K1. Intracellular EGFP and mKATE fluorescence were recorded two days post-transfection. Arrows are promoters, semi-circles are Kozaks, rounded rectangles are genes, and "T" represents polyA signals. This figure was taken from Di Blasi *et al.*[151] under a Creative Commons Attribution 4.0 International Licence (http://creativecommons.org/licenses/by/4.0/).

2.2.2. Capacity monitor sensitivity to resource loading

We first carried out a proof-of-concept experiment to test the sensitivity of the transient capacity monitor to test plasmid designs of increasing resource footprint. We co-transfected the capacity monitor with *(i)* an empty plasmid devoid of any coding sequence or regulatory element (pEmpty) as control of zero resource footprint, *(ii)* a constitutively-expressed tetracycline transactivator (rtTA), *(iii)* a TET-ON system in its uninduced state, or *(iv)* a fully induced TET-ON system (Figure 2.2a). The TET-ON system is a genetic switch composed of a constitutively-expressed rtTA that, in the presence of doxycycline (dox), binds to its cognate TETp promoter, promoting transcription. While the rtTA-expressing cassette *(ii)* and the uninduced TET-ON system *(iii)* are different in composition, they both only express the rtTA module and should therefore have similar resource footprints. However, the induced TET-ON system *(iv)* enables both constitutive expression of the rtTA module and dox inducible expression of the TETp-controlled EGFP and should – in theory – require more cellular resources.

Based on the capacity monitor fluorescence, we were able to quantify the resource footprint of the different constructs (Figure 2.2b-c). To make quantitative observations, we calculated the resource capacity for transient gene expression and the %output upon transfection of each test plasmid design (Figure 2.2c). Capacity for a specific test plasmid design was calculated as a percentage of the total capacity monitor expression measured upon co-transfection of the pEmpty (i.e. in the absence of competition). %output was calculated as a percentage of the total expression of the fully-induced TET-ON system. While capacity informed us on the resource footprint of each test plasmid design - with lower capacity indicating higher resource footprint -, the %output reported on the gene expression performance of each design - with lower %outputs indicating lower gene expression performance. In both HEK293T and CHO-K1, the capacity monitor expression faced a step-wise decrease as the resource demand of the test plasmids increased (Figure 2.2b). While co-transfection of test plasmids *(ii)* and *(iii)* expressing one transcriptional unit resulted in a decrease in capacity of a similar order of magnitude – with 44%-59% remaining capacity in HEK293T and 67% in CHO-K1 -, we observed a further reduction upon expression of the induced TET-ON system *(iv)*, with only 15% and 27% remaining capacity in HEK293T and CHO-K1, respectively (Figure 2.2b-c). This is consistent with the expression of two transcriptional units from the induced TET-ON system *(iv)*, as opposed to only one in *(ii)* and *(iii)*. Of note, we observed that CHO-K1 maintained higher capacity than HEK293T for equal test plasmids, suggesting that cell lines might have unique resource allocation profiles that ultimately result in different robustness to resource fluctuations (Figure 2.2b-c).

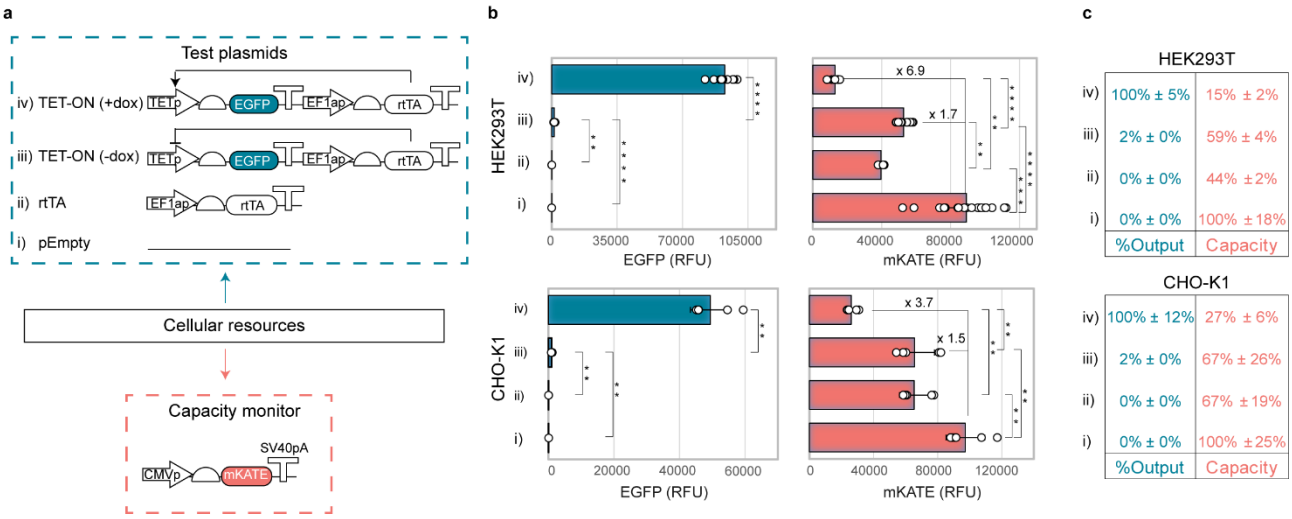

Figure 2.2 The resource footprint of genetic designs can be measured by co-transfection with a transient capacity monitor. **(a)** The transient capacity monitor was co-transfected with a library of genetic constructs with increasing resource footprints. These included: *(i)* a pEmpty, as control of zero resource footprint, *(ii)* an rtTA-expression cassette, *(iii)* an uninduced TET-ON system (dox 0ng/uL), *(iv)* an induced TET-ON system (dox 1ng/uL). **(b)** Test plasmid (EGFP) and capacity monitor (mKATE) expression levels in HEK293T (top panels) and CHO-K1 (bottom panels). **(c)** Remaining capacity and %output for transient gene expression for the test plasmid designs in HEK293T (top panel) and CHO-K1 (bottom panel). Test plasmid and capacity monitor expression levels are reported as mean RFU ± std. Number of independent experiments ≥ 3 (depending on the condition); two technical repeats per experiment. Two-sided Mann-Whitney test P value: ****<0.0001, ***<0.0005, **<0.005, *<0.05. Data analysis is described in Materials and methods (Chapter 7).

The coupling observed between the test plasmids and the capacity monitor in Figure 2.2 might be due to the efficient intracellular delivery of expression plasmids in high dosage, typical of transient transfections. Whether resource competition also arises for lower gene expression dosages is unclear. Should resource competition not be a problem at lower gene expression doses, a balance in the expression of different cassettes could be achieved by optimising their relative dosages, making the development of more complex mitigation strategies unnecessary and limiting the relevance of resource competition work in mammalian cells. Despite initial evidence demonstrating resource competition-based coupling at low plasmid doses[72], and the dosage of transfected plasmids used in Figure 2.2 being 10-50% of what is used in similar published works[48,72,82], we believe this is an important point. Therefore, we tested the effect of lower plasmid dosages on the resource footprint of the TET-ON system in HEK293T (Figure 2.3). Coupling effects were still present at the lowest tested plasmid dosage (i.e. 0.02pmol), leading to suppressed output of the capacity monitor following induction of the TET-ON system. Decreasing plasmid dosage led to suppressed output of both the test plasmid and the capacity monitor, proving successful modulation of intracellular plasmid copy number (Figure 2.3).

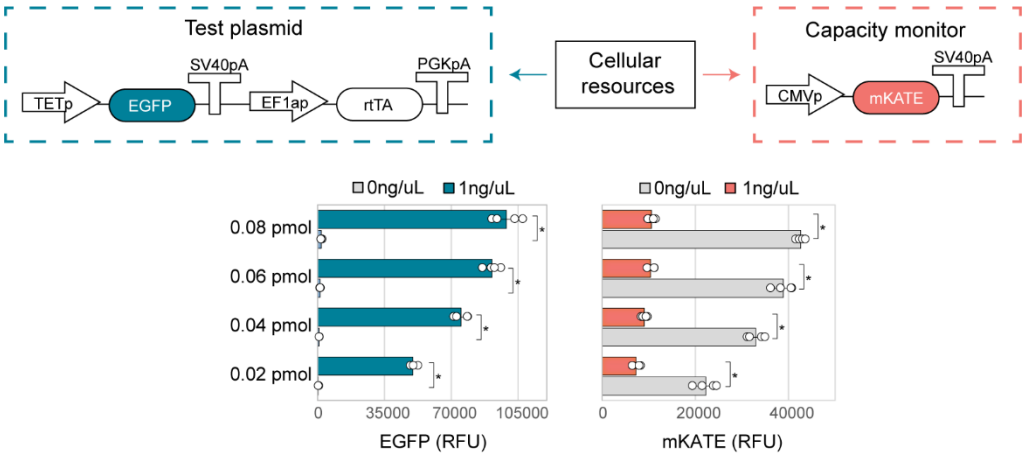

Figure 2.3 Resource competition al low gene expression dosage in HEK293T. Co-transfection of the TET-ON system and the capacity monitor at different plasmid dosages in HEK293T. Test plasmid (EGFP) and capacity monitor (mKATE) expression at uninduced (dox 0ng/uL) and induced (dox 1ng/uL) conditions. Test plasmid and capacity monitor expression levels are reported as mean RFU ± std. Number of independent experiments = 2; two technical repeats per experiment. Two-sided Mann-Whitney test P value: ****<0.0001, ***<0.0005, **<0.005, *<0.05. Data analysis is described in Materials and methods (Chapter 7).

To further corroborate our findings, we tested whether the capacity monitor could detect the resource footprint of a single-copy integrated gene. We used a HEK293T cell line with a single copy integration of a TET-ON inducible BFP (TETp-BFP), kindly gifted to us by Prof. Douglas Fowler[152] (Figure 2.4). This system is very similar to the one used in our previous experiments, enabling dox-regulated expression of a fluorescence gene, a BFP in this case. The TETp-BFP cell line was transfected with the capacity monitor plasmid and cultured in the presence or absence of dox. Induction of the genomically-integrated BFP led to decreased expression of the capacity monitor (Figure 2.4).

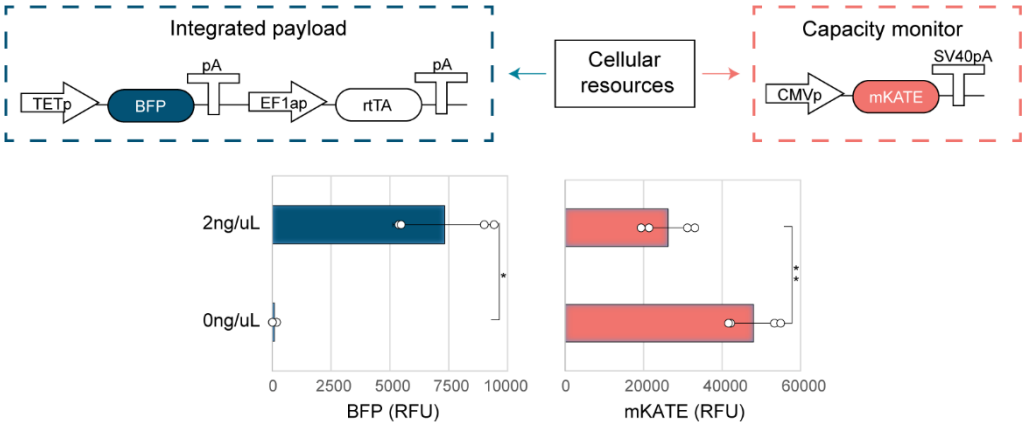

Figure 2.4 Resource footprint of a single copy integrated cassette. An integrated cell line with a single copy integration of a TET-ON-controlled BFP[152] was transfected with the capacity monitor. Integrated payload (BFP) and capacity monitor (mKATE) expression at uninduced (dox 0ng/uL) and induced (dox 2ng/uL) conditions. Integrated

payload and capacity monitor expression levels are reported as mean RFU ± std. Number of independent experiments = 2; two technical repeats per experiment. Two-sided Mann-Whitney test P value: ****<0.0001, ***<0.0005, **<0.005, *<0.05. Data analysis is described in Materials and methods (Chapter 7).

With this experiment, we proved for the first time that genes in single copies can interfere with the expression of a competing expression cassette. We also demonstrated that resource loading is not limited to co-transfected systems, opening to considerations on the effect of resource competition on integrated systems. Complete coverage of this topic will be given in Chapter 4.

2.2.3. Resource footprint of biological parts

In the previous section, we confirmed that transient capacity monitors effectively quantify the resource footprint from a competing expression cassette. Co-transfection of the capacity monitor with libraries of genetic constructs of diverse composition can inform how to design synthetic constructs with optimal efficiency (i.e. minimal resource footprint and maximised expression). We first focused on test plasmid designs differing for biological parts essential to mammalian gene expression systems, such as promoters, Kozak sequences, and polyA signals. Promoters control the binding of RNA polymerases and transcription factors, enabling transcription[153]; Kozak sequences function as translation starting sites in eukaryotes[154]; polyA signals are involved in transcriptional termination[155]. Based on their role, we reasoned that these three biological parts would represent critical parameters influencing the resource footprint of genetic constructs at the transcriptional and translational levels. We built a modular backbone architecture to easily generate a library of EGFP-expressing designs bearing seven constitutive promoters (pJB42CAT5, PGKp, hACTBp, SV40p, UBp, CMVp, EF1ap) and one inducible promoter (TETp); three Kozak sequences with different translational efficiency (Kz1, Kz2, and Kz3); and six polyA signals (SV40pA, SV40pA in reverse orientation – SV40pA_rv-, PGKpA, BGHpA, RBpA, HGHpA). The modular backbone could also accommodate the insertion of a second transcriptional unit, which is needed to clone inducible constructs (Figure 2.5a-b).

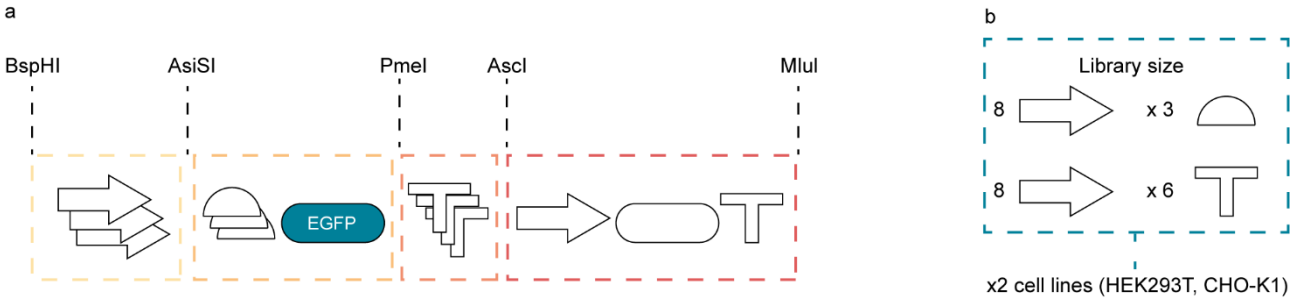

Figure 2.5 Modular backbone architecture and library size. (a) We designed a modular plasmid architecture to easily substitute biological parts of interest. This can accommodate a promoter flanked by BspHI and AsiSI restriction sites, a Kozak-EGFP flanked by AsiSI and PmeI restriction sites, a polyA signal flanked by PmeI and AscI restriction sites, and a second transcriptional unit flanked by AscI and MluI restriction sites. This way, biological parts can be substituted by digestion with combinations of restriction enzymes. **(b)** We aimed to clone 8x3 promoter-Kozak and 8x6 promoter-

polyA test plasmid combinations. Although we obtained most of these combinations, some did not clone correctly and were left out. Each design was then individually co-transfected with the capacity monitor in HEK293T and CHO-K1. Arrows are promoters, semi-circles are Kozaks, rounded rectangles are genes, and "T" represents polyA signals.

These three biological parts – promoters, Kozaks, and polyA signals - will be first presented in isolation to clarify their effect on resource footprint. Then the library will be discussed in its entirety, opening to considerations on informed construct design.

2.2.3.1. Promoters

To analyse the contribution of promoter sequences to the resource footprint of genetic constructs, we cloned a library of test plasmids, each bearing one of eight promoter configurations. We then co-transfected individual test plasmids with the transient capacity monitor (Figure 2.6a). Although our promoter library effectively tuned EGFP expression up to approximately two logarithmic decades in HEK293T, we observed that the rank of the promoters across cell lines was not maintained (Figure 2.6b-c). For instance, PGKp and SV40p were weaker promoters than hACTBp in HEK93T but behaved as stronger promoters in CHO-K1. This reinforces previous reports showing the variability of promoter efficiency across cell lines[45,46]. We also noticed that stronger promoters had higher resource footprints than weaker promoters (Figure 2.6b-c), and we found a negative linear correlation between the test plasmid and the capacity monitor expression levels in both cell lines (Figure 2.6d).

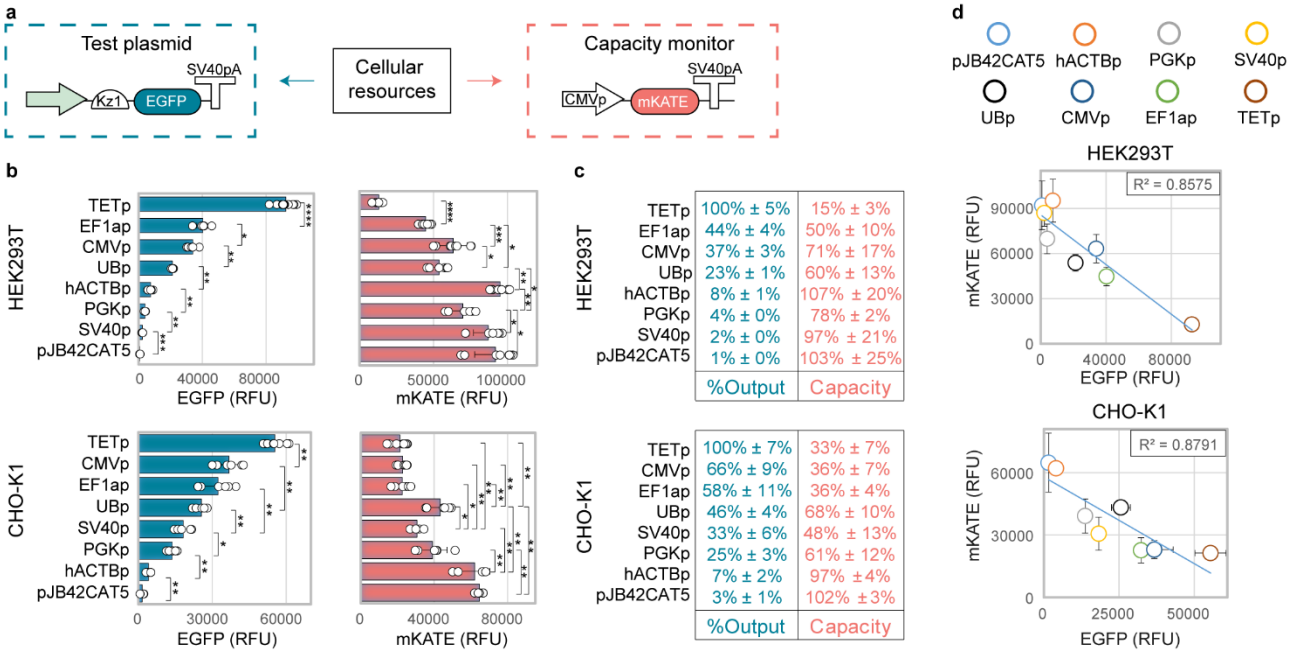

Figure 2.6 Promoters impact the resource footprint of genetic constructs. (a) We cloned a library of test plasmids, each carrying one of eight promoter configurations (dox 1ng/uL was used to induce the expression of the TETp). Each test plasmid was co-transfected with the capacity monitor in HEK293T and CHO-K1. **(b)** Test plasmid (EGFP) and capacity monitor (mKATE) expression in HEK293T (top panels) and CHO-K1 (bottom panels). **(c)** %output and remaining capacity for the test plasmid designs in HEK293T (top panel) and CHO-K1 (bottom panel). **(d)** Negative linear

correlation between the test plasmid output (EGFP) and the capacity monitor output (mKATE) in HEK293T (top panel) and CHO-K1 (bottom panel). Pearson coefficient -0.926 and -0.861 in HEK293T and CHO-K1, respectively. Test plasmid and capacity monitor expression levels are reported as mean RFU ± std. Number of independent experiments ≥ 3 (depending on the condition); two technical repeats per experiment. Two-sided Mann-Whitney test P value: ****<0.0001, ***<0.0005, **<0.005, *<0.05. Data analysis is described in Materials and methods (Chapter 7).

These findings suggest that promoters are crucial in shaping the resource footprint of genetic constructs, with stronger promoters requiring more cellular resources than weaker ones. As an exception to this general behaviour, we found promoter sequences that are more efficient than others, allowing higher test plasmid expression for a lower resource footprint (i.e. higher capacity monitor expression). For instance, in CHO-K1, UBp showed higher test plasmid outputs than SV40p – 46% and 33% of %output, respectively - at a reduced resource footprint – 68% and 48% remaining capacity, respectively (Figure 2.6c). Similarly, CMVp is more efficient than UBp in HEK293T (Figure 2.6c).

Similar to testing a library of promoters, titration of the inducer concentration results in different transcriptional rates from the TETp in a TET-ON system (Figure 2.7a). As expected, increasing the dox concentration led to higher test plasmid expression and increased resource footprint (Figure 2.7b-c). This is in accordance with our previous results in Figure 2.6 and indicates that increasing the transcriptional rate has a resource cost for mammalian cells. Based on our results, we conclude that one or more bottlenecks involving gene expression resources are present in engineered mammalian cells, in accordance with previous findings[48,72]. Since changes in transcriptional rates also modify resource consumption of downstream translational and post-translational processes, we cannot be sure as to which types of resources are undergoing limitations. This aspect will be further explored in this Chapter, and new insights will be provided.

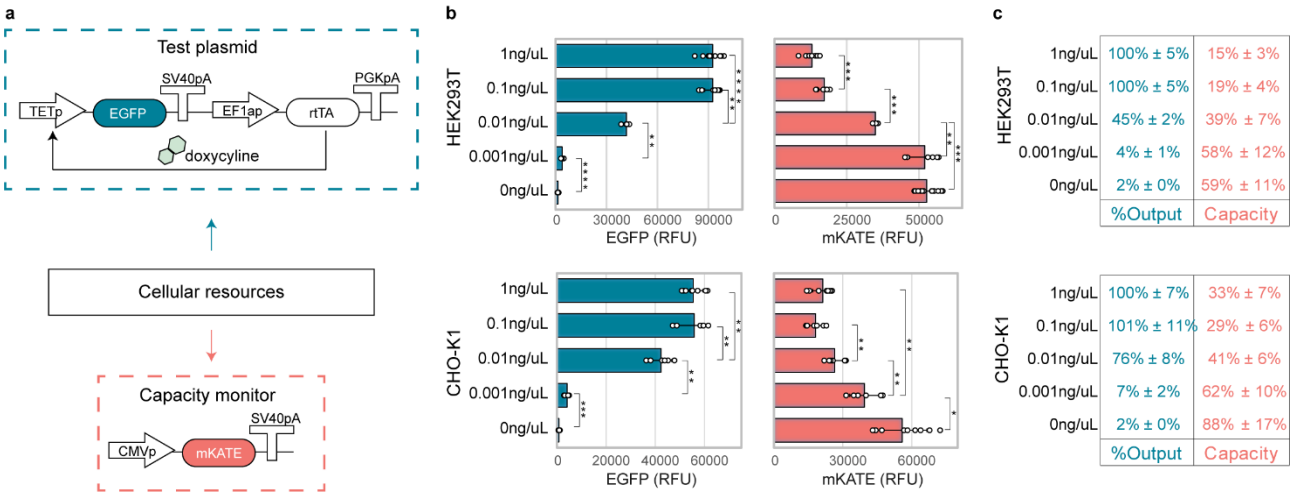

Figure 2.7 Resource footprint of a TET-ON system at different inducer concentrations. (a) We tested the TET-ON system in competition with the capacity monitor at different inducer (dox) concentrations. **(b)** Test plasmid (EGFP) and capacity monitor (mKATE) expression at different dox concentrations in HEK293T (top panels) and CHO-K1 (bottom

panels). **(c)** %output and remaining capacity for the tested conditions in HEK293T (top panel) and CHO-K1 (bottom panel). Test plasmid and capacity monitor expression levels are reported as mean RFU ± std. Number of independent experiments ≥ 3 (depending on the condition); two technical repeats per experiment. Two-sided Mann-Whitney test P value: ****<0.0001, ***<0.0005, **<0.005, *<0.05. Data analysis is described in Materials and methods (Chapter 7).

To prove that the behaviour we observed was not dependent on the specific capacity monitor design, we tested three additional transient capacity monitors in which the mKATE is controlled by different promoters, namely SV40p, UBp, and EF1ap (Figure 2.8). When tested with the promoter library, the additional monitor designs showed similar behaviour to the original capacity monitor, with increasing transcriptional rate from the test plasmid leading to decreased capacity monitor expression. Of note, upon co-transfection of the UBp-mKATE capacity monitor and the UBp-EGFP test plasmid in HEK293T, we observed an unexpectedly high resource footprint compared to what would be inferred from the test plasmid output (Figure 2.8, red arrow). This was not observed in CHO-K1 or for other pairs of competing identical promoters (i.e. CMV-mKATE vs CMV-EGFP, SV40p-mKATE vs SV40p-EGFP, EF1ap-mKATE vs EF1ap-EGFP) in HEK293T. Such behaviour can be explained by a resource bottleneck resulting from the saturation of UBp-specific transcription factors in HEK293T but not in CHO-K1.

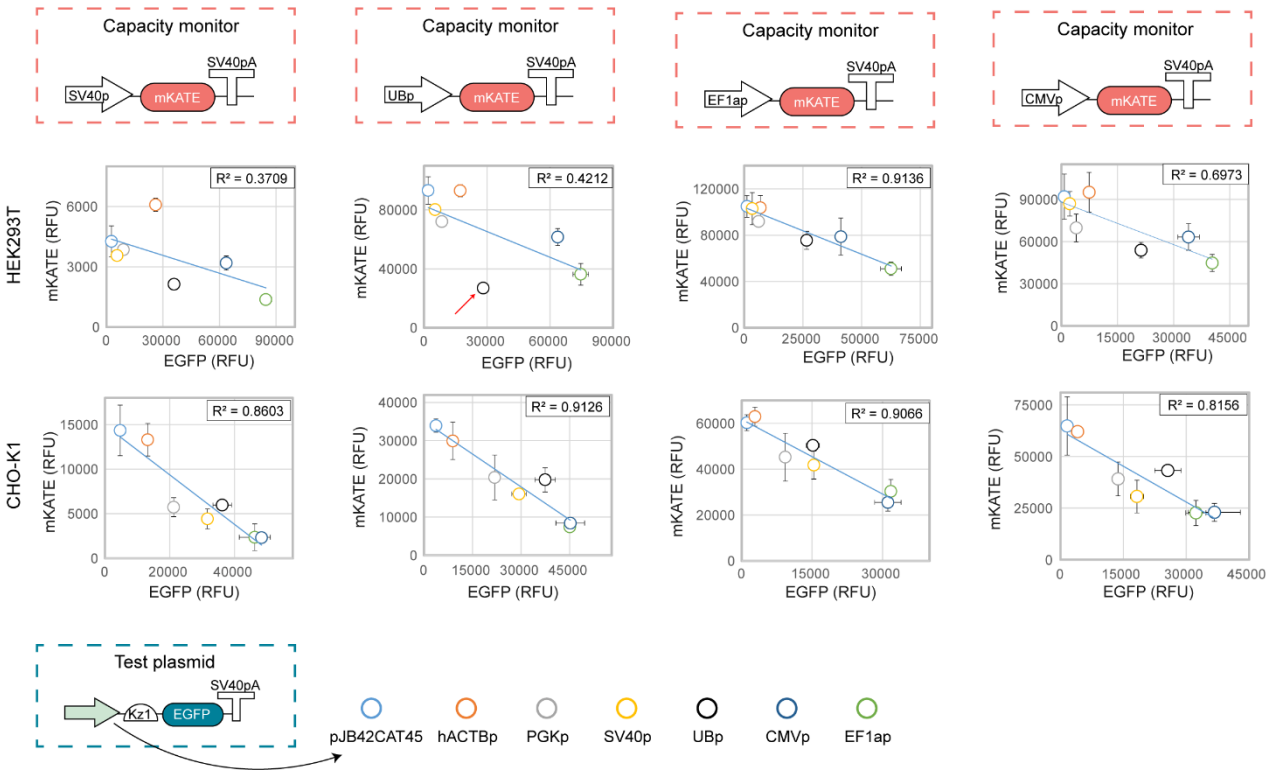

Figure 2.8 Testing of four capacity monitor designs with different promoters. We tested the resource footprint of a library of test plasmids, each carrying one of seven promoter configurations, with four capacity monitor designs in HEK293T (top panels) and CHO-K1 (bottom panels). The red arrow indicates competition between UBp-mKATE (capacity monitor) and UBp-EGFP (test plasmid). Test plasmid and capacity monitor expression levels are reported as

mean RFU ± std. Number of independent experiments ≥ 2 (depending on the condition); two technical repeats per experiment. Data analysis is described in Materials and methods (Chapter 7).

This adds to the basal competition for gene expression resources caused by the co-expression of the test plasmid and the capacity monitor, exacerbating the coupling effects between the two. Further work is needed to prove this claim. Nevertheless, our findings provide a clear example of how promoter-specific effects add complexity to resource competition profiles, calling for a thorough characterisation of resource allocation and abundance in different cell lines.

2.2.3.2. PolyA signals

PolyA signals play a crucial role in eukaryotic gene expression, directly mediating transcriptional termination and indirectly modulating mRNA transport, stability, and translation by adding polyA tails[156]. We assembled six polyA signals (SV40pA, SV40pA_rv, PGKpA, BGHpA, RBpA, HGHpA) in combination with six promoters (PGKp, SV40p, hACTBp, UBp, CMVp, EF1ap) and co-transfected each test variant with the capacity monitor in HEK293T and CHO-K1 (Figure 2.9a). Although we tried to obtain all possible promoter-polyA combinations, some variants were not cloned successfully and were left out of the analysis (i.e. SV40p-BGHpA, PGKp-BGHpA, hACTBp-BGHpA, UBp-BGHpA, EF1ap-BGHpA, UBp-HGHpA). When paired with the same promoter, changing the polyA signal dramatically impacted the expression of the upstream EGFP (Figure 2.9b, Figure 8.1). This could be due to multiple reasons, including differences in transcriptional termination, mRNA stability, and translational efficiency resulting from the specific polyA signal in use. Although other works have highlighted polyA-dependent modulation of gene expression in mammalian cells[157], by assessing the effect of polyA signals on the expression of a competing capacity monitor, we can uncover unique insights on their resource footprint. In HEK293T, we observed distinct resource footprints for the six polyA signals when paired with any promoters (Figure 2.9b, Figure 8.1). However, in CHO-K1, interference with the capacity monitor expression mediated by polyA signals was mostly observed when combined with strong promoters, such as CMVp and EF1ap (Figure 2.9b, Figure 8.1). When considering all the genetic designs together, we noticed a negative linear correlation between the test plasmid and capacity monitor expression levels only in CHO-K1 (Figure 2.9c). This is due to some polyA signals consistently resulting in suppressed test plasmid output while at the same time having a high resource footprint in HEK293T. For example, PGKpA and SV40pA have poor gene expression performance and a high resource cost for all promoter combinations, as indicated by suppressed EGFP and mKATE outputs. The absence of correlation in HEK293T implies that some promoter-polyA combinations are more efficient than others in HEK293T.

As stated before, polyA signals are involved in many cellular processes, regulating gene expression and interacting with many regulatory proteins. Understanding the effect of resource competition imposed by polyA signals in these scenarios is not trivial, as the observed behaviour is likely to result from the

interaction of many cellular players, all capable of facing shortage when heterologous proteins are overexpressed. In addition, cell lines are likely to have unique resource availability and allocation profiles, providing a speculative explanation for the variability in behaviour across cell lines.

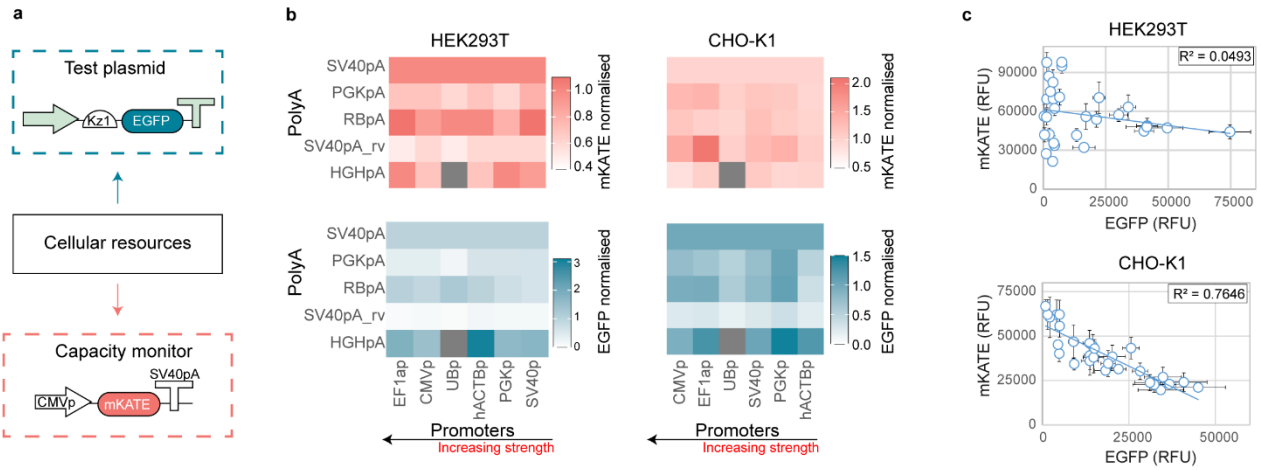

Figure 2.9 Resource footprint of promoter-polyA combinations. (a) We assembled six promoters in combination with six polyA signals and co-transfected each test plasmid variant with the capacity monitor in HEK293T and CHO-K1. **(b)** Test plasmid (EGFP) and capacity monitor (mKATE) normalised expression levels. Normalised expression levels were calculated by dividing the average RFU for each combination against the average RFU of the combination with equal promoter and SV40pA (i.e. EF1ap-HGHpA was normalised against EF1ap-SV40pA; hACTB-PGKpA was normalised against hACTB-SV40pA, and so on). The grey square indicates the UBp-HGHpA combination, which did not clone correctly and was left out of the analysis. BGHpA was excluded from these heatmaps because we only obtained it in combination with CMVp; data for this sample can be found in Figure 8.1. **(c)** A negative linear correlation between the test plasmid output (EGFP) and the capacity monitor output (mKATE) was observed in CHO-K1 (bottom panel) but not in HEK293T (top panel). Pearson coefficient -0.222 and -0.874 in HEK293T and CHO-K1, respectively. Test plasmid and capacity monitor expression levels are reported as mean RFU ± std unless otherwise specified. Number of independent experiments ≥ 3 (depending on the condition); two technical repeats per experiment. Data analysis is described in the Materials and methods (Chapter 7).

To gain more insights on the role of polyA signals in resource competition, we cloned four additional transient capacity monitor designs, each bearing one of four polyA signals, namely PGKpA, RBpA, SV40pA_rv, and HGHpA. We tested the novel monitor designs in competition with shortlisted test designs bearing EF1ap-Kz1-EGFP and one of five polyA signals previously adopted. In CHO-K1, the competition experiment with the novel capacity monitors gave similar results to our initial dataset in Figure 2.9, with increasing test plasmid outputs leading to decreased capacity monitor expression levels. However, in HEK293T, we observed different behaviours based on the polyA signal. While increasing test plasmid outputs led to increased capacity monitor expression for the SV40pA and SV40pA_rv capacity monitor designs, the opposite trend was observed for the other polyA designs (Figure 2.10). Speculating on the cause of these differences in behaviour is difficult without further experiments. Nevertheless, we provide

the first evidence of sequence and cell line dependencies of polyA-based resource competition profiles. This introduces an additional layer of complexity to the current understanding of resource competition in mammalian cells, adding polyA signals to promoter elements as key biological parts influencing the resource footprint of genetic constructs.

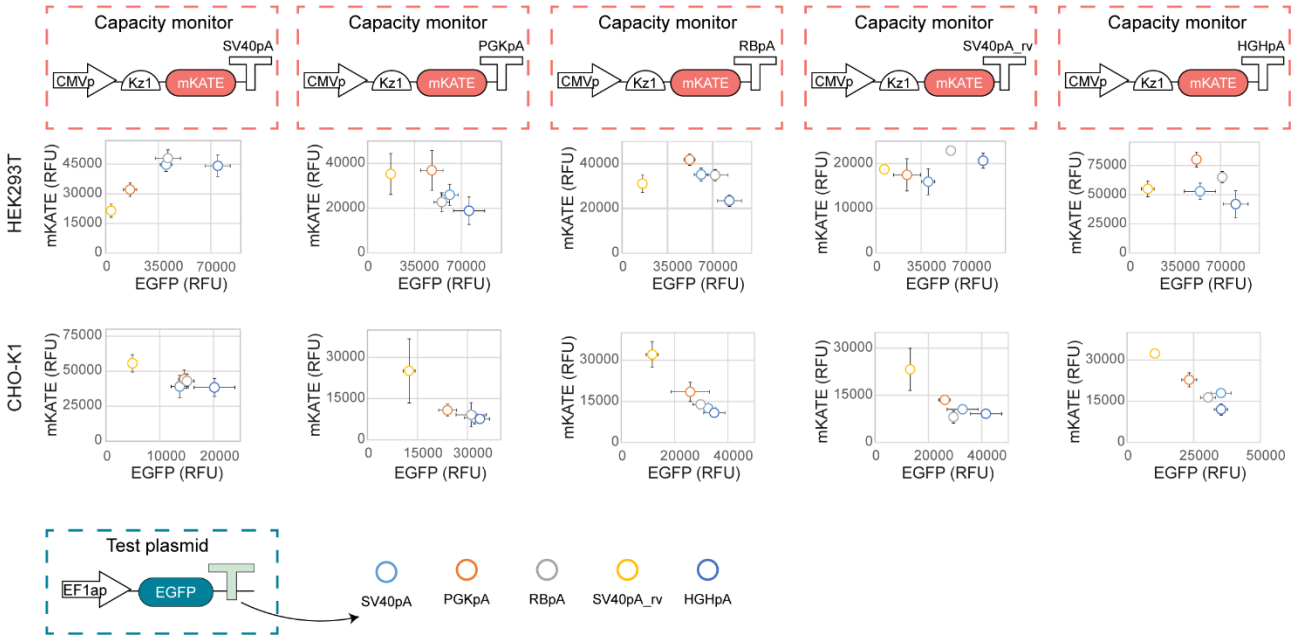

Figure 2.10 Testing four capacity monitor designs with different polyA signals. We evaluated the resource footprint of test constructs, each containing one of five polyA configurations, with four capacity monitor designs bearing different polyA signals in HEK293T (top panels) and CHO-K1 (bottom panels). Test plasmid (EGFP) and capacity monitor (mKATE) expression levels are reported as mean RFU ± std. Number of independent experiments = 2; two technical repeats per experiment. The exact number of repeats for each sample is reported in Di Blasi et al.[151]. Data analysis is described in Materials and methods (Chapter 7).

2.2.3.3. Kozak sequences

Although both promoters and polyA signals impact the resource footprint of genetic designs, it is unclear whether they do this by overloading transcriptional resources or impacting downstream translational and post-translational resources. Contrasting evidence exists on the importance of transcriptional and translational resources in shaping resource competition profiles in mammalian cells[48,72]. To settle this controversy, we analysed the impact of Kozak sequences on the resource footprint of genetic constructs.

Kozak sequences are directly linked to translational resources as they represent the translation start sites in eukaryotes. Kozak motifs with different translational efficiency are often adopted to tune gene expression outputs[158,159]. We selected three previously reported Kozak sequences - the Kozak consensus (Kz1) and two Kozak variants with low and high translational efficiencies (Kz2 and Kz3, respectively) - and tested their behaviour when assembled downstream of the promoter library (Figure 2.11a)[159]. While Kz1 and Kz3 showed similar translational efficiencies for most promoter combinations, as evidenced by the test plasmid

outputs, Kz2 dramatically suppressed EGFP expression (Figure 2.11b-c, Figure 8.2). When comparing promoter-Kozak combinations, most of the variation in resource footprint was explainable by the change of promoter sequence. A moderate impact of Kozak sequences on resource footprint was mostly observed in combination with strong promoters (Figure 2.11b-c, Figure 8.2). This should be evident from Figure 2.11b-c, where we plotted few selected combinations to highlight the relative importance of promoters and Kozak sequences in shaping the resource footprint of genetic constructs. These findings prove that Kozak sequences only minorly impact the resource footprint of genetic constructs, ultimately suggesting that translational resources do not undergo significant limitations in our experiments. If translational resources were the limiting factor in our experiments, we would observe a more pronounced impact on the capacity monitor expression upon dramatic changes in the translational efficiency of the test plasmid. However, the impact of Kozak sequences on resource footprint was mainly observed when paired with strong promoters. This reinforces our conclusion, as translational resources start to be saturated only upon sustained mRNA production from strong promoters. Ultimately, our findings indicate that transcriptional resources are more limiting than translational resources in the tested cell lines. Specifically, the latter include any resources acting downstream of ribosome recruitment (i.e. ribosomes, tRNAs, aminoacyl-tRNA synthetases, aminoacids, etc.).

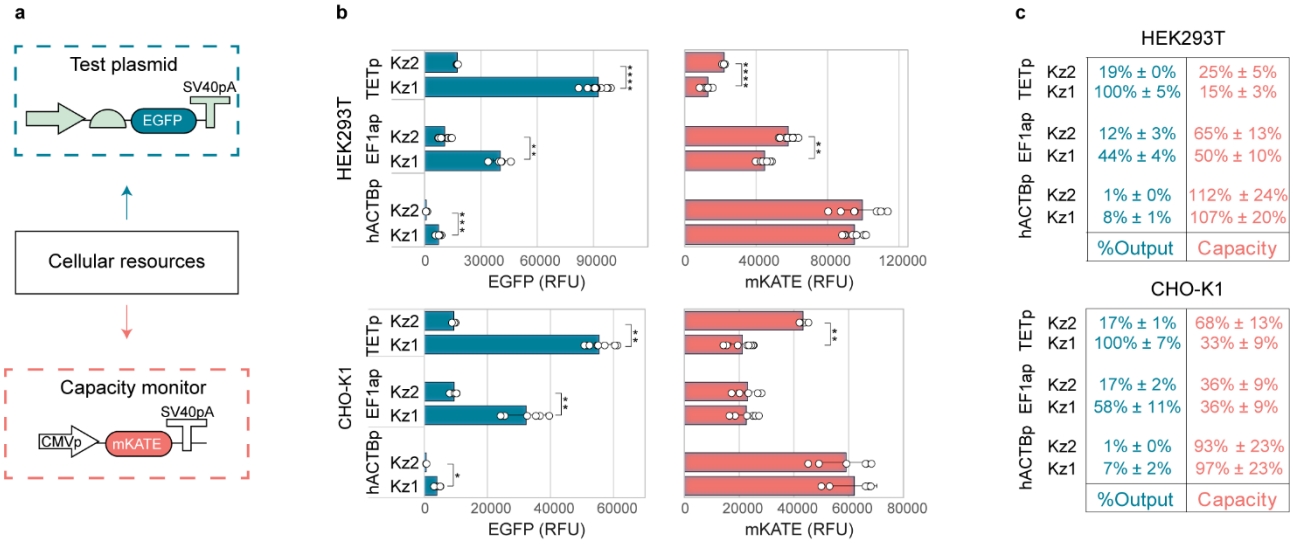

Figure 2.11 Resource footprint of promoter-Kozak combinations. (a) We assembled eight promoters in combination with three Kozak sequences and co-transfected each test plasmid variant with the capacity monitor in HEK293T and CHO-K1. **(b)** Test plasmid (EGFP) and capacity monitor (mKATE) expression of six shortlisted promoter-Kozak combinations in HEK293T (top panels) and CHO-K1 (bottom panels). The full dataset can be found in Figure 8.2 **(c)** %output and remaining capacity for the tested combinations in HEK293T (top panel) and CHO-K1 (bottom panel). Test plasmid and capacity monitor expression levels are reported as mean RFU ± std. Number of independent experiments ≥ 3 (depending on the condition); two technical repeats per experiment. Two-sided Mann-Whitney test P value: ****<0.0001, ***<0.0005, **<0.005, *<0.05. Data analysis is described in Materials and methods (Chapter 7).

Since we believe this is an important finding, we obtained further proof by performing additional experiments. First, we used online tools to design three 5'UTR RNA hairpins of increasing minimum free energy (MFE) – H1, H2, and H3, respectively - that work by sequestering the Kozak sequences (please refer to Materials and methods – Chapter 7 - for information on how these hairpins were designed). Since 5'UTR RNA structures have been shown to tune translational efficiency in mammalian cells[160], we thought this would be a good complementary approach to validate our previous results. The three RNA structures were cloned at the 5' UTR of EF1ap-Kz1-SV40pA and worked as expected, with hairpins of increasing MFE causing a higher suppression of the test plasmid translation (Figure 2.12a). Comparing the strongest hairpin (EF1ap + H3), which almost completely repressed the test plasmid expression, with the no-hairpin construct (EF1ap), we only observed a minor rescue in the capacity monitor expression, in accordance with our previous Kozak results (Figure 2.12a). More importantly, when similar changes in test plasmid expression were achieved by either changing promoter sequence or introducing hairpin structures, the impact on resource footprint was very different (Figure 2.12a, red asterisks). For instance, the test plasmids EF1ap+H2 and hACTBp had similar EGFP expression levels but very different resource footprints. In other words, while changing the EF1ap promoter to a weaker hACTBp dramatically decreased the resource footprint, introducing hairpin H2 to EF1ap did not change the resource footprint significantly (Figure 2.11, red asterisks). These results reiterate that transcriptional resources are more limiting than translational resources in engineered HEK293T and CHO-K1, as the same variation in test plasmid output causes *(i)* decreased resource footprint when transcriptional resources are released (i.e. by changing the strong EF1ap to a weaker hACTBp), *(ii)* minimal impact on the resource footprint when translational resources are released (i.e. by introducing RNA hairpins in the 5' UTR to block translation).

As a second approach, we performed RNA transfection experiments. We reasoned that transfecting RNA instead of DNA would allow us to isolate the impact of genetic constructs on translational resources. We set up a protocol for *in vitro* transcription of capped and poly-adenylated mammalian mRNAs using commercially-available kits (please refer to Materials and methods – Chapter 7 - for information on the protocol). We synthesised two test mRNA bearing Kozak sequences with different translational efficiency (Kz1-EGFP and Kz2-EGFP) and one monitor mRNA (Kz1-mKATE). We co-transfected each test mRNA individually with the capacity monitor in HEK293T, and, despite dramatic differences in the EGFP outputs of the test mRNAs, we observed no alteration in the capacity monitor expression levels (Figure 2.12b). Altogether, our results convincingly demonstrate that translational resources are not limiting in the engineered HEK293T and CHO-K1. This is in striking contrast with what was previously reported in engineered bacteria, where translational resources play a major role in shaping resource competition profiles[61,68].

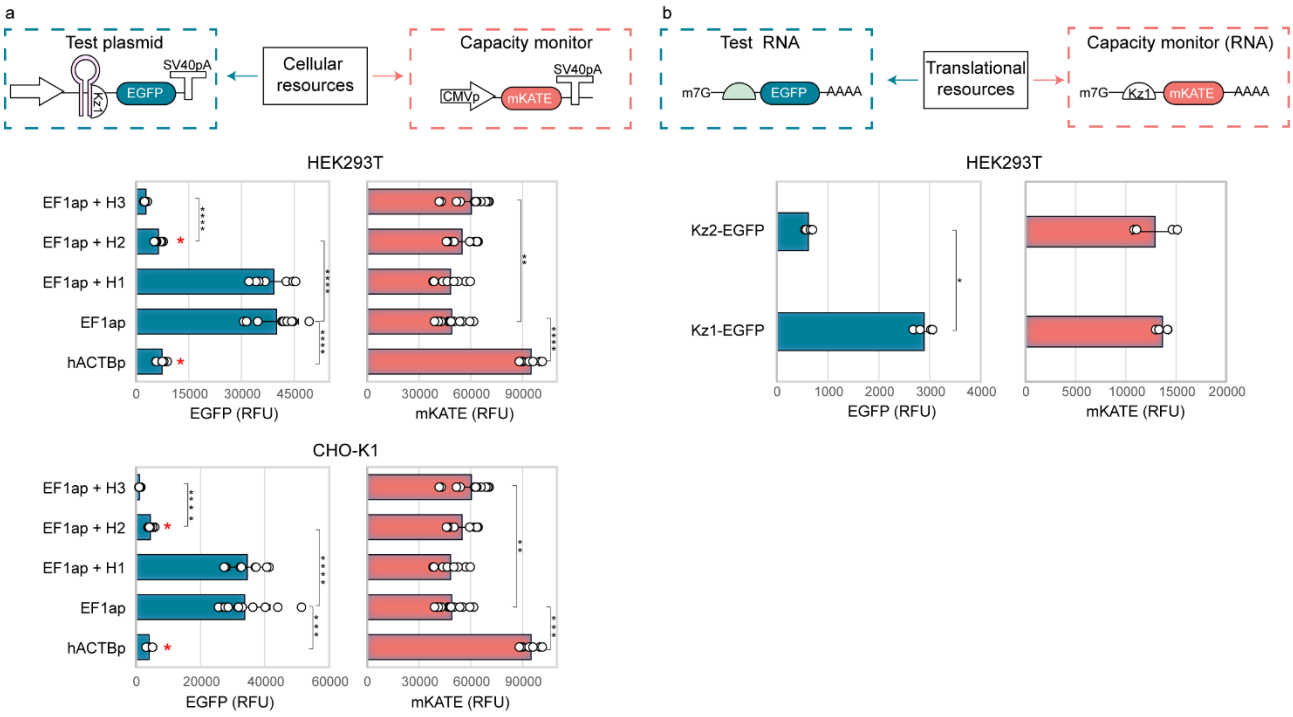

Figure 2.12 Translational resources are not limiting in engineered mammalian cells. (a) Three hairpins (H1, H2 and H3) were cloned at the 5`-UTR of the EF1ap-Kz1-EGFP-SV40pA construct. Test plasmid (EGFP) and capacity monitor (mKATE) expression levels in HEK293T (top panels) and CHO-K1 (bottom panels). Red asterisks identify two constructs with similar test plasmid outputs: one with a stronger promoter and H2, and one with a weaker promoter and no hairpin. **(b)** Two test mRNAs (Kz1-EGFP and Kz2-EGFP) and one monitor mRNA (Kz1-mKATE) were generated by *in vitro* transcription. All mRNAs were capped and polyadenylated. Test plasmid (EGFP) and capacity monitor (mKATE) expression levels in HEK293T. Test plasmid and capacity monitor expression levels are reported as mean RFU ± std. Number of independent experiments ≥ 2 (depending on the condition); two technical repeats per experiment. Two-sided Mann-Whitney test P value: ****<0.0001, ***<0.0005, **<0.005, *<0.05. Data analysis is described in Materials and methods (Chapter 7).

2.2.4. Resource-aware construct design identifies alternatives with optimal efficiency

Combining all the library designs can clarify the overall resource competition landscape and inform on the presence of more efficient design options (Figure 2.13a). Although we observed that greater output from the test plasmids was generally matched by decreased expression of the capacity monitor – hence, increased resource footprint -, some designs appeared more efficient than others. This enabled the identification of design regions where resource-aware design can be implemented towards optimal expression performance at a minimal resource cost (Figure 2.13b, shaded areas). We then conducted a pairwise comparison of genetic constructs differing in one of the three design elements considered in this study (promoters, Kozaks, or polyA signals) (Figure 2.13c). We measured the differences in EGFP and mKATE (ΔEGFP and ΔmKATE) between each pair of genetic constructs bearing *(i)* different promoters but identical Kozaks and polyA signals; *(ii)* different Kozaks but identical promoters and polyA signals; and *(iii)* different polyA signals but identical promoters and Kozaks. These pairs were then grouped into quartiles

based on their ΔEGFP, revealing how similar changes in the test plasmid expression performance come at different resource costs depending on the type of part being modified (Figure 2.13c). In both cell lines, the promoter is the genetic component with the greatest impact on resource footprint. Kozak efficiency minorly alters a construct's resource footprint, and only for high ΔEGFP values, in accordance with our prior observations. The role of polyA signals is cell-line dependent, and while in HEK293T, positive ΔEGFP across all quartiles led to decreased resource footprint (i.e. positive ΔmKATE), in CHO-K1, we observed the opposite behaviour. We noted that while changing the three design elements can lead to the same gain in test plasmid expression, promoters do so at a higher resource cost than Kozaks and polyA signals (Figure 2.13d-e). These insights are of great relevance for the cell engineering community as they represent higher-order design rules enabling the generation of genetic constructs with optimal efficiency. For instance, increasing gene expression performance by changing Kozaks or polyA signals is a better strategy than promoter modification, resulting in the generation of higher-efficiency genetic designs.

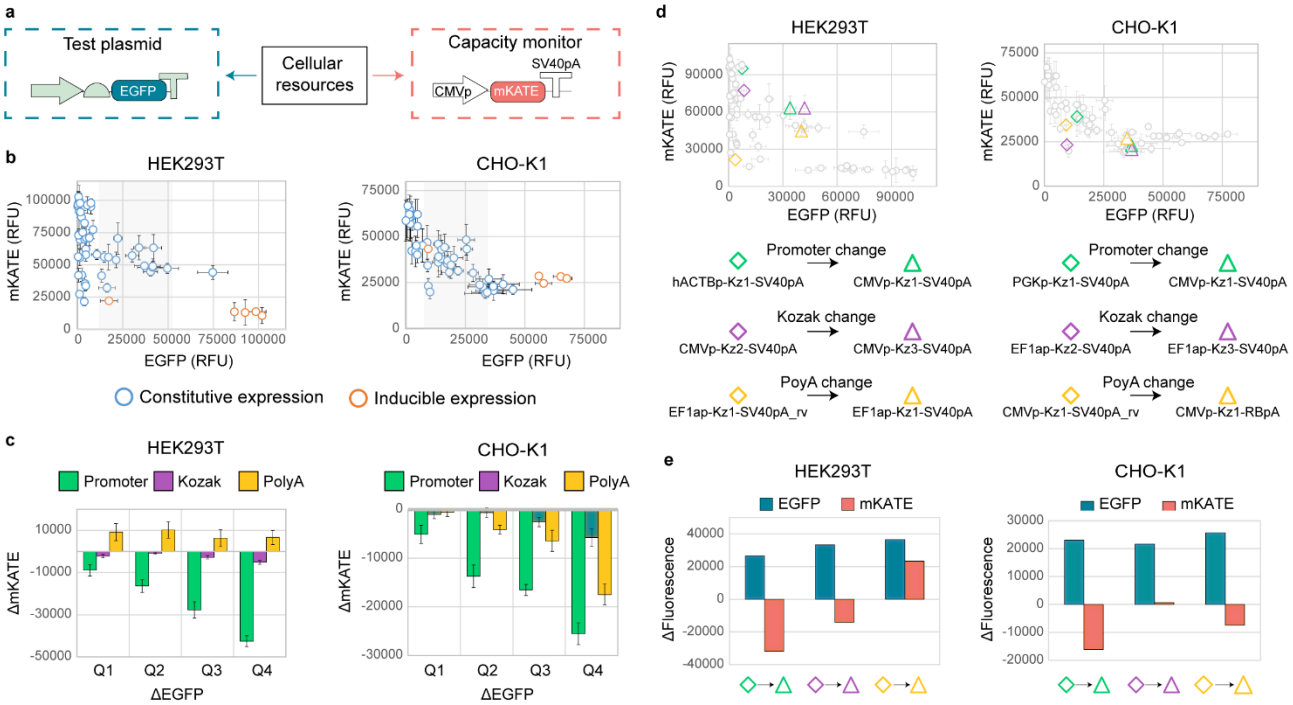

Figure 2.13 Resource-aware construct design identifies alternatives with optimal efficiency. (a) We co-transfected each test plasmid of the library individually with the capacity monitor in HEK293T and CHO-K1. **(b)** Test plasmid (EGFP) and capacity monitor (mKATE) expression levels for the test plasmid library in HEK293T (left panel) and CHO-K1 (right panel). Each dot represents one test plasmid design. Shaded areas indicate regions where test designs with optimal expression performance and reduced resource footprint can be identified. **(c)** Contribution of promoters, Kozaks and polyA signals to resource footprint. Vertical bars show the average change in mKATE (ΔmKATE) in each ΔEGFP quartile due to changing promoter, Kozak, or polyA. ΔmKATE and ΔEGFP due to promoter change were obtained by comparing all pairs of circuits with identical Kozak and polyA but different promoter sequences. The same method was used to define ΔmKATE and ΔEGFP due to Kozak and polyA modification. This analysis was performed by Simone Furini, a co-author of the paper published in Nature Communications[151]. **(d)** Similar variations in the test plasmid gene expression

performance (EGFP) achieved by changing promoter, Kozak or polyA have a different impact on the capacity monitor expression levels (mKATE). **(e)** ΔFluorescence for the construct pairs in Figure 3d. Test plasmid (EGFP) and capacity monitor (mKATE) expression levels in HEK293T. Test plasmid and capacity monitor expression levels are reported as mean RFU ± std. Number of independent experiments ≥ 2 (depending on the condition); two technical repeats per experiment. Data analysis is described in Materials and methods (Chapter 7).

2.2.4.1. Optimisation of mammalian co-expression systems

Our findings can prove very useful for all those applications requiring the co-expression of genetic constructs. One prime example is the mammalian-based production of therapeutic proteins relying on the expression of two or more genes (such as antibodies and virus-like-particles), which will inevitably compete for cellular resources. However, so far, there is no evidence that the design rules that we highlighted in previous discussions hold true when the EGFP is replaced with other genes of interest. Should they not hold, this would greatly limit the usefulness of our resource-aware framework, as an initial *ad hoc* characterisation would be needed for any gene that one might want to use.

To address this major concern, we selected eight test designs based on our previous characterisation (Figure 2.13a) and replaced the EGFP with the Epstein-Barr virus nuclear antigen 1 (EBNA1) gene (Figure 2.14a). EBNA1 is a viral gene often adopted in bioprocessing and gene therapy applications to improve the efficiency of plasmid transfections[161–166]. The eight EBNA1 designs were individually co-transfected with the capacity monitor in HEK293T. We kept the capacity monitor in our experiments to have an easy way of assessing the resource footprint of the EBNA1 designs. However, we stress that the capacity monitor could be replaced by any gene of commercial or therapeutic value. Although we observed no correlation between the EBNA1 and EGFP expression levels of the selected designs, this is in accordance with previous work outlining the lack of modularity when the genetic context is changed[167] (Figure 2.14b, Figure 8.3). However, the expression levels of the capacity monitor did show a good correlation when in competition with either EGFP or EBNA1 (Figure 2.14b). This indicates that the capacity monitor output is a good predictor of the resource footprint of genetic designs independently of the gene in use, thus confirming the validity and translatability of our resource-aware framework.

Five of the eight shortlisted test designs were selected to yield similar test plasmid output while displaying different resource footprints (Figure 2.14c – left panel). In accordance with our previous characterisation, the EBNA1-expressing design carrying strong promoters resulted in a higher resource footprint than the design with weaker promoters, with 13% remaining capacity for TETp, 53% for CMVp, and 68-85% for the UBp and hACTBp designs (Figure 2.14c-d). Importantly, based on our previous EGFP characterisation, we managed to predictably design EBNA1-expressing constructs with reduced resource footprint, increasing the expression of the co-transfected capacity monitor by 6.7 times (i.e. compare TETp-Kz-SV40pA and UB-Kz1-SV40pA). As discussed before, the EBNA1 expression levels of the five designs did not match our

predictions based on the previous EGFP characterisation. The two constructs skewing our predictions both carry the UBp, suggesting that combining UBp and EBNA1 might be particularly performant in promoting gene expression. Consequently, the two UBp designs display decreased resource footprint and increased EBNA1 expression output, achieving co-optimisation of both competing cassettes.

Finally, four designs bearing CMVp-Kz1 and one of four polyA configurations were selected to yield different test plasmid outputs while maintaining similar resource footprints (Figure 2.14e-f). In line with previous findings, the change of polyA to more favourable configurations allowed for higher EBNA1 production without increasing the resource footprint.

Overall, by applying our framework to the design of EBNA1-expressing constructs, we were able to generate synthetic constructs with *(i)* increased EBNA1 levels at a fixed resource footprint, *(ii)* decreased resource footprint at a fixed EBNA1 expression, and *(iii)* increased EBNA1 expression and decreased resource footprint.

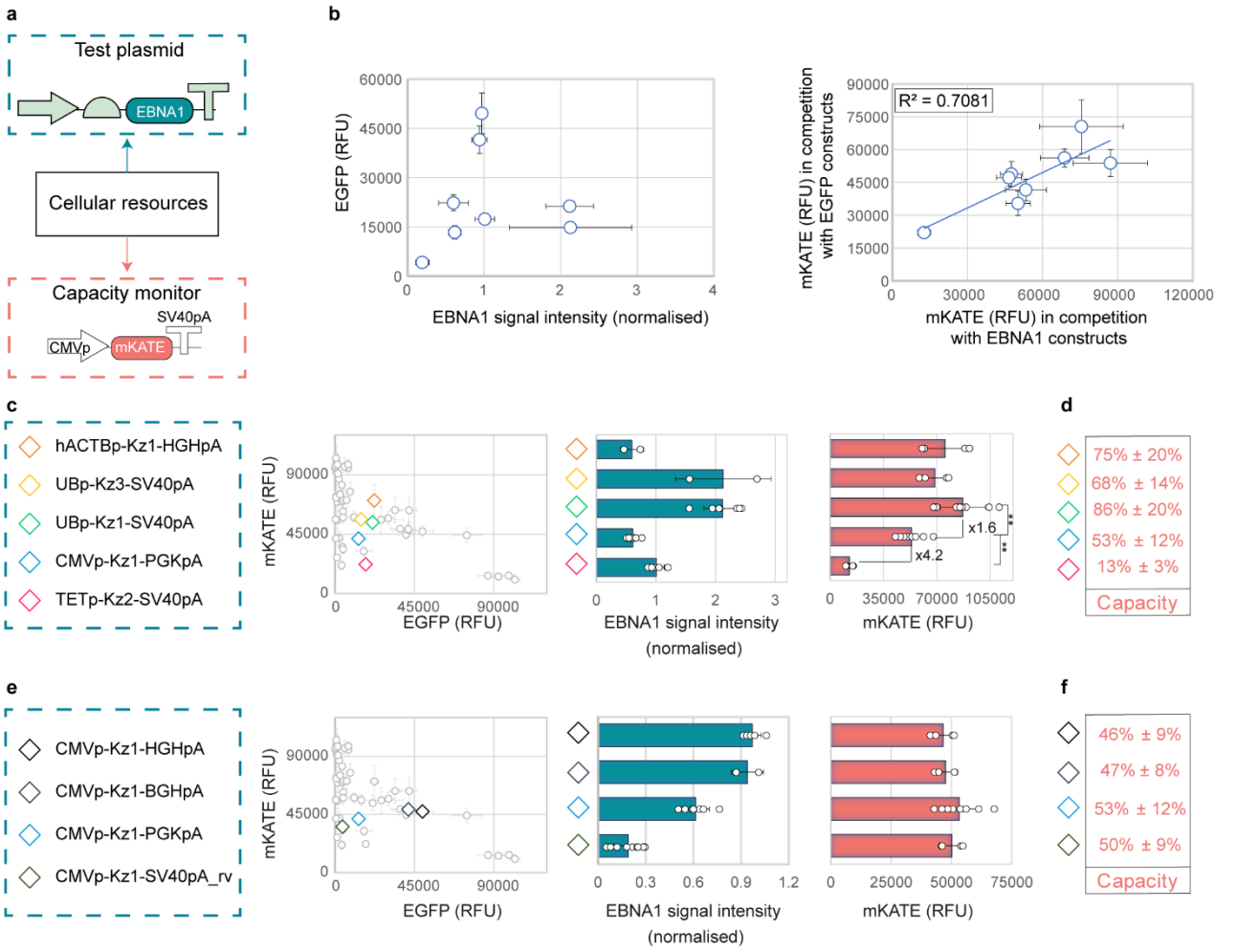

Figure 2.14 Resource-aware design of EBNA1-expressing genetic constructs in HEK293T. (a) We cloned eight EBNA1-expressing designs based on our previous characterisation. Each of these designs was co-transfected with the capacity

monitor in HEK293T. **(b)** Absence of linear relation between EGFP expression (RFU) and EBNA1 expression (normalised signal intensity) for the eight selected designs HEK293T (left panel, Pearson coefficient 0.08). Positive linear relation between mKATE expression in competition with EGFP (RFU) and mKATE expression in competition with EBNA1 (RFU) for the selected designs in HEK293T (right panel, Pearson coefficient 0.84). **(c)** We selected five designs with similar gene expression outputs and different resource footprints (diamonds, first two panels on the left). EBNA1 (middle panel) and capacity monitor (mKATE, right panel) expression levels for the five selected designs. **(d)** Remaining capacity for the five EBNA1-expressing designs in panel c. **(e)** We selected four designs with different gene expression outputs and similar resource footprints (diamonds, first two panels on the left). EBNA1 (middle panel) and capacity monitor (mKATE, right panel) expression levels for the four selected designs. (**f)** Remaining capacity for the four EBNA1-expressing designs in panel e. EGFP and capacity monitor expression are reported as mean RFU ± std. EBNA1 expression is reported as normalised signal intensity and was calculated as described in Materials and methods (Chapter 7). Number of independent experiments ≥ 2 (depending on the condition); two technical repeats per experiment. Data analysis is described in Materials and methods (Chapter 7).

2.2.5. Resource-aware design of multi-gene constructs

Having successfully generated genetic constructs with reduced resource footprint by acting on three design parameters – promoters, Kozaks, and polyA signals -, we decided to add other commonly-used biological parts to our characterisation. As mentioned, co-expression of multiple genes is often needed in mammalian cell engineering applications. While the genes of interest can be arranged in different transcriptional units, one-transcriptional-unit configurations are also adopted. For instance, using internal ribosome entry sites (IRESes) or 2A viral peptides enables the co-expression of two or more genes from one promoter[168]. Since from previous experiments we found that promoters are key design features impacting the resource footprint of genetic constructs, we reasoned that reducing the number of transcriptional units – that is, reducing the number of promoters - by using IRESes or 2A peptides would still enable multi-gene expression but at a lower resource cost.

We assembled a small library of constructs for the co-expression of EGFP and mTagBFP2 (referred to as *BFP* from now on) either from two separate transcriptional units or from one transcriptional unit using 2A peptides or IRES sequences. In particular, we tested two 2A peptides - P2A (porcine teschovirus-1 2A) and T2A (thosea asigna virus 2A)[169] – and one IRES2 sequence (from encephalomyocarditis virus)[170]. We also included an empty plasmid and a single-gene cassette for EGFP expression (Figure 2.15a). We observed a step-wise decrease of the capacity monitor expression upon competition with an increasing number of transcriptional units, with 48% - 26% and 38% - 16% of remaining capacity for the 1TU and 2TU constructs in HEK293T and CHO-K1, respectively (Figure 2.15b-d). This is expected, as a higher number of transcriptional units will inevitably result in a higher resource footprint. When EGFP and BFP are arranged in a bicistronic configuration using IRES and 2A peptides, capacity is rescued to levels higher than the 2TU constructs. In some cases, the capacity of bicistronic arrangements is comparable to, or higher than, the

expression of a single EGFP-expressing TU (i.e. ≥48% in HEK293T and ≥38% in CHO-K1), despite the latter producing only one fluorescent protein (Figure 2.15b-d). This reinforces our findings, as reducing the number of promoters in the circuit frees up transcriptional resources and dramatically reduces the resource footprint of bicistronic multi-gene configurations. Although IRES2 has a lower resource footprint in both cell lines than 2A peptides, this comes at the cost of reduced output levels of the coding sequence in second position (Figure 2.15b-c). This is in line with previous work demonstrating that IRES-dependent translation of the gene in second position is lower than CAP-dependent translation of the first gene[171]. Therefore, when designing bicistronic configurations (IRES vs 2A peptides), the trade-offs between resource footprint and expression levels need to be considered.

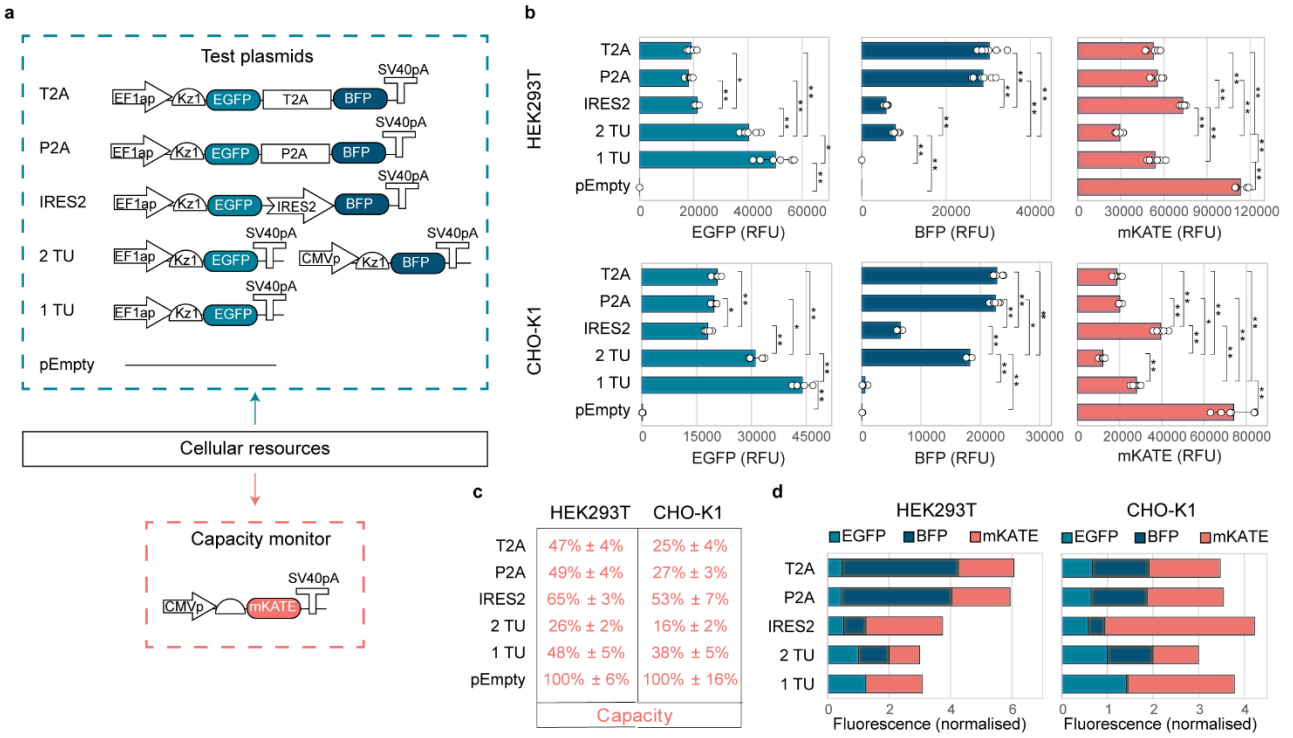

Figure 2.15 Resource footprint of multi-gene constructs. (a) We cloned a library of test constructs for the co-expression of EGFP and BFP either from two transcriptional units (2TU) or from one transcriptional unit (IRES2, P2A, T2A). We included in the library a pEmpty as control of zero resource footprint and a one-transcriptional-unit EGFP-expressing construct. **(b)** Test plasmid (EGFP, BFP) and capacity monitor (mKATE) expression levels in HEK293T (top panels) and CHO-K1 (bottom panels). **(c)** Remaining capacity for tested designs in HEK293T and CHO-K1. **(d)** Test plasmid (EGFP, BFP) and capacity monitor (mKATE) expression levels normalised to the RFU value of the corresponding reporter in the 2TU configuration in HEK293T (left panel) and CHO-K1 (right panel). Test plasmid and capacity monitor expression are reported as mean RFU ± std unless otherwise specified. Number of independent experiments = 3; two technical repeats per experiment. Two-sided Mann-Whitney test P value: ****<0.0001, ***<0.0005, **<0.005, *<0.05. Data analysis is described in Materials and methods (Chapter 7).

We also observed that changing gene syntax on the test plasmid resulted in different resource footprints (Figure 2.16a-d). We arranged two transcriptional units (EF1ap-Kz1-EGFP-SV40pA and CMVp-Kz1-BFP-SV40pA) in tandem on the same plasmid (2 TU - linked) or on two separate plasmids (2 TU – unlinked). The linked arrangement had a remaining capacity comparable to the expression of a one-transcriptional-unit EGFP-expressing plasmid (41% and 44% in HEK293T and CHO-K1, respectively), as opposed to the unlinked arrangement, which came at a higher resource footprint and lower capacity (26% and 16% in HEK293T and CHO-K1, respectively) (Figure 2.16b-c). However, in the linked arrangement, the expression of one of the two genes was suppressed compared to the unliked arrangement in both cell lines. Transcriptional interference between adjacent transcriptional units is a well-reported phenomenon[172,173] and might potentially explain why our two-gene configurations have different resource footprints.

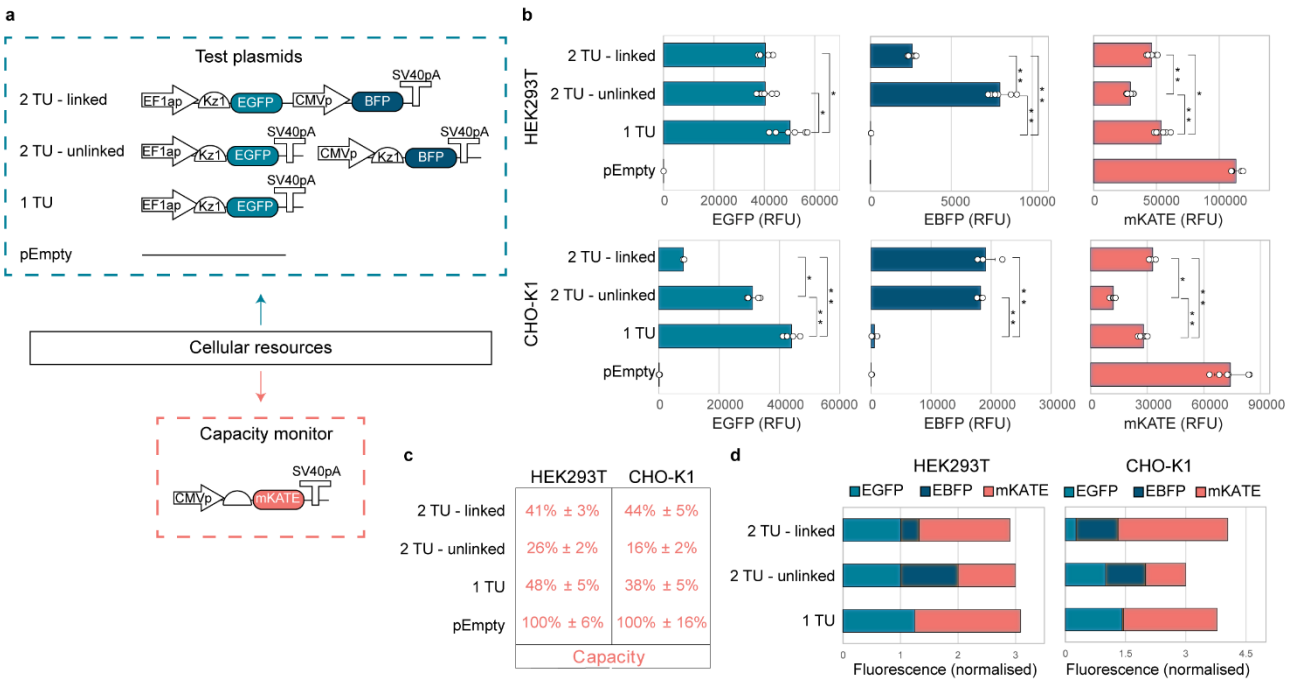

Figure 2.16 Resource footprint of multi-gene constructs with different syntax. (a) We cloned two transcriptional units (EF1ap-Kz1-EGFP-SV40pA and CMVp-Kz1-BFP-SV40pA) in tandem on the same plasmid (2TU – linked) or on two separate plasmids (2TU – unlinked). We included in the library a pEmpty as control of zero resource footprint and a one-transcriptional-unit EGFP-expressing construct. **(b)** Test plasmid (EGFP, BFP) and capacity monitor (mKATE) expression levels in HEK293T (top panels) and CHO-K1 (bottom panels). **(c)** Remaining capacity for tested designs in HEK293T and CHO-K1. **(d)** Test plasmid (EGFP, BFP) and capacity monitor (mKATE) expression levels normalised to the RFU value of the corresponding reporter in the "2TU – unlinked" configuration in HEK293T (left panel) and CHO-K1 (right panel). Test plasmid and capacity monitor expression are reported as mean RFU ± std unless otherwise specified. Number of independent experiments = 3; two technical repeats per experiment. Two-sided Mann-Whitney test P value: ****<0.0001, ***<0.0005, **<0.005, *<0.05. Data analysis is described in Materials and methods (Chapter 7).

Interestingly, while in HEK293T the expression of the gene in second position (BFP) was affected, in CHO-K1, the situation was reversed, with the gene in first position (EGFP) facing output suppression compared to the unlinked configuration (Figure 2.16b,d). This is interesting as it proves that transcriptional interference is bidirectional and can potentially open to spatial considerations of resource competition. In the linked arrangement, the two transcriptional units are in spatial vicinity and might face higher competition than when placed in an unlinked arrangement. The position of the gene facing output suppression across the two cell lines might be explained by the relative efficiency of the promoters of the two transcriptional units (CMVp and EF1ap) in hijacking cellular resources, which might reasonably differ across cell lines. So far this is just a speculation, but if confirmed, it would explain the observed behaviour. Nevertheless, we provide the first report on the effects of gene syntax on resource footprint, showing how different configurations can impact the expression of competing genes.

2.2.5.1. Optimisation of a multi-gene switch

To demonstrate the applicability of our findings, we optimised the design of the TET-ON system towards decreased resource footprint. The TET-ON system is a gene switch widely used in mammalian cell engineering to enable inducible gene expression. As we showed before (Figure 2.2), the TET-ON system is very demanding on cellular resources, dramatically suppressing the expression of competing cassettes. While the extent of the problem is underreported within the community, the TET-ON system is rarely used in isolation, often being part of more complicated pieces of genetic circuitry. In these scenarios, resource competition causes coupling effects that can alter the behaviour of genetic systems.

The TET-ON system is often used in a two-transcriptional-unit configuration. One transcriptional unit constitutively expresses the rtTA transactivator, which, in the presence of dox, binds to the TETp cognate promoter in the other transcriptional unit, activating the expression of the TETp-controlled gene. Alternative configurations where the rtTA and the controlled gene are arranged in a bicistronic configuration and expressed from the same TETp have been developed[174]. However, the resource footprint of similar arrangements has not been characterised. We reasoned that reducing the number of transcriptional units from two to one would decrease the resource footprint of the TET-ON system. We cloned six bicistronic configurations where transcription of both the rtTA and the controlled gene (an EGFP in our case) is initiated from a single TETp through an IRES2 or 2A peptide and terminated by different polyA signals selected to yield different test plasmid outputs based on our previous characterisation (Figure 2.17a). In accordance with our hypothesis, we observed increased capacity monitor expression levels in both the uninduced (dox 0ng/uL) and induced (dox 1ng/uL) conditions when using a one-transcriptional-unit configuration compared to the two-transcriptional-unit arrangement (Figure 2.17b). Although the resource-efficient bicistronic designs showed lower TET-ON-based expression under induced conditions – that is, lower EGFP expression - an improvement in fold change, defined as the ratio of EGFP output in the

presence and absence of dox, was observed in HEK293T cells (Figure 2.17b). Additionally, we noticed that designs incorporating the more efficient RBpA and HGHpA exhibited higher EGFP expression and fold changes than corresponding designs carrying the low-performing PGKpA, once more confirming the applicability of our previous findings to other contexts and applications (Figure 2.17b).

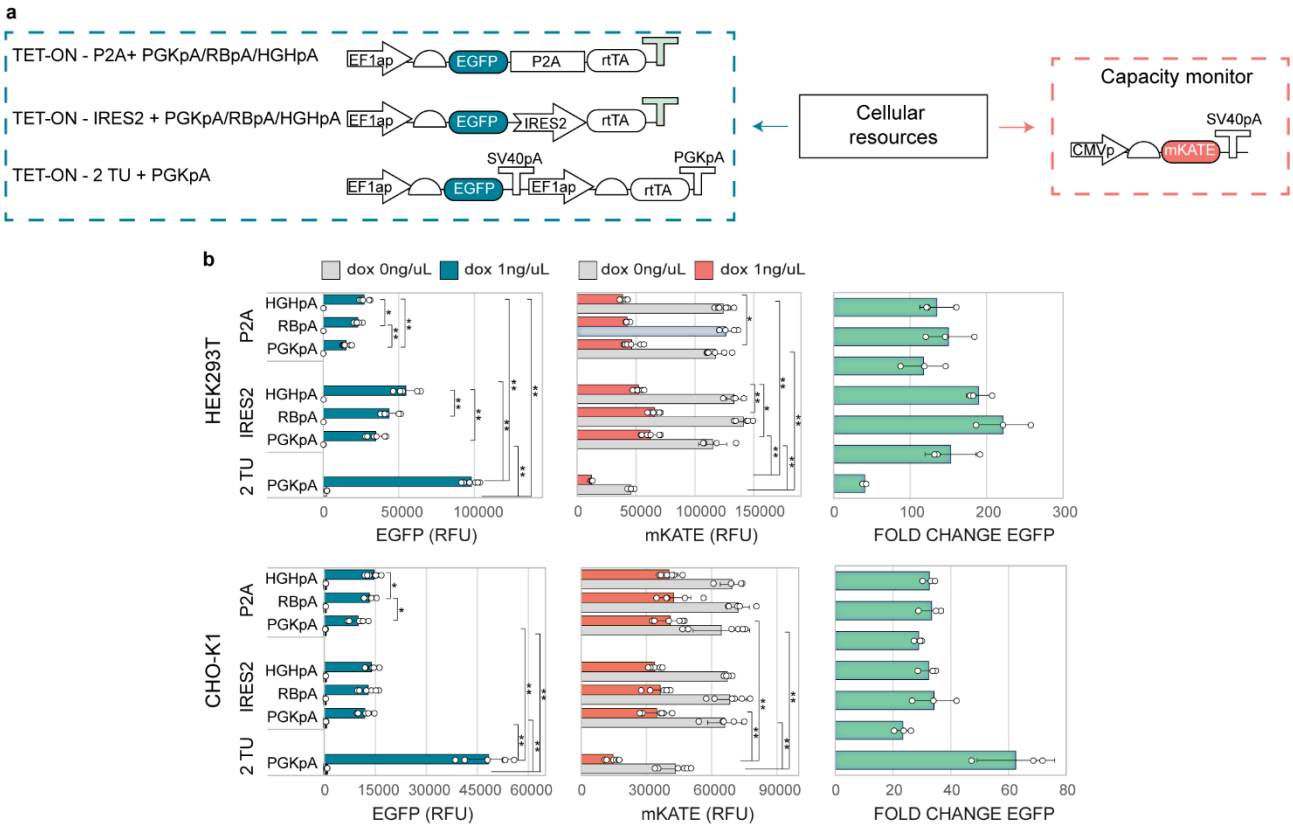

Figure 2.17 Resource-aware design of the TET-ON gene switch. (a) We cloned six bicistronic configurations for the TET-ON system. In these designs, transcription of the rtTA and EGFP is initiated from the same TETp – using an IRES2 or a P2A peptide - and terminated by one of three polyA signals (PGKpA, RBpA, HGHpA). The resource footprint of the bicistronic configurations was evaluated and compared to the two-transcriptional-unit TET-ON configuration. **(b)** TET-ON (EGFP – left panels) and capacity monitor (mKATE – middle panels) expression levels at uninduced (dox 0ng/uL) and induced (dox 1ng/uL) conditions in HEK293T (top panels) and CHO-K1 (bottom panels). Fold change was calculated as the ratio of EGFP output in the presence and absence of dox induction (right panels). TET-ON and capacity monitor expression levels are reported as mean RFU ± std. Number of independent experiments = 3; two technical repeats per experiment. Two-sided Mann-Whitney test P value: ****<0.0001, ***<0.0005, **<0.005, *<0.05. Data analysis is described in Materials and methods (Chapter 7).

A possible explanation of the lower expression of the bicistronic TET-ON designs is that a higher inducer concentration is needed to fully induce the bicistronic configurations. Another interpretation is that the bicistronic designs do not express enough transactivator; as for these, the rtTA gene is part of a positive feedback loop that relies on the leaky expression of the TETp to initiate. Arguably, this might lead to decreased rtTA levels when compared to the two-transcriptional-unit configuration, where the expression

of the rtTA gene is controlled by a strong constitutive promoter. We tested these hypotheses on two shortlisted bicistronic designs (IRES2+HGHpA and P2A+HGHpA) (Figure 2.18a). First, we showed that a 10-fold increase in the inducer concentration did not improve the expression of the bicistronic configurations in HEK293T and CHO-K1 (Figure 2.18b).

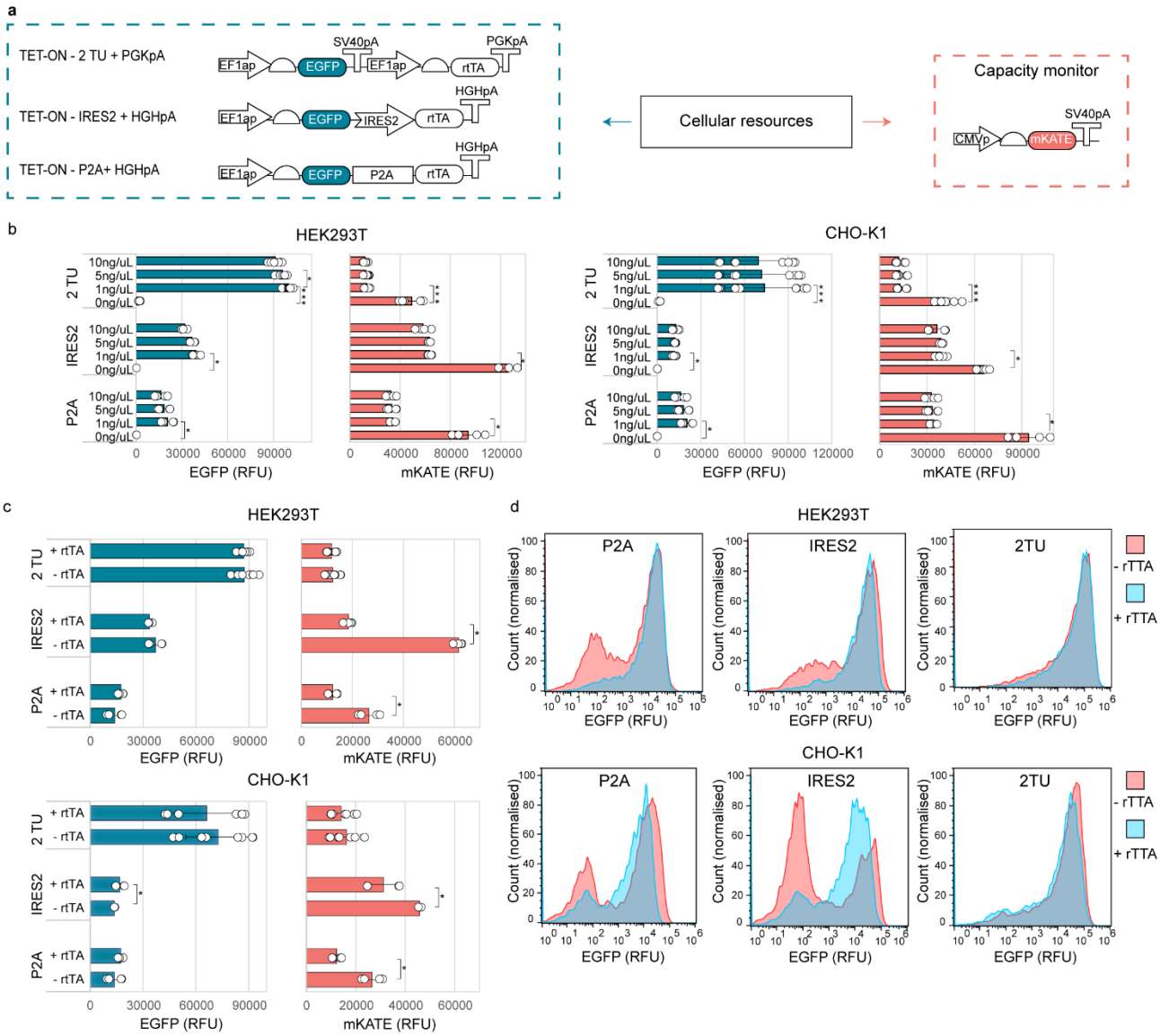

Figure 2.18 Attempts to improve the expression of bicistronic TET-ON configurations. (a) We attempted to improve the expression of two shortlisted bicistronic designs (IRES2+HGHpA and P2A+HGHpA). As control of optimal expression, we included the two-transcriptional-unit TET-ON configuration. **(b)** TET-ON (EGFP) and capacity monitor (mKATE) expression levels at increasing dox concentration (0ng/uL, 1ng/uL, 5ng/uL, 10ng/uL) for the selected TET-ON designs in HEK293T (left panels) and CHO-K1 (right panels). **(c)** TET-ON (EGFP) and capacity monitor (mKATE) expression levels in the presence or absence of an rtTA-expressing plasmid for the three selected designs in HEK293T (top panels) and CHO-K1 (bottom panels). For the "-rtTA" condition, a pEmpty was transfected instead of the rtTA plasmid to maintain the same molar amount of total transfected DNA. **(d)** EGFP fluorescence histograms in the presence or absence of an rtTA-expressing plasmid for the three selected designs in HEK293T (top panels) and CHO-K1

(bottom panels). TET-ON and capacity monitor expression levels are reported as mean RFU ± std. Number of independent experiments ≥ 2 (depending on the condition); two technical repeats per experiment. Two-sided Mann-Whitney test P value: ****<0.0001, ***<0.0005, **<0.005, *<0.05. Data analysis is described in Materials and methods (Chapter 7).

To test the alternative hypothesis, we investigated the effect of additional rtTA concentration on the behaviour of our designs. This was done by repeating the competition experiments in the presence or absence of an additional rtTA-expressing plasmid. Contrary to our hypothesis, additional rtTA did not improve the expression of the bicistronic circuits, increasing the resource footprint instead (Figure 2.18c). This is expected as the extra rtTA cassette will consume additional cellular resources, exacerbating the coupling effects on the capacity monitor. However, we noticed an increased percentage of EGFP-expressing cells in the presence of the rtTA-expressing plasmid for the bicistronic configurations in HEK293T and CHO-K1 (Figure 2.18d). This indicates that additional rtTA is required to improve the activation dynamics of these configurations, although it does not result in increased overall expression levels.

2.3. Conclusions

In this Chapter, we established a resource-aware framework to screen libraries of transfected genetic designs. This allowed us to quantify the resource footprint of genetic parts essential to assemble functional mammalian gene expression systems. We demonstrated that the resource bottlenecks in the gene expression pathway localise at the transcriptional level and that, contrary to published evidence[71,72], translational resources undergo saturation only upon high concentrations of intracellular transcripts from heterologous genes. We then used our resource-aware framework to infer higher-order design rules, which resulted in the generation of genetic constructs with higher efficiency, displaying optimal gene expression performance at a reduced resource footprint. We leveraged our findings to predictably generate EBNA1-expressing designs with defined resource footprints, showing how our work can support bioproduction and biotherapeutic applications. Finally, we investigated the resource footprint of multi-gene configurations, individuating resource-efficient arrangements and redesigning a TET-ON system towards reduced resource cost.

Chapter 3 - Engineering robustness to resource loading

3.1. Introduction

In Chapter 2, we used a resource-aware framework to streamline the generation of genetic constructs with reduced resource footprints. We demonstrated that tuning the design of genetic constructs by acting on key biological parts directly affects their resource footprint. A question that remains to be answered by our previous characterisation is how different genetic constructs respond to resource competition and if it might be possible to design genetic constructs that show robustness to resource limitations. Jones *et al.* already showed that the expression outputs of genetic constructs with varying promoter composition were differentially affected by the competition with a library of trans-activator domains[48]. The only available strategy to date to buffer the output of genetic constructs to fluctuations in the resource pool is through the engineering of adaptive circuits[48,72,132]. These often entail complex genetic architectures, and more straightforward approaches to engineer robustness to resource availability could prove useful to the community.

In this Chapter, we discuss our efforts to design genetic constructs that are more robust to resource shortages. First, we characterised the effects of resource competition on the library of genetic constructs presented in Chapter 2. We identified design parameters influencing the robustness of genetic constructs to resource fluctuations, and we showed that our results were consistent upon change of the competing module. Lastly, we demonstrated how these findings have implications for the adoption of transfection controls.

Overall, this Chapter presents solid evidence that robustness to resource availability can be implemented by modifying the design of genetic constructs and without the need to engineer complex circuit architectures. We also show the usefulness of our approach for foundational biology experiments that use transfection controls.

3.2. Results and discussion

3.2.1. Introducing robustness to resource loading

In Chapter 2, we co-transfected a library of test genetic constructs and a transient capacity monitor to gain information on the impact that each test construct has on the capacity monitor output, which we termed resource footprint. However, resource competition is bidirectional, with the capacity monitor also having a resource footprint and impacting the expression of the library of test plasmids. We reasoned that by changing the observation point, we could use the same type of experiments to investigate how the different genetic designs of the library respond to the resource cost of the capacity monitor expression. To be coherent with the change of observation point, in this Chapter, we refer to the capacity monitor as a *loading module* (or simply, *loader*) imposing a *resource load* on the library of test genetic constructs. Based on this, we define the concept of *robustness to resource load* as the deviation between the expression levels of a test construct in the presence of the loading module and the expression level of the same construct in the absence of resource load. For instance, genetic constructs that maintain similar expression levels despite the presence of competing loading modules are said to be *robust to resource load*, as opposed to genetic constructs that face higher suppression of their expression outputs.

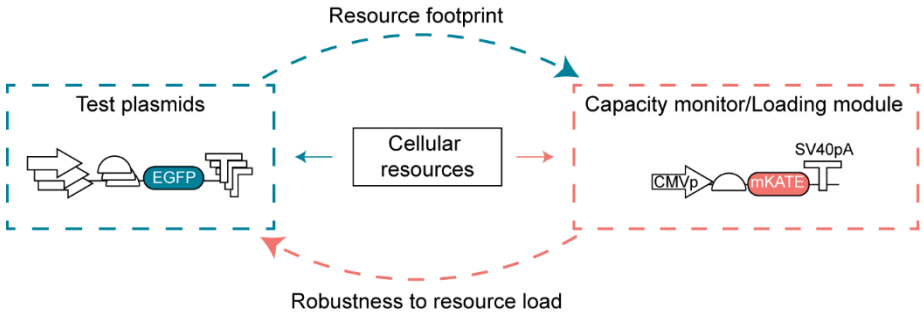

Figure 3.1 Defining robustness to resource load. The effects of resource competition are bidirectional. As each test construct of the library has a resource footprint and impacts the capacity monitor expression, similarly, the capacity monitor places a resource load on the library of test constructs. In this Chapter, the capacity monitor will be referred to as the loading module or loader. The loading module imposes a resource load on the library of test constructs. Each construct displays unique robustness profiles to the resource load placed by the loading module.

3.2.2. Characterising robustness to resource loading

3.2.2.1. A global picture

As a first experiment, we co-transfected the test plasmid library in the presence or absence of a loading module, a CMVp-controlled mKATE gene or an empty plasmid, respectively (Figure 3.2a). Two days after transfection, we quantified the expression of the test plasmids at the flow cytometer. We reasoned that robustness to resource load could be visualised by plotting the expression of the test plasmid library in the presence of the loading module (EGFP resource-loaded) over the expression in the absence of the loading

module (EGFP unloaded) (Figure 3.2a). In this plot, test plasmids robust to resource loading would localise close to the y=x line, indicating similar expression levels for the two conditions. On the other hand, constructs that do not display robustness would sit below the y=x line, meaning that their expression in the presence of the loading module is suppressed compared to their expression in the absence of resource loading (Figure 3.2a). As expected, the experimental data matched our expectations, with the constructs of the library displaying different degrees of robustness (Figure 3.2b).

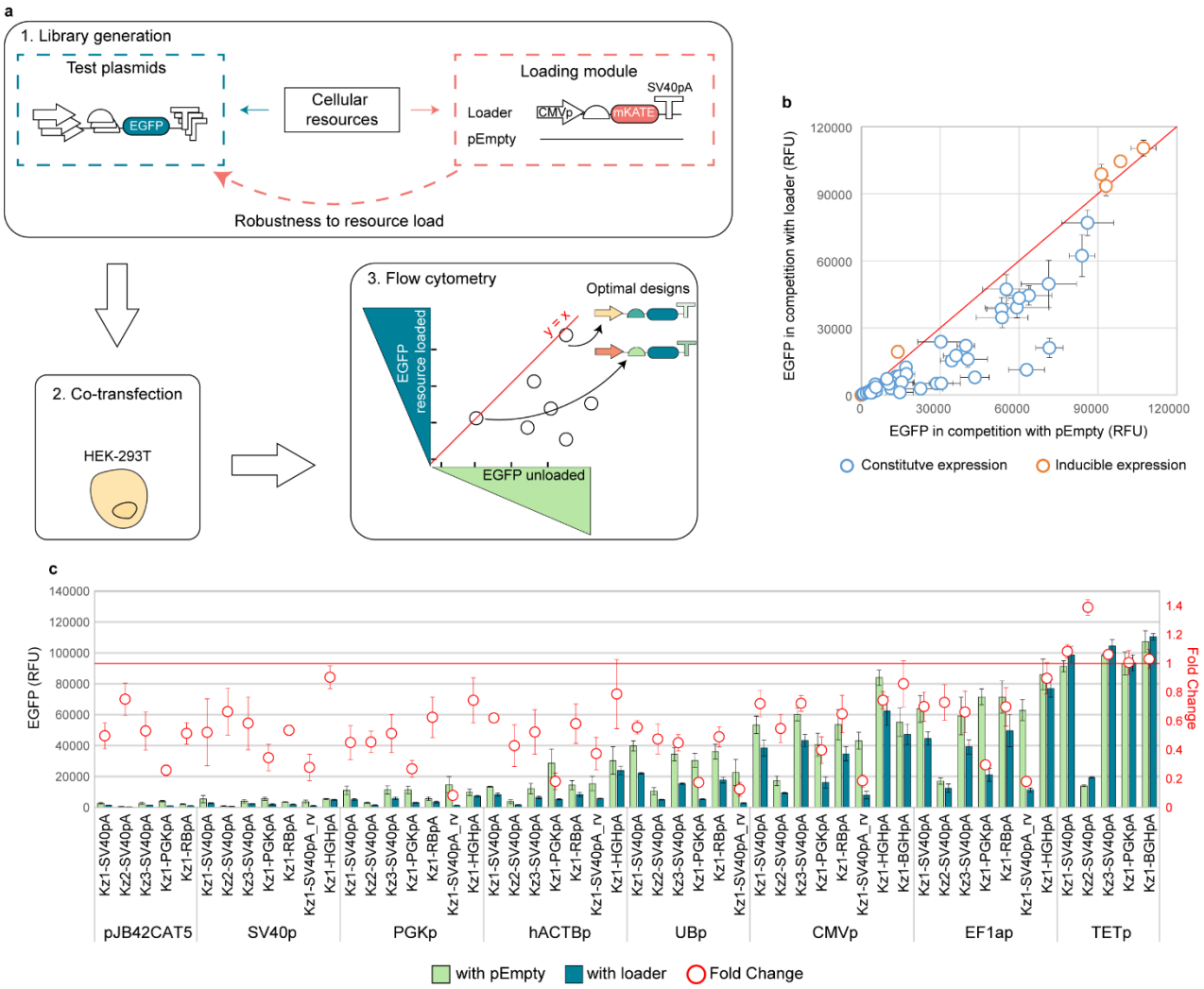

Figure 3.2 Genetic constructs design influences robustness to resource load. (a) Schematics of the experimental workflow. Each test construct of the library was individually co-transfected in the presence or absence of a loading module in HEK293T. In our experiments, the loading module is a CMVp-controlled mKATE-expressing construct, while the unloaded condition is achieved by transfecting an empty plasmid (pEmpty). The robustness to resource load was visualised by plotting the expression levels of the library constructs in the presence of resource load (i.e. EGFP resource-loaded) over their expression levels in the absence of resource load (i.e. EGFP unloaded). The constructs close to the y=x line in this plot display high robustness profiles. **(b)** Test plasmids (EGFP) expression levels for the workflow outlined in panel a. **(c)** Test plasmids (EGFP) expression levels in competition with a pEmpty or a loader. Fold change values for each construct were calculated by dividing the expression levels in the presence of resource load (i.e. with loader) over the expression levels in the absence of resource load (i.e. with pEmpty). Test plasmids

expression levels are reported as mean RFU ± std. Number of independent experiments ≥ 3 (depending on the condition); two technical repeats per experiment. Data analysis is described in Materials and methods (Chapter 7).

Interestingly, all five constructs for inducible expression displayed perfect robust behaviour, regardless of their expression output (Figure 3.2b). This suggests that they are very efficient in hijacking cellular resources and that their expression is favoured in a scenario of resource competition, consistent with previous observations of the high resource footprint of these inducible designs. We also noticed that robustness to resource load does not depend on the expression levels of the library constructs, as we could individuate low-expressing high-robust and high-expressing low-robust designs. To add a quantitative description, we calculated the robustness to load for each construct as the fold change of expression between the loaded and unloaded condition (Figure 3.2c).

3.2.2.2. Contribution of biological parts to resource loading

Proved that the genetic constructs of the library had different robustness to resource load, we investigated the role played by promoters, Kozaks, and polyA signals in shaping robustness profiles in mammalian cells.

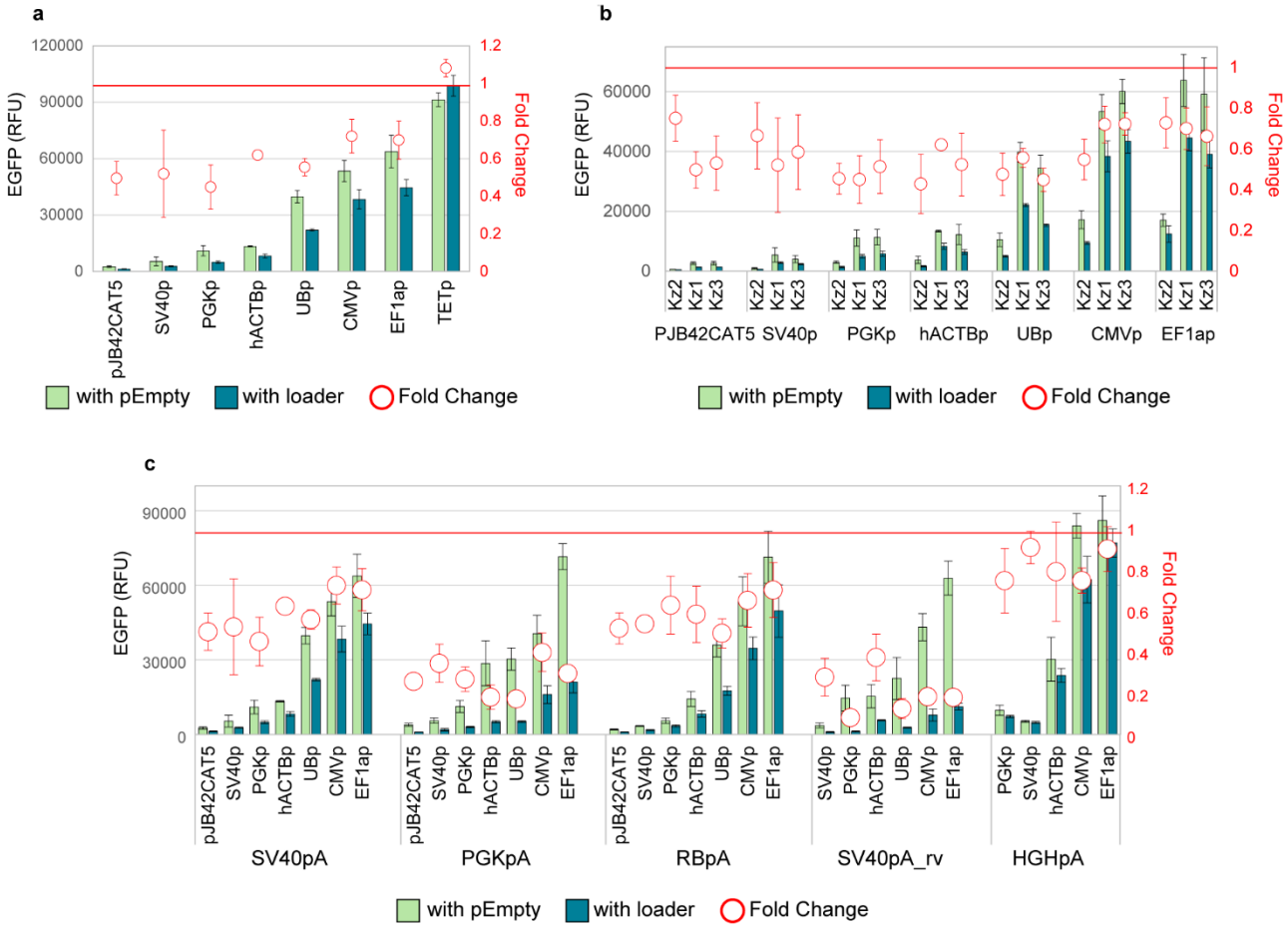

Figure 3.3 Contribution of three biological parts to resource load. (a) Test plasmid (EGFP) expression levels and fold changes for eight test constructs with different promoters. **(b)** Test plasmid (EGFP) expression levels and fold changes

for a library of test constructs bearing different promoter-Kozak configurations. **(c)** Test plasmid (EGFP) expression levels and fold changes for a library of test constructs bearing different promoter-polyA configurations. Test constructs expression levels are reported as mean RFU ± std. Fold change values for each construct were calculated by dividing the expression of the construct in the presence of resource loading (i.e. with loader) over their expression in the absence of resource load (i.e. with pEmpty). Number of independent experiments ≥ 3 (depending on the condition); two technical repeats per experiment. Data analysis is described in Materials and methods (Chapter 7).

While changes in the promoter or Kozak sequence led to different fold changes (i.e. different degrees of robustness), we could not identify any meaningful trend (Figure 3.3a-b). However, we noticed that polyA signals had a decisive effect on the robustness of the tested designs, irrespective of the promoter they were paired with (Figure 3.3c). For instance, while PGKpA and SV40pA_rv consistently resulted in poor robustness profiles of genetic constructs, SV40pA and RBpA increased fold changes significantly. Overall, HGHpA proved to be the best polyA signal in our panel, resulting in highly robust constructs in all the tested promoter configurations (Figure 3.3c). While explaining these observations is difficult without further experiments, these data further highlight the complex role of polyA signals in resource competition, which we also outlined in Chapter 2. Further study will be needed to understand this complexity towards more predictable engineering of genetic constructs for mammalian expression.

Based on these results, we shortlisted few test constructs bearing different promoter-polyA combinations for further analysis. We selected constructs bearing one of three strong promoters (CMVp, EF1ap, and TETp) in all the polyA configurations present in our library. For these constructs, we analysed the response at various degrees of resource load either by titration of the plasmid ratio of the loading module (Figure 3.4) or by testing four loading modules imposing a different resource load (Figure 3.5).

First, we performed a titration experiment where we varied the ratio between an empty plasmid and the loader (going from pEmpty:loader of 3:0 to 0:3) while keeping the ratio of the transfected test plasmid constant. This experiment allowed us to visualise the response of the selected test plasmids to increasing load from the same loading module. Overall, the robustness profiles observed in our initial experiments were maintained, with TETp inducible constructs displaying the highest robustness and clear polyA-dependent effects (Figure 3.4). Interestingly, we observed that while for some constructs the output was completely suppressed at the lowest concentration of the loading module (1:2:1), for others, we observed a step-wise decrease in their expression levels with increasing concentration of loader. For instance, constructs with SV40pA_rv and PGKpA variants faced dramatic output suppression at the lowest concentration of the loading module, while SV40pA and RBpA variants displayed a more gradual decrease (Figure 3.4).

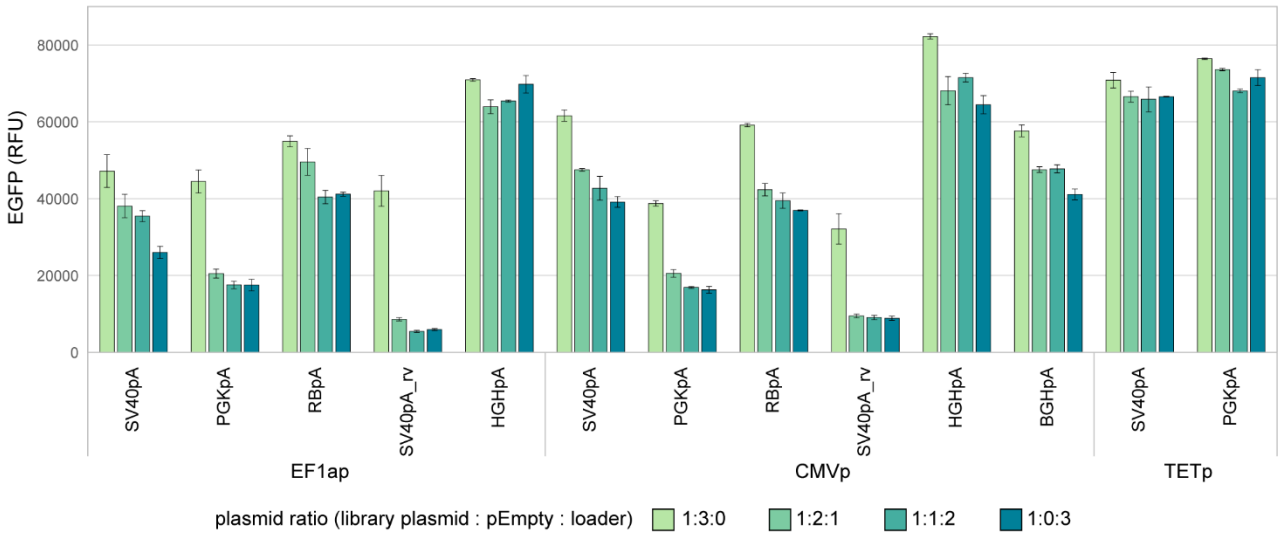

Figure 3.4 Robustness to increasing concentration of the loading module. We shortlisted few test constructs from the library for further analysis. We selected constructs bearing strong promoters (CMVp, EF1ap, and TETp) in all polyA configurations. Test constructs (EGFP) expression levels at different concentrations of the loader. Test constructs expression levels are reported as mean RFU ± std. Number of independent experiments = 1; two technical repeats per experiment. Data analysis is described in Materials and methods (Chapter 7).

In addition to this, we also tested the selected test plasmids in competition with four loading modules imposing a different resource load (Figure 3.5). We used loading modules that we knew from previous characterisation in Chapter 2 to impose very different resource loads (in order of increasing load: SV40p-mKATE, UBp-mKATE, CMVp-mKATE, and EF1ap-mKATE). Again, the robustness profiles were in accordance with previous results (Figure 3.5).

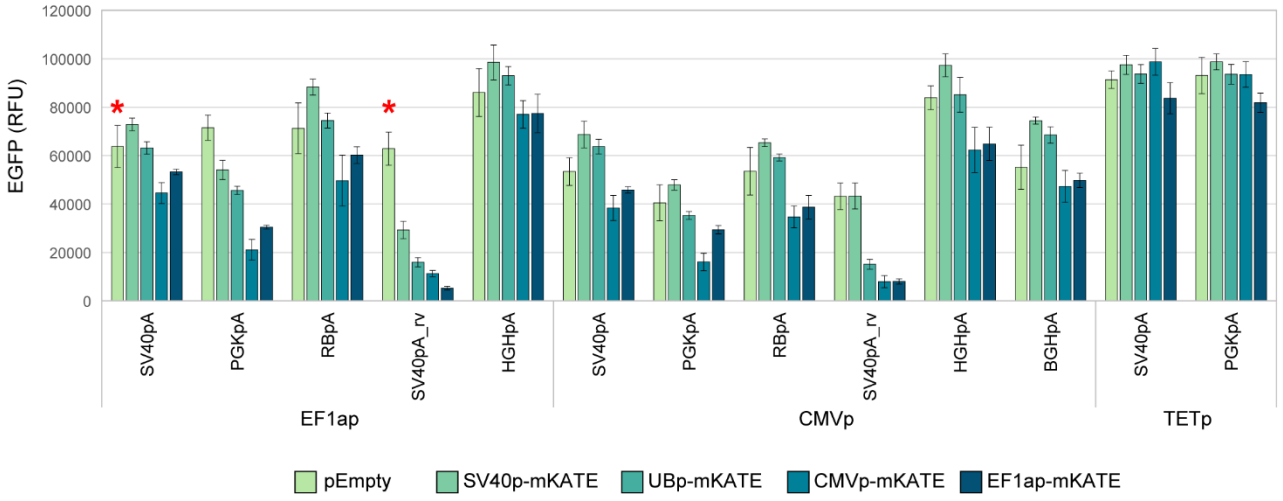

Figure 3.5 Robustness to increasing resource load from different loading modules. We shortlisted few test constructs from the library for further analysis. We selected constructs bearing strong promoters (CMVp, EF1ap, and TETp) in all polyA configurations. Test constructs (EGFP) expression levels at different degrees of resource load imposed by four different loading modules. Red asterisks indicate two constructs with similar expression output without resource loading but different robustness behaviour. Test constructs expression levels are reported as mean RFU ± std.

Number of independent experiments ≥ 3 (depending on the condition); two technical repeats per experiment. Data analysis is described in Materials and methods (Chapter 7).

We also noticed that two of the shortlisted constructs displayed similar expression output in the absence of resource load but very different robustness behaviours (Figure 3.5, red asterisks). We selected these two designs for further considerations outlined in the next section.

3.2.3. Individuation of better transfection controls

Our results could be particularly relevant for designing transfection controls for use in mammalian cell biology. Transfection controls allow to control for variability in plasmid copy number, typical of transfection experiments, by normalisation of the expression output of the genetic construct of interest over the expression output of the co-transfected transfection control. When transfection controls are used to control variability across a panel of genetic constructs, their expression must be robust to resource load. Should this not be the case, the expression of the transfection controls would fluctuate based on the degree of resource load imposed by each sample. This would introduce artefacts and bias in experimental results.

In Figure 3.5, we noticed that two test constructs had similar expression output in competition with the empty plasmid but displayed different robustness profiles in the presence of resource load, with EF1ap-Kz1-SV40pA being more robust than EF1ap-Kz1-SV40pA_rv (Figure 3.5 red asterisks, Figure 3.6a). We reasoned that these two genetic constructs could be considered alternative designs of transfection controls that are co-transfected with a library of mKATE-expressing constructs bearing different promoters. Treating these two samples as transfection controls can help us uncover how artefacts originate when robustness to resource load is not included in the experimental design. When comparing the raw expression levels of the promoter library (mKATE) with the expression levels normalised over the transfection control (mKATE normalised), we noticed two major biases (Figure 3.6b-c). First, we observed that using the low-robust transfection control design (EF1ap-SV40pA_rv) led to a dramatic overestimation of normalised expression values of genetic constructs with high resource load. This is evident when comparing the normalised and raw expression levels of the EF1ap-mKATE design. While the raw expression values are comparable to the CMVp-mKATE and UBp-mKATE designs, the normalised values are much higher. This is not observed when using the high-robust alternative design (EF1ap-Kz1-SV40pA), for which the promoter library's raw and normalised expression levels have a similar trend (Figure 3.6b-c). In addition to this, we also observed that bias is introduced when comparing normalised expression values obtained using different transfection controls. For instance, while the raw expression values of CMVp-mKATE and EF1ap-mKATE are higher when in competition with EF1ap-Kz1-SV40pA compared to when in competition with EF1ap-Kz1-SV40pA_rv, the normalised expression values have an opposite trend (Figure 3.6b-c). These results highlight how bias can originate and issue a warning when using transfection controls for normalisation procedures. Ideally,

normalisation should not be performed altogether to eliminate biases. When this is impossible, a robust transfection control design should be used.

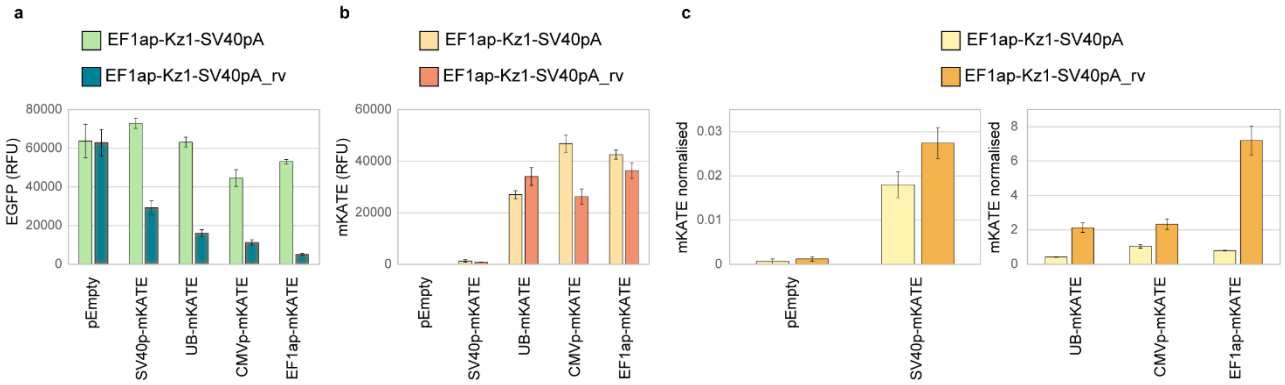

Figure 3.6 Experimental biases originating from the use of transfection controls. (a) Transfection controls (EGFP) expression levels when co-transfected with a library of genetic constructs with different promoters. Expression levels are reported as mean RFU ± std **(b)** Promoter library (mKATE) raw expression levels when co-transfected with two transfection controls. Expression levels are reported as mean RFU ± std **(c)** Promoter library (mKATE) normalised expression levels when co-transfected with two transfection controls. Normalised expression levels have been calculated by dividing the expression levels of each genetic construct (mKATE) over the expression levels of co-transfected control (EGFP). Number of independent experiments ≥ 3 (depending on the condition); two technical repeats per experiment. Data analysis is described in Materials and methods (Chapter 7).

3.3. Conclusions

In this Chapter, we explored strategies to design genetic constructs with increased robustness to resource load. To provide a starting point for this work, we assessed the expression of a library of genetic constructs in the presence of resource load from a constitutively-expressed loading module or in the absence of resource load using a co-transfected empty plasmid. We demonstrated that genetic constructs of different designs have varying robustness and that increased robustness profiles are not dependent on the expression level of genetic constructs. We noticed that constructs for TET-ON inducible expression show perfect robustness profiles and that, contrary to promoters and Kozak sequences, polyA signals dramatically impact the robustness of genetic constructs. Finally, we used our findings to highlight two sources of bias when using transfection controls for normalisation procedures.

Chapter 4 - Resource-depleted mammalian cells: effects on integrated genes and cellular growth

4.1. Introduction

In the previous chapters, we characterised the effect of resource competition in transiently-transfected systems. Understanding whether resource competition is limited to co-transfected systems or affects integrated genes is of paramount importance for the mammalian cell engineering community, given the widespread use of integrated payloads in bioproduction and biotherapy applications. It is also unclear what is the effect of resource competition on the physiology of mammalian cells. While scattered evidence showed some impact on cellular growth[65], a comprehensive experimental investigation is still lacking.

In bacteria, integrated capacity monitors have proved effective in sensing and quantifying the resource load from competing plasmids[61,89]. Ceroni *et al*. developed a capacity monitor to quantify how much gene expression capacity is taken up by genetic constructs. The system is based on a constitutively expressed GFP integrated inside the *E. coli* genome. By tracking the GFP production rate, the authors managed to quantify the demand that genetic constructs place on the pool of gene expression resources. They also observed a positive linear relation between the capacity monitor production rate and the growth rate of the engineered bacteria, meaning that the monitor can be used as a proxy measure of the burden that genetic constructs impose on the bacterial host. Integrated capacity monitors are the best option to assess the effect of heterologous expression on endogenous genes and, ultimately, on the cell's physiology.

This Chapter summarises a consistent body of work to develop fluorescence-based integrated capacity monitors in mammalian cells. First, we developed a landing pad master cell line to streamline the integration of capacity monitor designs. We then tested four design iterations: two constitutive capacity monitor designs and two inducible capacity monitor designs, ultimately individuating an integrated capacity monitor design with optimal features. Importantly, in this iterative process, we elucidated useful insights for the design of integrated biosensors in mammalian cells. Finally, we used the optimal capacity monitor design to test the resource footprint of a library of genetic constructs, proving that the capacity monitor output can effectively recapitulate growth profiles of resource-depleted engineered mammalian cells.

Overall, this Chapter provides evidence of the impact of resource competition on integrated genes. Here, we present the first example of integrated capacity monitor in mammalian cells. We prove that our capacity monitor reports on cellular resource availability and growth profiles, providing a tool for real-time monitoring of the effects of resource competition on the physiology of engineered mammalian cells.

4.1.1. Note to the reader

The data presented in this Chapter include some samples and experiments with only one biological repeat. This is because the experimental work was carried out over many years and required many optimisation experiments, especially during the first years of the PhD, where protocols had to be set up. This resulted in data that, although could not be pooled together and plotted in a single graph, could still give useful insights to inform future capacity monitor iterations. Despite these flaws, we ran three biological repeats for the experiments involving the final integrated capacity monitor design.

4.2. Results and discussion

4.2.1. Generation of a landing pad master cell line

In the early stages of this work package, we realised that testing multiple capacity monitor designs was likely needed before finding the optimal one. Since building mammalian stable cell lines is a long and laborious process, we decided first to generate a landing pad master cell line. Landing pads are genomically integrated platforms that can be used for easy and fast site-specific recombination of payloads of interest[152,175–177]. To enable recombination, complementary recombination sites must be included in the landing pad and the payload design, and the corresponding integrase gene must be expressed to catalyse the recombination reaction. Our landing pad design consists of a tricistronic transcriptional unit downstream of the strong constitutive promoter EF1ap (Figure 4.1). The promoter and the transcriptional unit are separated by the phage BxBI attP recombination site, enabling the recombination of genetic payloads carrying the complementary attB site in the presence of the BxBI integrase (Figure 4.1). The transcriptional unit encodes for the green fluorescence gene mNeonGreen (mNG), the BxBI phage integrase to enable site-specific recombination of payloads, and the puromycin resistance gene to allow easy selection of cell clones bearing at least one integration of the landing pad. The advantages of this arrangement are twofold. First, the transgene in the payload should only expressed after correct recombination inside the landing pad. Second, the cells undergoing correct recombination are marked by the loss of mNG expression, which is easily trackable.

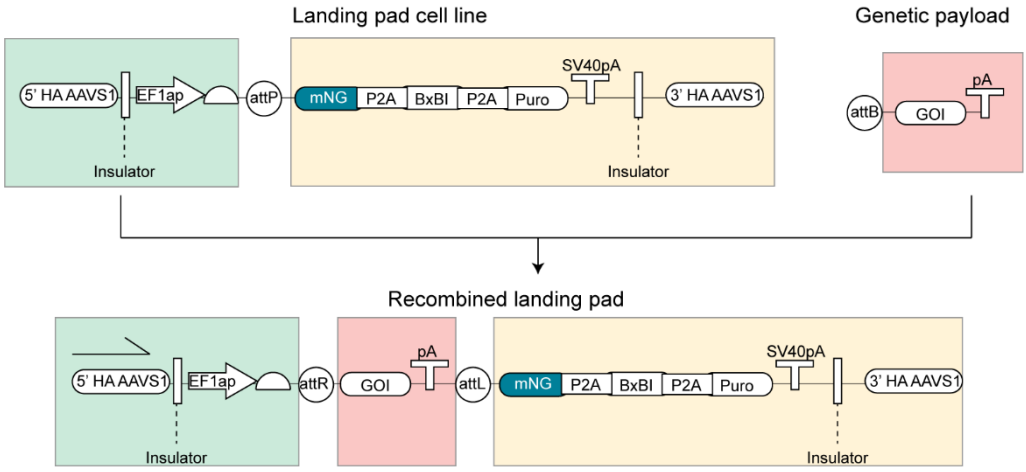

Figure 4.1 Landing pad design. The landing pad consists of a tricistronic transcriptional unit under the control of EF1ap. The transcriptional unit encodes for a fluorescence gene (mNG), the BxBI integrase gene, and the puromycin resistance gene. An attP site was included between the EF1ap promoter and the tricistronic transcriptional unit to enable recombination with a complementary attB site in the genetic payload. After recombination, the genetic payload will be inserted downstream of EF1ap, displacing the original landing pad tricistron, which will stop being expressed. We flanked the landing pad with homology arms for the AAVS1 locus (i.e. 5' HA AAVS1 and 3' HA AAVS1 in the figure).

Since we envisioned using the landing pad as a platform to generate capacity monitors, the site of integration was key to our experimental design. While we needed to select a transcriptionally active site to avoid epigenetic silencing over long-term culturing[55], it was also important to avoid disturbing the expression of essential genes, as this could bias the behaviour of our capacity monitor in later experiments. For these reasons, we selected the AAVS1 locus (chromosome 19) as the integration site and CRISPR-Cas9 as the integration method. We chose AAVS1 because it has been reported as a genomic locus allowing stable expression without known side effects for cellular physiology[54], and we selected CRISPR-Cas9 to have precise control over the integration site. To generate the landing pad cell line, we co-transfected HEK293T cells with two plasmids: the All-in-One plasmid encoding Cas9 and the gRNA; and the plasmid bearing the landing pad sequence flanked by the AAVS1 homology arms (Figure 4.2a). We passaged the transfected cell population for 10 days to allow complete dilution and degradation of the transfected plasmids, as suggested by previously published work[152]. At this point, assuming that the residual mNG fluorescence resulted from correct integration events, we single-cell sorted the fluorescent cellular population into individual wells and expanded the resulting clones. We isolated clones bearing only one copy at the AAVS1 locus to streamline future capacity monitor integrations. Having more than one landing pad copy complicates the selection step of subsequent recombination experiments, as one would need to screen out cells with partial recombination events (i.e. those cells where not all the landing pad copies have successfully recombined with the payload).

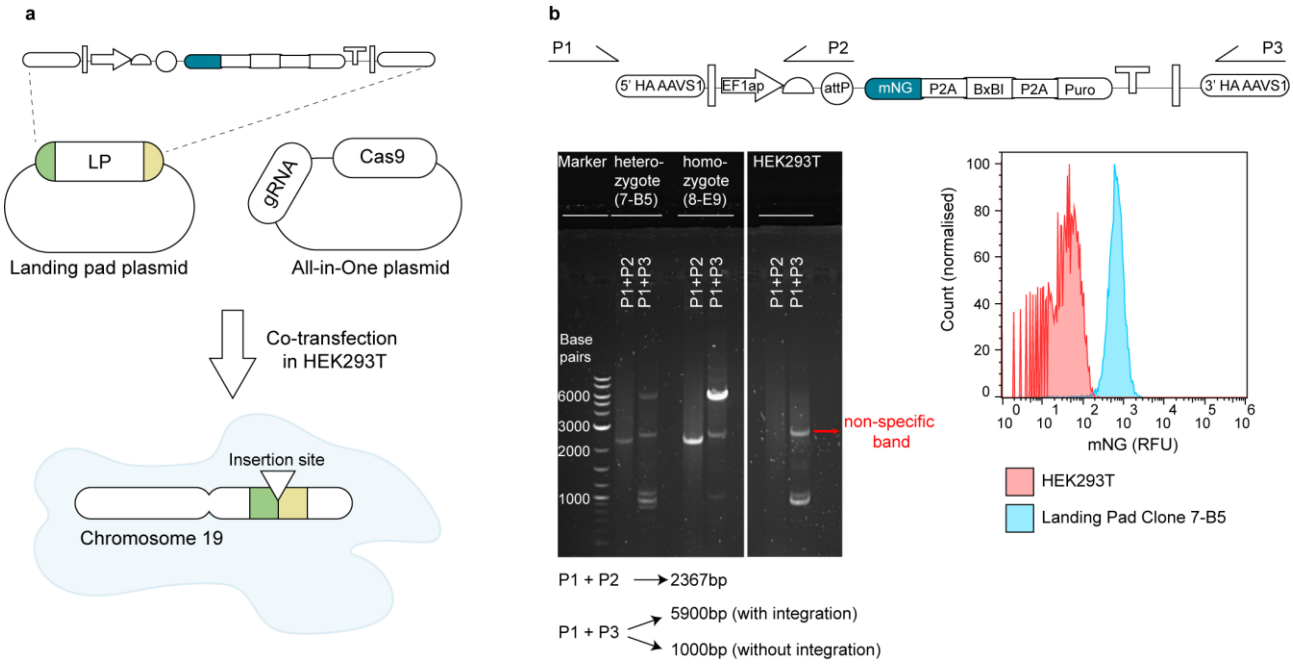

Figure 4.2 CRISPR-Cas9 integration of the landing pad design in HEK293T. (a) The landing pad plasmid was co-transfected with an All-in-One plasmid encoding both the Cas9 and the gRNA necessary for integration into the AAVS1 locus. Upon co-transfection in HEK293T, the Cas9+gRNA introduces a double-strand break at the AAVS1 locus, which is repaired by homologous recombination with the AAVS1 homology arms flanking the landing pad (coloured in the

figure). This results in the insertion of the landing pad at the target genomic location. **(b)** PCR genotyping of a heterozygote clone (7-B5) and a homozygote clone (8-E9), showing correct integration patterns (left panel). Non-specific band present at 2.5Kb following amplification with P1+P3. HEK293T were included as a negative control. Although all the samples were run together, the gel was cropped to eliminate unnecessary samples, which might have confounded the reader. Fluorescence (mNG) histogram of the selected clone 7-B5 and unmodified HEK293T (negative control), showing homogeneous expression of the integrated clone (right panel). Landing pad fluorescence is reported as RFU.

To isolate clones with a single integration, we first assessed the sorted clones using fluorescence microscopy and retained only those that showed homogeneous mNG fluorescence. We extracted the genomic DNA of these clones and performed PCR genotyping to gain information on the copy number. Eventually, we selected clone 7-B5 as our best candidate, as it carried a single-copy landing pad insertion– as evidenced by the PCR-genotyping (Figure 4.2b – left panel) – and homogeneous mNG fluorescence expression (Figure 4.2b – right panel).

4.2.2. Integrated capacity monitor v1 - testing of a constitutively expressed capacity monitor

After isolation and expansion of the landing pad master cell line (clone 7-B5), we tested it as the first version of capacity monitor. We transfected the landing pad cell line with test plasmids of increasing complexity, including a pEmpty, a CMVp-mKATE plasmid, and a CMVp-Cas9-T2A-mKATE plasmid (Figure 4.3a). Two days after transfection, intracellular fluorescence was recorded by flow cytometry (Figure 4.3b). Regardless of the type of test plasmid, no change in the landing pad fluorescence was observed compared to the untransfected condition (Figure 4.3c).

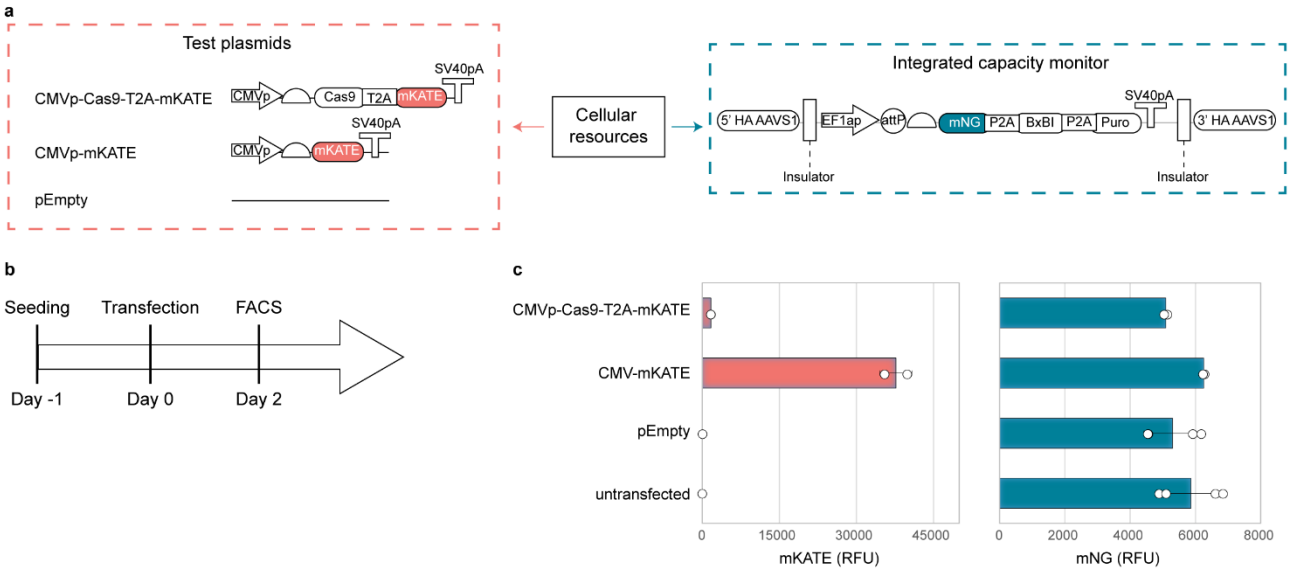

Figure 4.3 Testing of capacity monitor v1. (a) We transfected the landing pad cell line (clone 7-B5) with three test plasmids of diverse composition: a pEmpty as control of zero resource footprint, a plasmid encoding for the mKATE fluorescence gene under the control of the CMVp (CMVp-mKATE), a plasmid encoding for a bicistronic transcriptional

unit under the control of the CMVp (CMVp-Cas9-T2A-mKATE). **(b)** Timeline of the experiment. **(c)** Test plasmid (mKATE) and capacity monitor (mNG) expression levels were reported as mean RFU ± std. Number of independent experiments ≥ 1 (depending on the condition); two technical repeats per experiment. Statistical significance was not calculated as the number of independent experiments is less than two for some samples. Data analysis is described in the Materials and methods section.

Despite the landing pad having a similar design to our transient capacity monitor in Chapter 2, both constitutively expressing a fluorescence gene, the expression of the two at the time when resource competition is initiated is very different. While the transient monitor starts to be expressed in a condition of resource competition and reaches a steady state at the end of the two-day experiment, the integrated landing pad has already reached its steady state expression when resource competition begins and will have to readjust to a new steady state. This introduces a delay because the fluorescent proteins produced by the landing pad must be diluted and degraded before the new steady state can be visualised at the flow cytometer. Therefore, the landing pad design may not be dynamic enough to visualise resource competition during the two-day experiment.

4.2.3. Integrated capacity monitor v2 – generation and testing of a constitutively expressed capacity monitor with increased protein turnover

4.2.3.1. Generation of capacity monitor v2

Based on previous results, increasing the turnover of the readout fluorescent protein might result in improved dynamics of the capacity monitor. Therefore, we cloned a new capacity monitor payload for recombination inside the landing pad. The design is made of an attB site followed by a promoter-less cassette encoding for three mCherry genes fused together – to enhance the brightness of the readout gene – and appended with a C-terminal PEST degradation domain – to increase protein turnover (Figure 4.4a). We also included a blasticidin resistance gene to enable the positive selection of cells with correct recombination events (Figure 4.4a). When transfecting the landing pad master cell line with the capacity monitor payload, we failed to isolate recombinants (data not shown). We think this may be because the BxBI gene in the landing pad sequence was mistakenly not appended with a nuclear localisation signal. This might cause the recombination reaction to be inefficient. To obviate this, we co-transfected the master cell line with the capacity monitor payload and a plasmid constitutively expressing a BxBI gene tagged for nuclear localisation. In this case, an mCherry+/mNG- recombinant population has been identified ten days post-transfection (Figure 4.2b, red box). These results confirmed our previous claim that more than two days are needed before the residual landing pad fluorescence (mNG) is completely lost, suggesting that destabilised fluorescent proteins are crucial to see any burden-induced decrease in the capacity monitor expression. Interestingly, we observed the presence of a hybrid population of mCherry+/mNG+ cells that decreased from 29-30% on day 5 post-transfection to 1-2% on day 10 in the landing pad cell line (Figure

4.4b). This was unexpected, as according to our payload design, the only way for the landing pad cells to express mCherry would be upon correct recombination of the payload, which would also entail loss of mNG fluorescence (i.e. mCherry+/mNG-). This behaviour can be explained assuming that the transfected payload

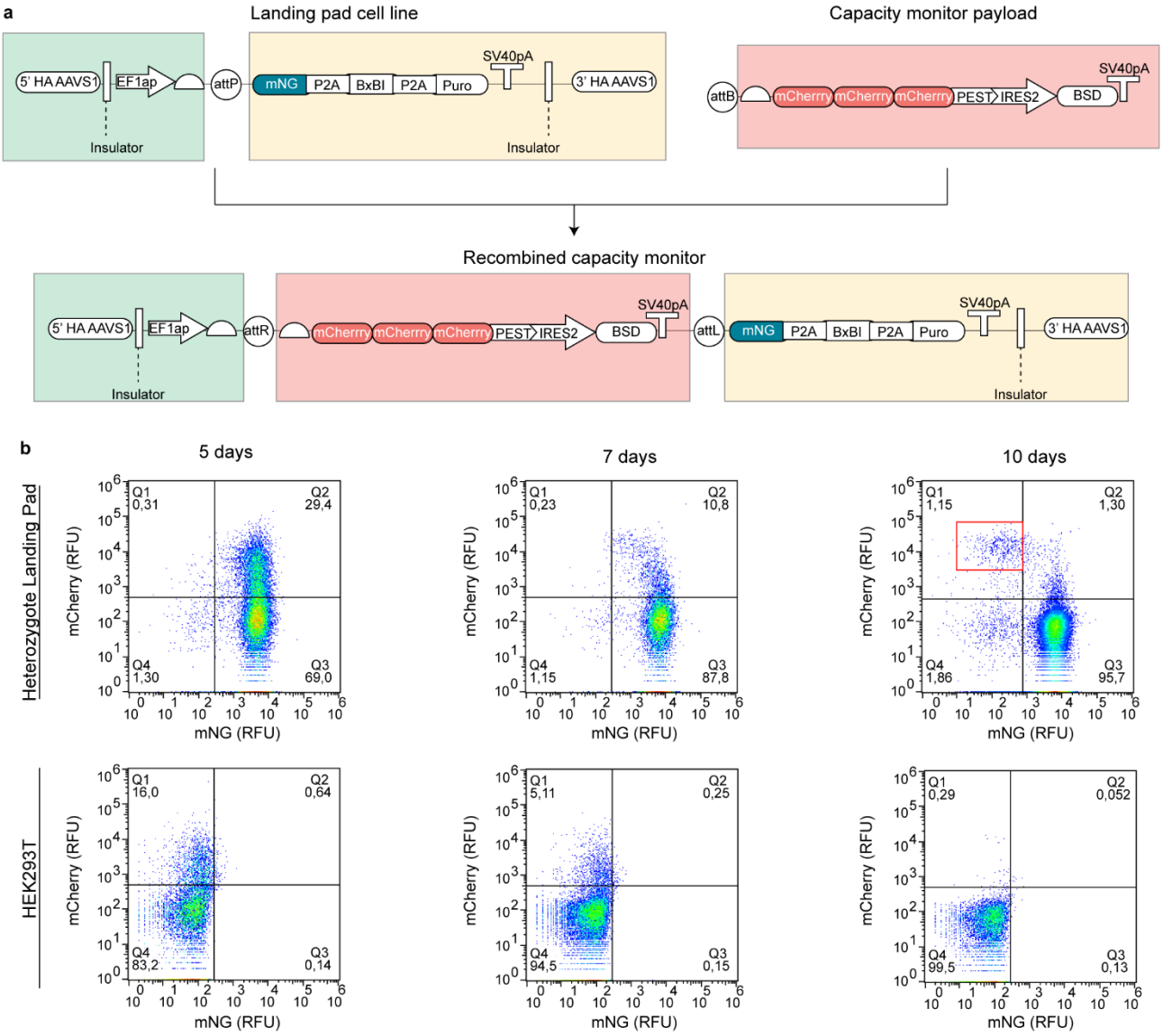

Figure 4.4 Recombination of a capacity monitor payload inside the landing pad. (a) We cloned a capacity monitor payload for recombination inside the landing pad. The capacity monitor carries an attB site; a promoter-less cassette encoding three mCherry fused together and appended with a PEST domain; a blasticidin resistance cassette, whose translation is initiated by an IRES sequence. The capacity monitor payload and an NLS-tagged BxBI-expression plasmid are co-transfected inside the landing pad master cell line. Correct recombinant events can be isolated by tracking the loss in mNG fluorescence **(b)** Landing pad (mNG) and capacity monitor payload (mCherry) fluorescence spectra at day 5, 7, and 10 days after the transfection experiment outlined in panel a. HEK293T have been included as a negative control. Fluorescence is reported as RFU.

produces mCherry prior recombination and gets diluted over the course of 10 days, suggesting that the capacity monitor payload was somehow expressed before landing pad recombination despite it not bearing

an obvious promoter sequence. To gain a better understanding of these results, we tested the recombination of a second genetic payload, simpler in design compared to our capacity monitor (Figure 8.4). In this case, only the correctly recombined population (mKATE+/mNG-) was present and no hybrid cells were detected (i.e. mKATE+/mNG+). Based on these results, we speculate that cryptic promoter sequences present in the capacity monitor payload – and absent in the simpler payload - might initiate unwanted transcriptional events. This was not a problem for us, as to isolate correct recombination events we could either track the loss of landing pad fluorescence (mNG) or wait for the plasmid-produced mCherry fluorescence to be completely diluted (Figure 4.4b). Since the recombination efficiency was approximately 1% at day 10 post-transfection, we sorted the mCherry+/mNG- recombinant cells into a homogenous population (Figure 4.4b, red box). Unexpectedly, the sorted population faced quick progressive silencing of the expression payload, regardless of the original landing pad cell line showing stable expression profiles over long-term culturing (Figure 8.5). To our knowledge, similar behaviour has never been reported in the literature, and we did not anticipate it. Despite this represented a major setback, we used the unstable recombined capacity monitor cell line to gain insights to optimise subsequent capacity monitor design iterations.

4.2.3.2. Testing of capacity monitor v2

To limit the effect of post-recombination instability, we sorted the high mCherry-expressing population few days before performing the experiments. Although this reduced the spread of the fluorescence distribution, the fluorescent population was still not homogeneous (data not shown), and the data presented below are affected by this issue. As a first experiment, we transfected the capacity monitor with either a pEmpty or a plasmid encoding a TET-ON-controlled mNG (Figure 4.5a-b). Upon transfection, we noticed high variability in the capacity monitor expression levels amongst experiments performed on different days (Figure 4.5c – top panels), which was dramatically reduced following normalisation over the untransfected control (Figure 4.5c – bottom panels).

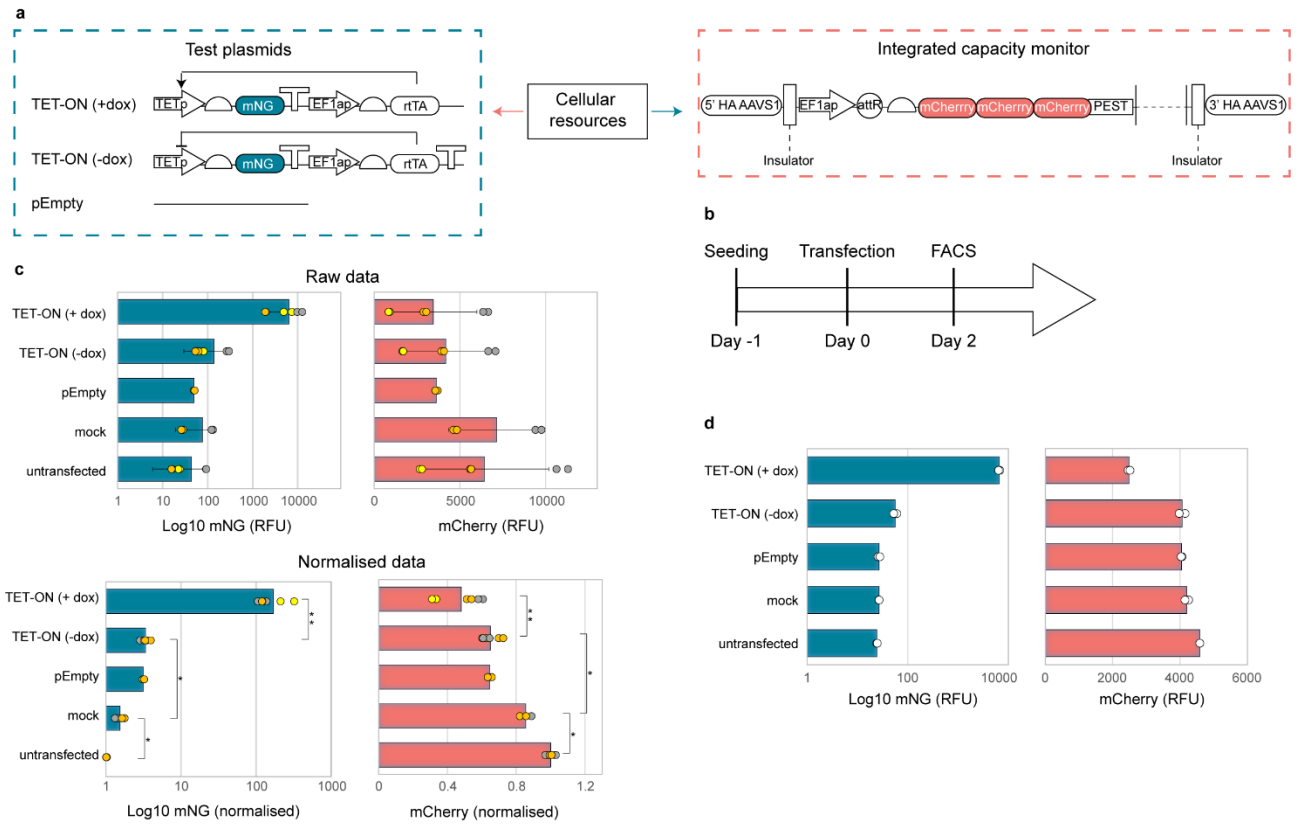

Figure 4.5 Testing of capacity monitor v2. (a) We transfected the capacity monitor cell line with test plasmids of diverse composition: a pEmpty as control of zero resource footprint, an uninduced TET-ON system (dox 0ng/uL), an induced TET-ON system (dox 1ng/uL). **(b)** Timeline of the experiment. **(c)** Test plasmid (log10 mNG) and capacity monitor (mCherry) expression levels reported as mean RFU ± std (top panels) or as normalised RFU (bottom panels). Normalised values were calculated by dividing the RFU of each data point by the averaged RFU of the untransfected condition run on the same day. Data points of different colours have been acquired in different experiments. **(d)** We repeated the experiment in panel c using FugeneHD as transfection reagent. Test plasmid (log10 mNG) and capacity monitor (mCherry) expression levels reported as mean RFU ± std. Number of independent experiments ≥ 1 (depending on the condition); two technical repeats per experiment. Two-sided Mann-Whitney test P value: ****<0.0001, ***<0.0005, **<0.005, *<0.05. Statistical significance was not calculated for samples with number of independent experiments less than two. Data analysis is described in Materials and methods (Chapter 7).

Overall, we observed a decrease in the capacity monitor expression when comparing the pEmpty condition to the untransfected sample, with a further minor decrease observed for the induced TET-ON system (Figure 4.5c – bottom panels). This indicates that much of the decrease in the capacity monitor expression is due to plasmid DNA uptake rather than to the resource footprint of the test plasmids (Figure 4.5c – bottom panels). To solve this issue, we repeated the experiment using FugeneHD as transfection reagent, which was shown to have minimal impact on cell physiology and endogenous gene expression profiles[178]. This time, we observed a decrease in the capacity monitor expression levels only upon transfection of the induced TET-ON plasmid (Figure 4.5d), establishing proof of concept of the use of integrated capacity

monitors to visualise the resource footprint of transfected payloads in mammalian cells. This has never been reported in mammalian cells so far.

4.2.4. Integrated capacity monitor v3 - testing of an inducible capacity monitor

Although we established proof of concept, using our current design as the final capacity monitor was impractical because of the silencing mentioned before. Since the landing pad master cell line showed stable expression over time, we reasoned we had to integrate the capacity monitor from scratch using CRISPR-Cas9 to solve the stability problem, similarly to what we did for the landing pad. This is a time-consuming process, and before embarking on it, we wanted to be sure to have individuated the best capacity monitor design for the integration.

An inducible capacity monitor might perform better than a constitutive one, as the expression of the monitor can be induced on demand as soon as resource competition is imposed. This situation more closely resembles what we experienced using our transient capacity monitor. To test our claim, we used an integrated cell line gifted to us from Fowler's laboratory[152]. This cell line was similar in design to our previous monitor version, as it displayed a single-copy integration at the AAVS1 locus of HEK293T cells. However, in this case, the integrated cassette encoded for the expression of a TET-ON-controlled BFP without appended degradation domains. We used this cell line as our first version of inducible capacity monitor and transfected it with a library of test plasmids of increasing resource footprints (Figure 4.6a). One day after transfection, we induced the capacity monitor expression, thus ensuring that the monitor was produced only after the test plasmids had migrated into the nucleus and started being expressed (Figure 4.4b). Interestingly, also in this case, we observed high variability in the capacity monitor expression amongst different experiments, which was dramatically reduced after normalisation (Figure 4.6c). Overall, we observed a progressive decrease in the capacity monitor expression with increasing outputs from the test plasmids. We also noticed a further decrease when the capacity monitor was transfected with the two-transcriptional-unit test plasmid (Figure 4.6c – 2 TU), which has a higher resource footprint than payloads bearing one transcriptional unit. These were our best results based on the range of suppression of the capacity monitor and the resolution achieved (i.e. the capacity monitor was able to discern between one-transcriptional-unit payloads carrying different promoters). This experiment confirmed that inducible capacity monitors are a better alternative to constitutive ones.

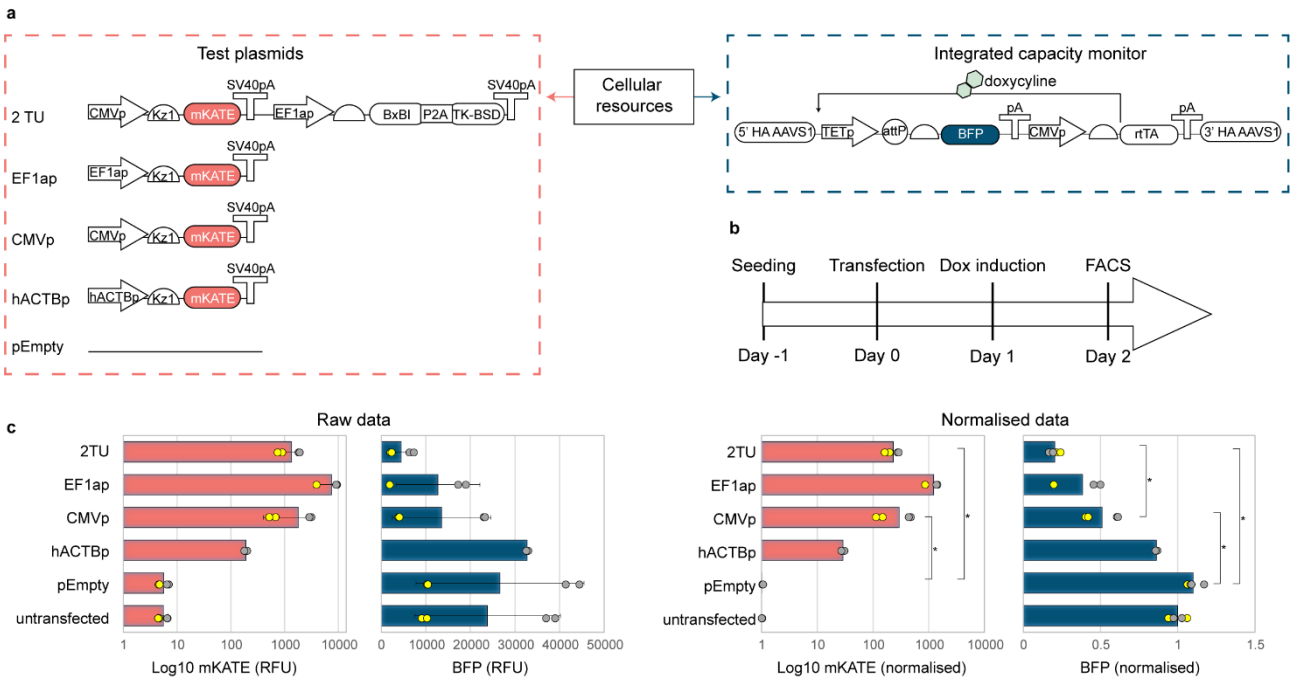

Figure 4.6 Testing of capacity monitor v3. (a) We transfected the capacity monitor cell line with test plasmids of increasing resource footprints, including a pEmpty, three one-transcriptional-unit plasmids bearing different promoters, and one plasmid carrying two transcriptional units. **(b)** Timeline of the experiment. The capacity monitor was induced one day after transfection. **(c)** Test plasmids (log10 mKATE) and capacity monitor (BFP) expression levels reported as mean RFU ± std (left panels) or as normalised RFU (right panels). Normalised values were calculated by dividing the RFU of each data point by the averaged RFU of the untransfected condition run on the same day. Data points of different colours have been acquired in different experiments. Number of independent experiments ≥ 1 (depending on the condition); two technical repeats per experiment. Two-sided Mann-Whitney test P value: ****<0.0001, ***<0.0005, **<0.005, *<0.05. Statistical significance was not calculated for samples with number of independent experiments less than two. Data analysis is described in Materials and methods (Chapter 7).

4.2.5. Integrated capacity monitor v4 – generation and testing of an inducible capacity monitor with increased protein turnover

4.2.5.1. Generation of capacity monitor v4

One key lesson learnt from previous iterations was that it is possible to optimise the dynamics of the capacity monitor by either increasing protein turnover or enabling inducible expression. Thus, we combined these approaches in our final capacity monitor design. We cloned a TET-ON-based inducible capacity monitor, where the TETp drives the expression of three mCherry genes fused together and appended with an N-terminal and C-terminal degradation domains (i.e. UbVR and PEST, respectively) (Figure 4.7a). We integrated this design at the AAVS1 locus in HEK293T, similar to what we did to generate the initial landing pad (Figure 4.7b). We screened for clones carrying a single-copy integration and selected clone 3-B7 for further experiments. This clone showed correct PCR-genotyping profile (Figure 4.7c), displayed the

expected induction behaviour in the presence of the inducer (Figure 4.7d), and was stable over prolonged cultivation (data not shown).

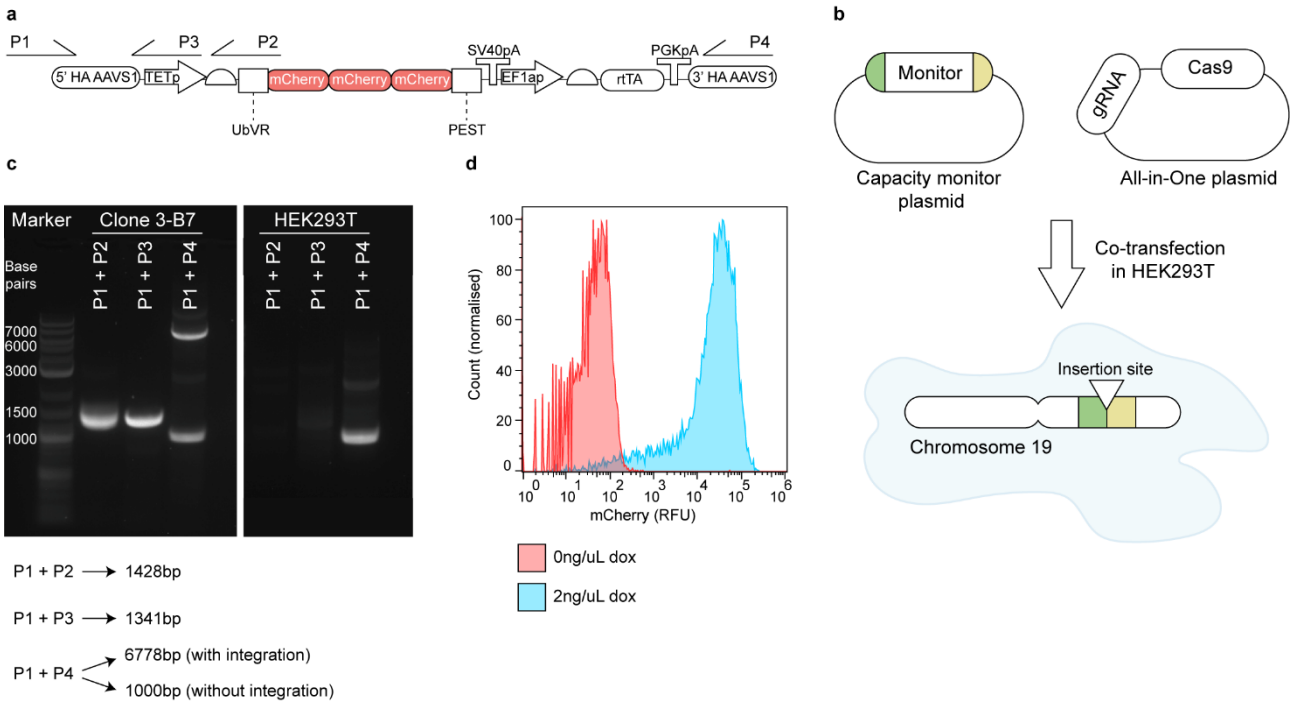

Figure 4.7 CRISPR-Cas9 integration of capacity monitor v4 in HEK293T. (a) Design of an AAVS1-integrative inducible capacity monitor with increased protein turnover. A TETp drives the expression of three mCherry fused together and appended with two degradation domains (UbVR and PEST). **(b)** The capacity monitor plasmid was co-transfected with an All-in-One plasmid encoding both the Cas9 and the gRNA necessary for integration into the AAVS1 locus. Upon co-transfection in HEK293T, the Cas9+gRNA introduces a double-strand break at the AAVS1 locus, which is repaired by homologous recombination with the AAVS1 homology arms flanking the capacity monitor (coloured in the figure). This results in the insertion of the capacity monitor at the target genomic location. **(c)** PCR genotyping of a heterozygote clone (3-B7) showing correct integration patterns. HEK293T were included as a negative control. Although all the samples were run together, the gel was cropped to eliminate unnecessary samples, which might have confounded the reader. **(d)** Fluorescence (mCherry) histogram of the selected clone 3-B7 at uninduced (dox 0ng/uL) and induced (dox 2ng/uL) conditions, showing homogeneous expression and correct induction behaviour of the integrated clone. Capacity monitor fluorescence is reported as RFU.

4.2.5.2. Testing of capacity monitor v4

As an initial test, we transfected the capacity monitor either with a pEmpty or an EGFP-expressing payload (Figure 4.8a) and induced capacity monitor expression one day after transfection (Figure 4.8b). The new capacity monitor design could distinguish between the pEmpty and the CMVp payload (Figure 4.8c). We also noticed that by doubling the amount of transfected plasmid from 0.08pmol to 0.16pmol, the change of capacity monitor expression between the pEmpty and the CMVp payloads was more dramatic (Figure 4.8c). Based on this, we transfected 0.16pmol in later experiments.

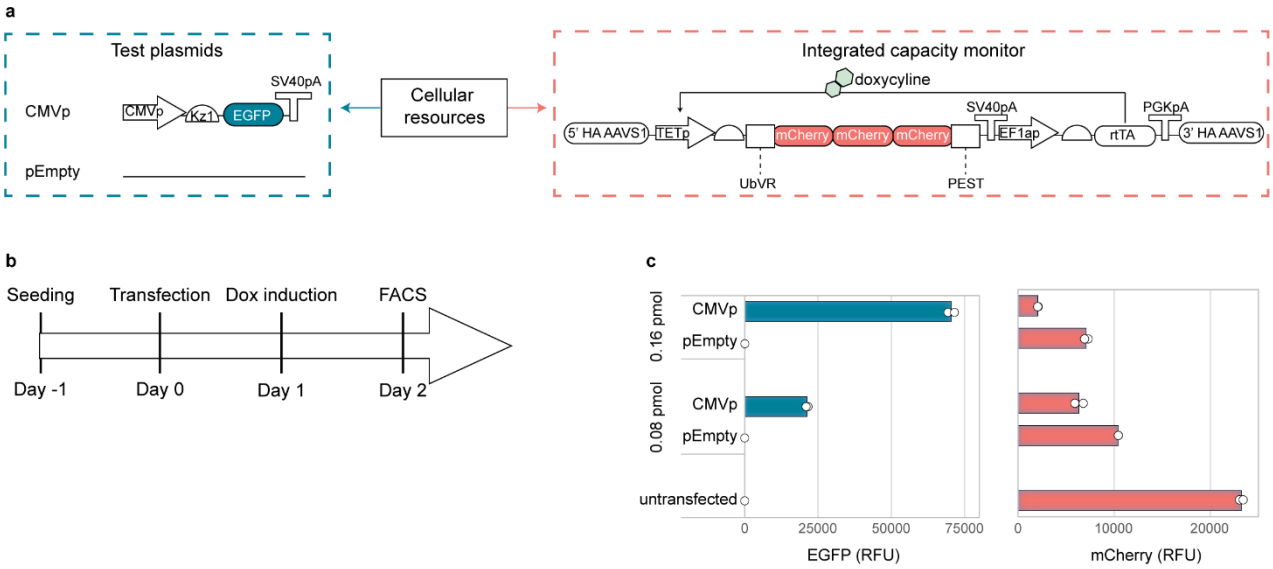

Figure 4.8 Testing of capacity monitor v4. (a) We transfected the capacity monitor cell line with either a pEmpty or a CMVp-driven EGFP. **(b)** Timeline of the experiment. The capacity monitor was induced one day after transfection. **(c)** Test plasmids (EGFP) and capacity monitor (mCherry) expression levels at different molar amounts of transfected DNA (0.08pmol and 0.16pmol). Expression levels are reported as mean RFU ± std. Number of independent experiments = 1; two technical repeats per experiment. Statistical significance was not calculated as the number of independent experiments is less than two for all samples. Data analysis is described in Materials and methods (Chapter 7).

4.2.6. Screening libraries of genetic designs and growth considerations

Proved that the new integrated capacity monitor design worked as expected, we used it to quantify the resource burden of a library of genetic constructs. While the following experiments are similar to those presented in Chapter 2, in this case, the integrated monitor provides insights on the impact that different resource-demanding genetic designs place on cellular physiology. This is fundamentally different from the experiments in Chapter 2, where only transient capacity monitors were used, and these considerations were not possible.

First, we tested the impact of our previously-characterised promoter library on the integrated capacity monitor (Figure 4.9a). As expected, we observed that increasing expression from the test plasmids led to decreased fluorescence output from the capacity monitor, with few exceptions (Figure 4.9b). For instance, the hACTBp and CMVp configurations both resulted in higher capacity monitor output than what would be inferred based on their expression levels. Interestingly, our results were in good accordance with the ones obtained using the transient capacity monitor in Chapter 2. We reported a positive linear correlation in the expression levels of the promoter library when in competition with the transient or the integrated capacity monitors (Figure 4.9c – left panel). Moreover, a positive linear correlation was also observed when comparing the transient and the integrated capacity monitor output levels when in competition with the promoter library (Figure 4.9c – right panel). This further corroborates our previous results and reiterates the importance of promoter selection for resource considerations.

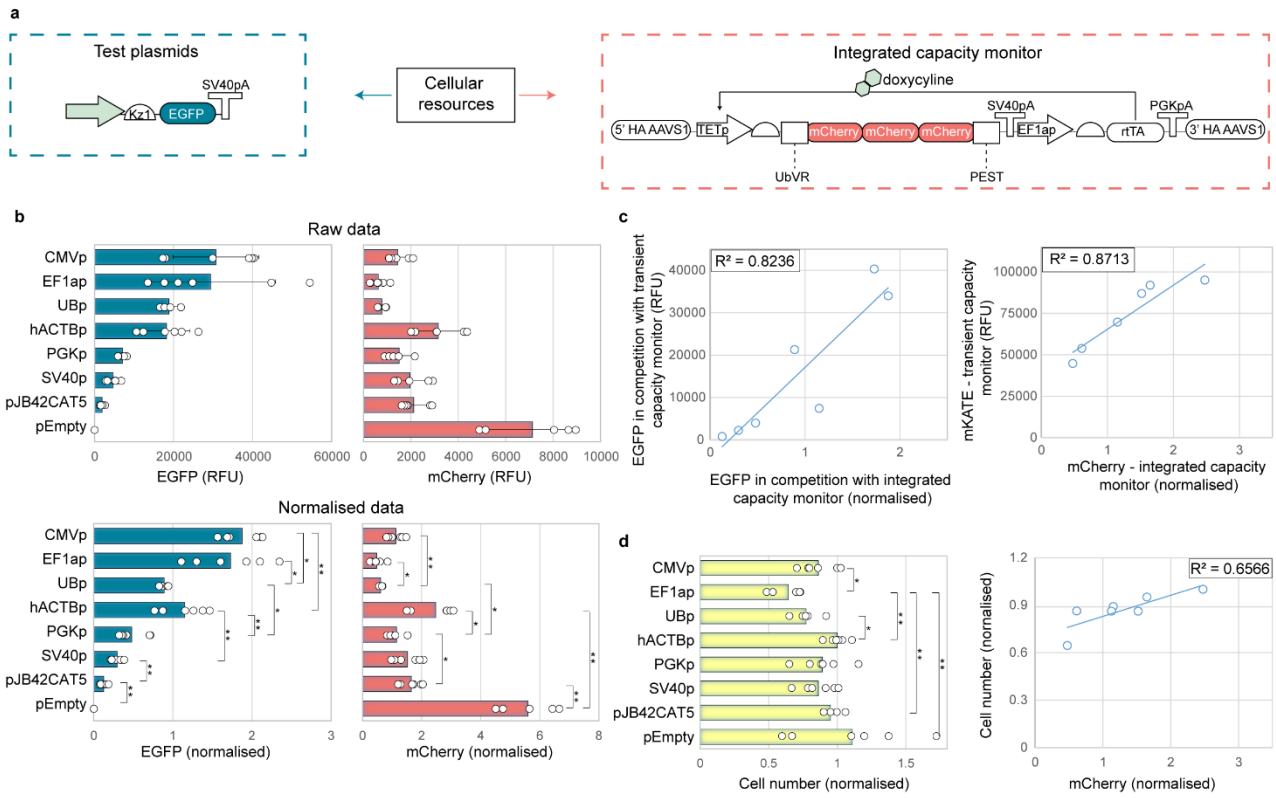

Figure 4.9 Relationship between capacity monitor expression and cellular growth for a library of test designs with different promoters. (a) We transfected the capacity monitor cell line with a library of test plasmids, each bearing one of seven promoter configurations. **(b)** Test plasmid (EGFP) and capacity monitor (mCherry) expression levels were reported as RFU ± std (top panels) or as or as normalised RFU (bottom panels). Normalised values have been calculated by dividing the RFU value for each data point against the averaged RFU of a control sample run on the same day. **(c)** Positive linear correlation between the expression levels of test plasmids in competition with the transient capacity monitor and in competition with the integrated capacity monitor (left panel – Pearson coefficient 0.907). Positive linear correlation between the expression levels of the transient and integrated capacity monitor when in competition with the tested designs (right panel – Pearson coefficient 0.933). **(d)** Normalised live cell number at the end of the transfection experiments for the tested designs (left panel). Cell counts for each sample were normalised to the averaged cell counts of the hACTB sample run on the same day. Positive linear correlation between capacity monitor expression levels and cell number (right panel, Pearson coefficient 0.81). Number of independent experiments = 3; two technical repeats per experiment. Two-sided Mann-Whitney test P value: ****<0.0001, ***<0.0005, **<0.005, *<0.05. Data analysis is described in Materials and methods (Chapter 7).

As stated before, one major advantage of integrated capacity monitors over transient capacity monitors is that they can inform on the burden of heterologous expression on endogenous genes and, ultimately, cell physiology. In previous research in bacteria, capacity monitors were good predictors of cellular growth, with lower production rates from the monitors indicating slower growth rates due to cellular burden[61]. To investigate whether similar considerations can be translated to mammalian cells, we performed cell counts at the end of the competition experiment. We observed different cell numbers across wells transfected

with different genetic constructs (Figure 4.9d – left panel). More importantly, we observed a positive linear correlation between the capacity monitor output and the cell number (Figure 4.9 – right panel, Pearson coefficient of 0.81), indicating that our integrated capacity monitor can be used as a proxy measure of mammalian cellular growth. This result is even more relevant if put into perspective. Over the course of my PhD, one of the main questions regarded what cellular burden would look like in mammalian cells. While initial evidence showed that heterologous expression leads to altered expression of endogenous genes and slowed cellular growth in mammalian cells, we here show that the design of genetic constructs directly affects cellular burden and growth profiles. Our integrated capacity monitor stands as a tool to enable thorough screening and characterisation of the cellular burden of libraries of genetic constructs, streamlining the identification of low-burden designs and reporting on the cellular stress caused by heterologous expression in real time. Ultimately, with our experiments, we proved that mammalian cells experienced cellular burden following the depletion of key gene expression resources and that this translates into growth defects that can be seen just after two days of culture. In addition to this, we transfected different promoter-polyA combinations into our capacity monitor cell line (Figure 4.10a). Overall, the differences in the capacity monitor expression due to polyA variations were minor compared to the effect of promoters, with few exceptions (i.e. CMVp-SV40pA and EF1ap-SV40pA) (Figure 4.10b).

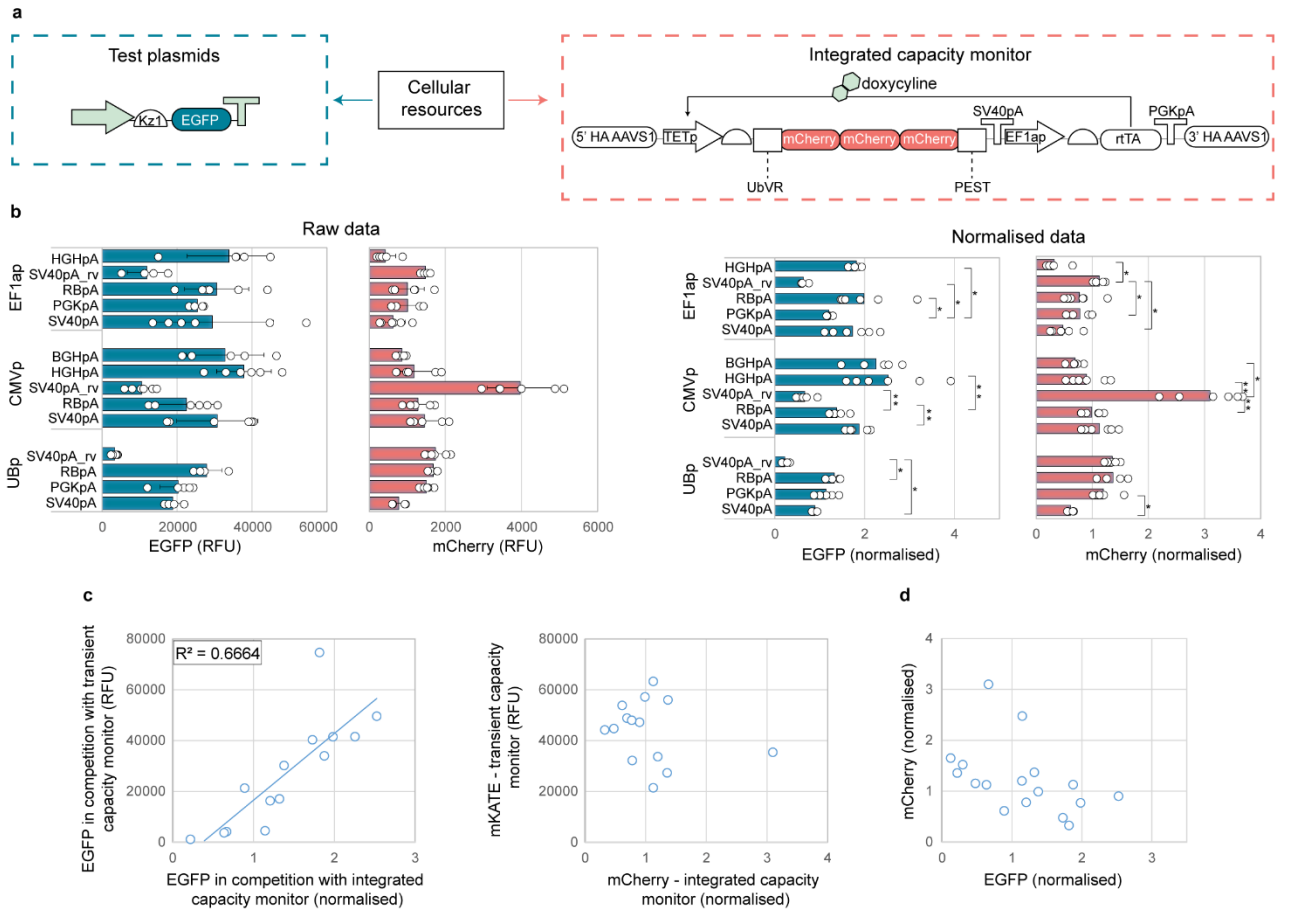

Figure 4.10 Impact of different promoter-polyA combinations on capacity monitor expression. (a) We transfected the capacity monitor cell line with a library of test plasmids bearing different promoter-polyA combinations. **(b)** Test

plasmid (EGFP) and capacity monitor (mCherry) expression levels reported as RFU ± std (left panels) or as or as normalised RFU (right panels). Normalised values have been calculated by dividing the RFU value for each data point against the averaged RFU of a control sample run on the same day. **(c)** Positive linear correlation between the expression levels of test plasmids in competition with the transient capacity monitor and in competition with the integrated capacity monitor (left panel – Pearson coefficient 0.816). No linear relationship was observed between the expression levels of the transient and integrated capacity monitor when in competition with the tested designs (right panel – Pearson coefficient -0.249). **(d)** General trend of increasing test plasmid output leading to decreased capacity monitor expression levels. No inverse linear correlation was observed (Pearson coefficient -0.478). Test plasmid and capacity monitor expression levels were reported as RFU ± or as normalised RFU. Number of independent experiments = 3; two technical repeats per experiment. Two-sided Mann-Whitney test P value: ****<0.0001, ***<0.0005, **<0.005, *<0.05. Data analysis is described in Materials and methods (Chapter 7).

Although the gene expression performance of the tested designs correlates when these compete with the transient or integrated capacity monitor (Figure 4.10c – left panel), the capacity monitor response to these designs is very different for the integrated and transient versions (Figure 4.10c - right panel). When considering all the test designs together, we noticed a general trend of increasing output from the test plasmids leading to decreased capacity monitor expression, although a linear inverse relation was not observed (Figure 4.10d). This revealed that some designs are more efficient than others, displaying increased gene expression performance and imposing a lower burden on the cell. While these considerations might seem similar to those in Chapter 2, they are different. In Chapter 2, we identified genetic designs with lower resource footprint and decreased impact on a co-transfected gene. In this Chapter, we identified genetic designs with decreased impact on the cellular host.

4.3. Conclusions

In this Chapter, we developed the first integrated capacity monitor for use in mammalian cells. Integrated capacity monitors are powerful tools to investigate the effects of resource shortages on endogenous gene expression and cellular physiology, a type of information inaccessible to simpler transient capacity monitors. We tested four design iterations before landing on a capacity monitor design with optimal characteristics. In doing so, we provided useful insights for designing integrative fast-responsive biosensing systems in mammalian cells, highlighting inducibility and modulation of protein turnover as complementary strategies to engineer dynamic systems. We demonstrated that the expression of integrated genes is affected by the expression of resource-demanding transient plasmids and showed that the capacity monitor expression levels can be used as a proxy measure of cellular growth. Finally, we used the integrated capacity monitor to screen the cellular burden of a library of transfected plasmids.

Chapter 5 - Competition for protein degradation and secretion resources

5.1. Introduction

In the previous chapters, we analysed the competition for gene expression resources in engineered mammalian cells. Other types of cellular resources have been shown to face shortages in engineered cells, including protein degradation resources in bacteria[69] and RNA degradation resources in mammalian cells[71]. Initial evidence on resource bottlenecks on the mammalian secretion pathway is also present in the literature[66], but clear, direct visualisation of competition effects is still missing. This greatly limits our understanding of the resource competition landscape in mammalian cells and prevents us from developing a global resource framework towards predictable cell engineering. To address this issue, first steps should be taken at a proper characterisation of resource bottlenecks within the protein degradation and secretion pathway, given that these two cellular processes are crucial for many applications. Protein degradation adds a layer of regulation to any gene expression process. Protein secretion resources are key for bioproduction applications, as most of the protein therapeutics produced in mammalian cells are secreted.

In this Chapter, we investigate competition effects amongst alternative pools of cellular resources. First, we obtained the first evidence of bottlenecks involving protein degradation resources in mammalian cells. We characterised the resource footprint of a library of degradation domains, putting forward considerations on the efficiency of different combinations of such domains. We proved that coupling due to resource limitations in the protein degradation pathway can also arise between integrated and transfected genes, opening to considerations for biotherapy and bioproduction. Finally, we showed clear evidence of competition for secretion resources and provided positional information on the resource bottlenecks within the secretion pathway.

Overall, this Chapter proposes a discussion on the competition for protein degradation and secretion resources in engineered mammalian cells, showing how resource-aware frameworks can be adopted to provide information on alternative pools of cellular resources.

5.2. Results and discussion

5.2.1. Competition for degradation resources

5.2.1.1. Proteasome-mediated protein degradation in mammalian cells

Intracellular protein degradation is a physiological process that cells use to maintain protein homeostasis. One of the key pathways that mammalian cells use to degrade cellular proteins is through the proteasome, which is a protein complex specialised in degrading short-lived and misfolded proteins. The main degradation route is through the ubiquitin-proteasome system (UPS), where three cascading enzymes (ubiquitin-activating enzyme E1, ubiquitin-conjugating enzyme E2, and ubiquitin-protein ligase E3) promote protein poly-ubiquitination and subsequent proteasomal degradation. Specifically, E1 binds the ubiquitin protein and transfers it to E2, which, in the presence of E3, attaches the ubiquitin to the protein of interest through a lysine isopeptide bond (Figure 5.1). In the human genome there are two genes encoding for E1 and approximately a hundred and a thousand genes encoding for E2 and E3, respectively[179]. The repeated action of these three proteins leads to the generation of poly-ubiquitin chains that will direct the target protein to proteasomal degradation[180].

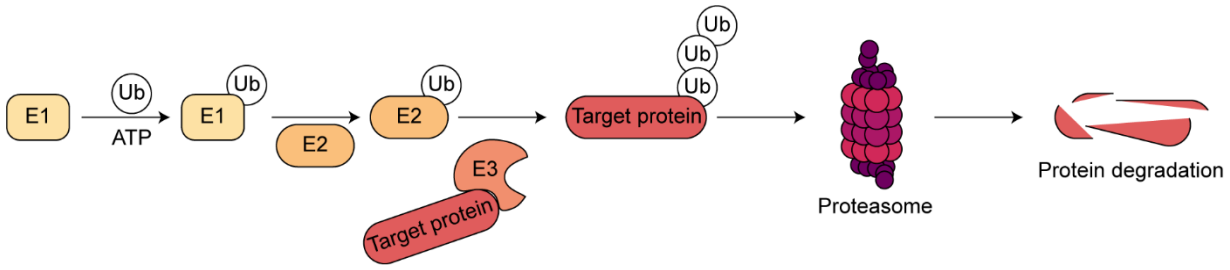

Figure 5.1 The ubiquitin-proteasome system (UPS) maintains intracellular protein homeostasis. E1 proteins bind the ubiquitin through an ATP-dependent reaction and transfer it to E2 proteins. E3 proteins will then catalyse the transfer of the E2-bound ubiquitin to the target protein through an isopeptide bond. The repeated action of this enzymatic cascade poly-ubiquitinates the target protein, marking it for proteasome-mediated degradation.

Many aminoacidic degradation domains that target proteins to proteasomal degradation have been described[181]. For instance, Chassin *et al.* characterised a library of degradation domains to tune protein turnover[181]. Most tags included an N-terminal ubiquitin protein with either an intact or a mutated isopeptidase site. Degradation domains with an intact isopeptidase site are recognised by deubiquitinating enzymes, which cleave ubiquitin, uncovering a stabilising or destabilising aminoacid at the N-terminus of the target protein. Degradation domains with a mutated isopeptidase site are not recognised by isopeptidase enzymes, leaving ubiquitinated proteins targeted for degradation. Eventually, both types of degradation domains lead to proteasome-mediated degradation. In this Chapter, we used a selection of degradation domains characterised by Chassin *et al.*[181], including two N-terminal ubiquitin tags with active

isopeptidase site (UbVR, 2xUbAV), two N-terminal ubiquitin tags with mutated isopeptidase site (UbR, UbM), and one C-terminal degradation domain (PEST).

5.2.1.2. Resource-aware framework to investigate competition for protein degradation resources

To investigate the effects of resource competition on protein degradation pathways, we reasoned that transient capacity monitors with increased protein turnover could represent appropriate visualisation tools. Therefore, we modified our previous transient capacity monitor design by appending to the readout mKATE protein one or more degradation domains (Figure 5.2). By doing so, we ensured that the mKATE actively engaged degradation resources, making its intracellular levels responsive to variations in the pool of such resources. Based on previous work in bacteria[69], we envisioned that if degradation resources were limiting in engineered mammalian cell lines, this would lead to coupling of the intracellular protein levels of the actively degraded test constructs and capacity monitor. The resource footprint of different degradation domains can then be inferred by co-transfection with the capacity monitor and monitoring of its intracellular accumulation levels through flow cytometry (Figure 5.2).

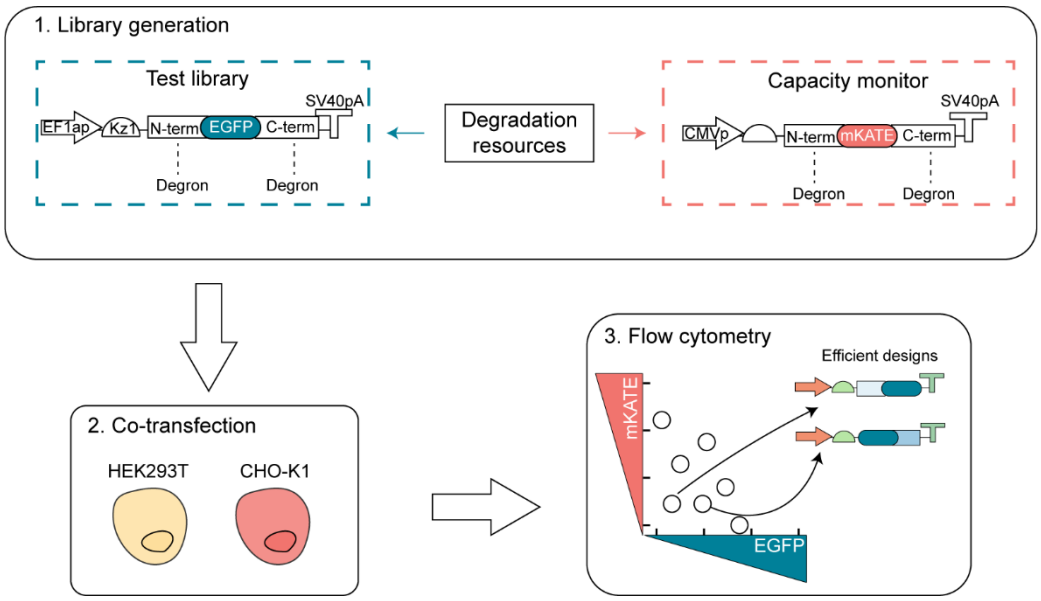

Figure 5.2 Resource-aware framework to screen libraries of degradation domains. Schematics of the experimental workflow. A library of test constructs with different degradation domains is co-transfected with a destabilised capacity monitor in HEK293T and CHO-K1. Intracellular protein levels are quantified two days post-transfection using flow cytometry. Competition for degradation resources can be visualised by plotting the capacity monitor protein levels (mKATE) over the test constructs protein levels (EGFP). In this plot, the constructs localising at the bottom left are the most efficient, as they are actively degraded (EGFP) while having a low resource footprint (i.e. they do not impact the degradation of the mKATE capacity monitor).

To obtain proof of concept, isolating the competition for protein degradation resources from the more general competition for gene expression resources was crucial to our experimental design. Therefore, we devised an experiment where a PEST-tagged capacity monitor was co-transfected either with a PEST-tagged

or an untagged EGFP-expressing test construct (Figure 5.3a). Importantly, the two test constructs shared the same plasmid backbone, promoter, Kozak, and polyA signal and differed only for the PEST domain. This ensured that they had a similar footprint on gene expression resources, with the only change being their footprint on degradation resources. We envisioned that coupling of the PEST-tagged test plasmid and the PEST-tagged capacity monitor outputs would be observed in a condition of competition. This would be consistent with the PEST-tagged EGFP demanding more protein degradation resources than its untagged version and fewer resources being available to the mKATE-PEST for efficient degradation at its usual rates. As expected, we observed a decrease in capacity monitor expression upon co-transfection with the EGFP-expressing plasmid compared to the pEmpty condition (Figure 5.3b). This is consistent with our previous findings, and it is due to resource bottlenecks in the gene expression pathway. However, co-transfection with the PEST-tagged EGFP-expressing plasmid partially rescued the capacity monitor fluorescence (Figure 5.3b). We argue that this rescue is not due to increased expression levels of the mKATE gene but is a result of decreased mKATE degradation, similar to previously published evidence in bacteria[69]. Our experiment suggests that the change in the capacity monitor degradation rate is dependent on the addition of a PEST signal on the competing test construct, pointing to a resource competition scenario. The PEST-tagged EGFP demands more protein degradation resources than its untagged counterpart. In doing so, fewer resources will be available to the PEST-tagged capacity monitor to be efficiently degraded at its usual rates, leading to positive intracellular accumulation of the mKATE protein.

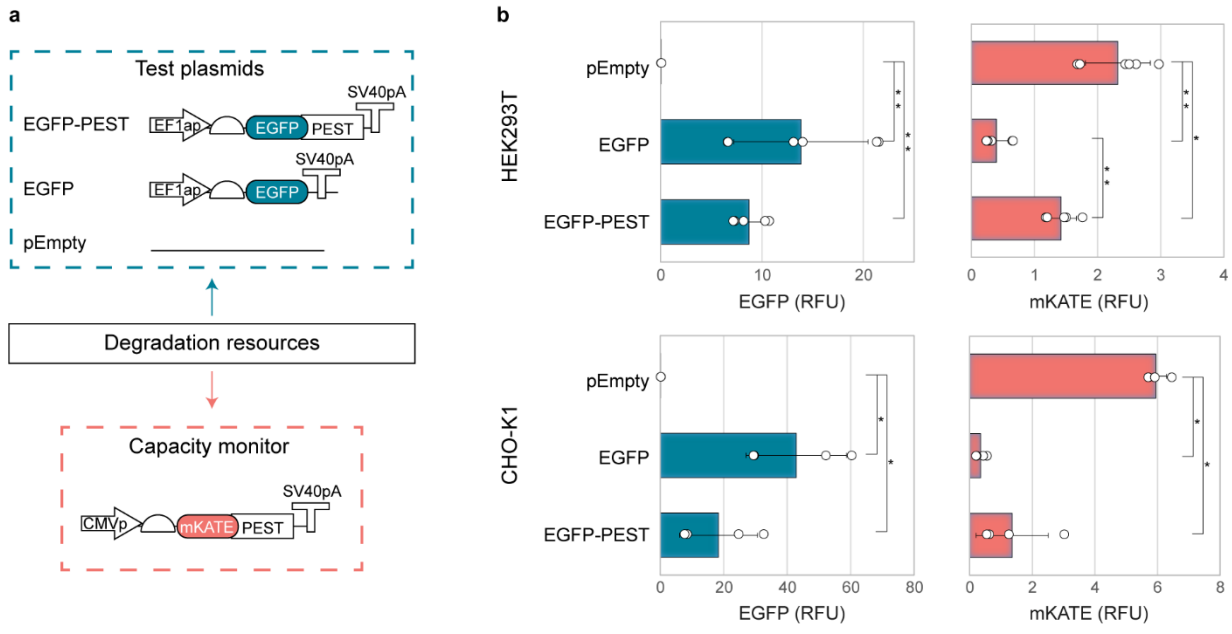

Figure 5.3 Competition for degradation resources can be visualised using destabilised transient capacity monitors. **(a)** We co-transfected a PEST-tagged capacity monitor either with a PEST-tagged or an untagged EGFP-expressing test construct. The two test constructs have been cloned starting from the same backbone. They share the same promoter, Kozak, and polyA signal and differ only in the presence or absence of a PEST degradation domain. **(b)** Test plasmid (EGFP) and capacity monitor (mKATE) intracellular protein levels in HEK293T (top panels) and CHO-K1 (bottom

panels). Test plasmid and capacity monitor intracellular protein levels are reported as mean RFU ± std. Number of independent experiments ≥ 2 (depending on the condition); two technical repeats per experiment. Two-sided Mann-Whitney test P value: ****<0.0001, ***<0.0005, **<0.005, *<0.05. Data analysis is described in Materials and methods (Chapter 7).

5.2.1.3. Resource footprint of a library of degradation domains

Once confirmed that protein degradation resources undergo limitations in engineered mammalian cells, we focused on two key questions opening to considerations on informed resource-aware design. The first question is whether different degradation domains had different resource footprints, and the second is whether orthogonal combinations of degradation domains with reduced coupling effects could be identified. To act on these points, we cloned a library of EGFP-expressing test plasmids appended with different degradation domains to investigate the resource footprint of different modifications. Our library contained two N-terminal ubiquitin tags with active isopeptidase site (UbVR, 2xUbAV), two N-terminal ubiquitin tags with mutated isopeptidase site (UbR, UbM), and one C-terminal degradation domain (PEST)[181] (Figure 5.4a). All these variants have been cloned starting from the same backbone to obtain a library of test constructs differing only for the degradation domain appended to the EGFP. In addition, we cloned a panel of three transient capacity monitors where the mKATE was linked to a C-terminal PEST domain, an N-terminal UbVR domain, or a combination of both (Figure 5.4a). To set up reference points in our analysis, we transfected the capacity monitors with a pEmpty or with an untagged EGFP. While the pEmpty gave us information on the intracellular levels of the tagged capacity monitors in the absence of competition for gene expression and protein degradation resources, the untagged EGFP gave us information on the intracellular levels of the tagged capacity monitors in the presence of competition for gene expression resources and in the absence of competition for protein degradation resources. We reasoned that the library of degradation-tagged EGFP constructs would lead to mKATE intracellular levels that are lower than the pEmpty condition and higher than the untagged EGFP. As expected, our library of degradation domains effectively modulated EGFP intracellular concentration, as evidenced by flow cytometry measurements (Figure 5.4b). In accordance with our previous results, increased degradation rates of the library of degradation domains (i.e. decreased EGFP accumulation) led to decreased degradation rates of all three competing capacity monitors (i.e. increased mKATE accumulation) (Figure 5.4b-c). Interestingly, when in competition with the test construct tagged with UbVR+PEST, the capacity monitors intracellular levels are rescued to a condition of no competition (pEmpty) regardless of the test constructs depleting gene expression resources. This shows the directionality and the strength of the competition for gene expression and protein degradation resources. While resource bottlenecks in the gene expression pathway result in output suppression of the capacity monitor, competition for degradation resources leads to increased intracellular accumulation of the capacity monitor. This introduces complexity to the resource competition landscape discussed in earlier chapters, extending its boundaries. To

accurately explain the intracellular levels of two competing proteins, the combined impact of resource limitations on the gene expression and protein degradation pathways must be considered, as it determines the net variation in intracellular protein accumulation.

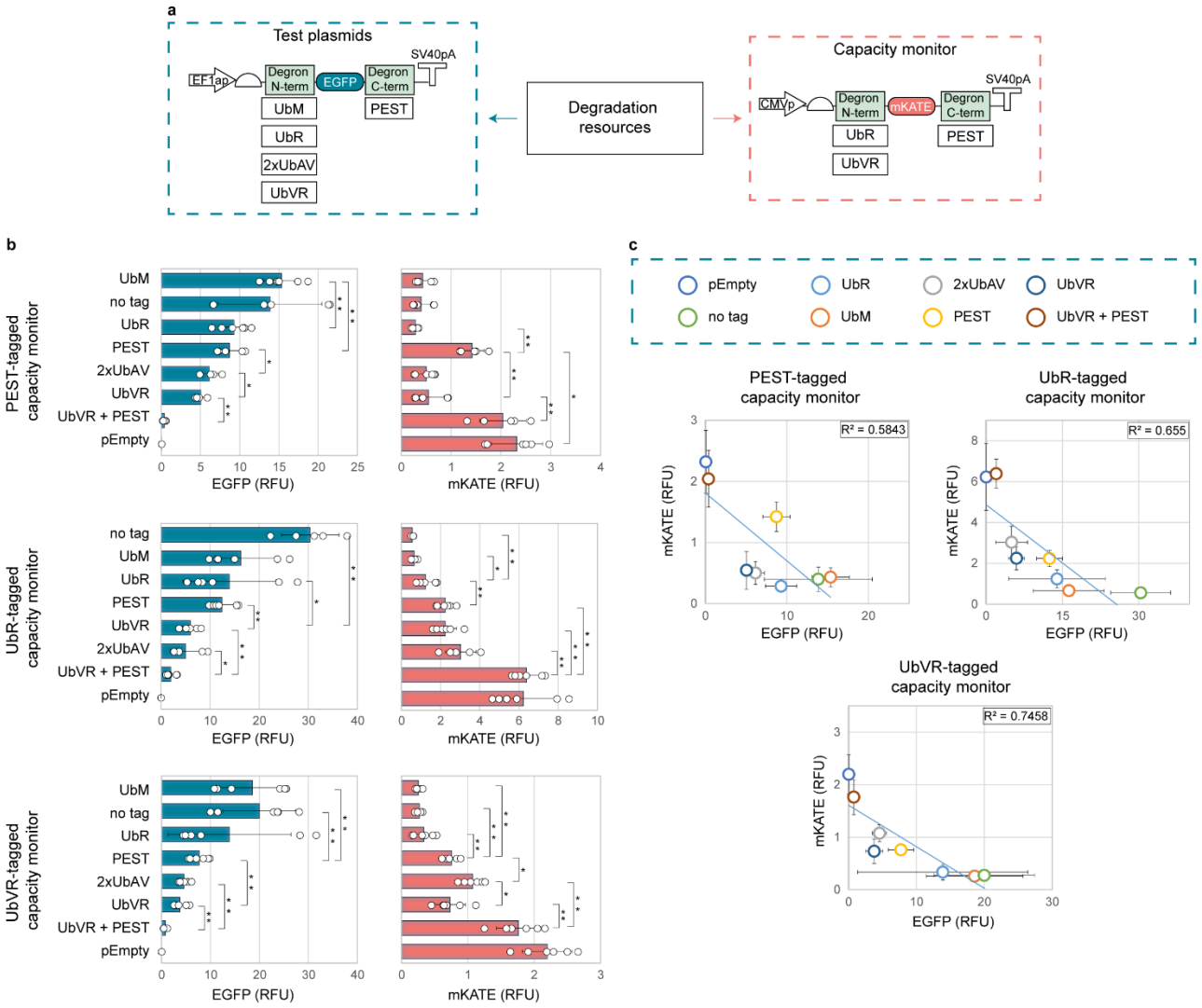

Figure 5.4 Resource footprint of a library of degradation domains in HEK293T. (a) We cloned a library of test constructs bearing different N-terminal and C-terminal degradation domains. We also generated three transient capacity monitors appended with one of two N-terminal tags (UbR and UbVR) and one C-terminal tag (UbVR). The test plasmid library was co-transfected with the panel of capacity monitor in HEK923T. **(b)** Test plasmid (EGFP) and capacity monitor (mKATE) intracellular protein levels for competition experiments with the PEST-tagged monitor (top panels), the UbR-tagged (middle panels), and the UbVR-tagged (bottom panels) capacity monitors in HEK293T. **(c)** Negative linear relation between the test plasmid (EGFP) and the capacity monitor (mKATE) intracellular protein levels for competition experiments with the three capacity monitors. Pearson coefficients -0.764, -0.809, and -0.864 for the PEST-tagged, the UbR-tagged, and the UbVR-tagged capacity monitors, respectively. Test plasmid and capacity monitor intracellular protein levels are reported as mean RFU ± std. Number of independent experiments = 3; two technical repeats per experiment. Two-sided Mann-Whitney test P value: ****<0.0001, ***<0.0005, **<0.005, *<0.05. Data analysis is described in Materials and methods (Chapter 7).

We observed inverse linear relations between the test constructs and the capacity monitor intracellular protein levels for all three capacity monitor configurations (Pearson coefficients of -0.764, -0.809, -0.864 for the PEST-tagged, the UbR-tagged, and the UbVR-tagged capacity monitors, respectively) (Figure 5.4c). This suggests that the resources undergoing limitations might be shared amongst all degradation domains (i.e. ubiquitin, proteasomal proteins), and no orthogonal combinations of domains can be identified. However, we noticed that some domain combinations were less efficient than others – where efficiency, in this case, entails simultaneous active degradation of the test construct and the capacity monitor proteins. For instance, an unusually high degree of coupling was recorded upon competition of the EGFP-PEST test construct and the mKATE-PEST capacity monitor. This is explainable by an additional bottleneck involving PEST-specific degradation resources. Despite all the tested degradation domains eventually converging to proteasomal degradation and recruiting a pool of common degradation resources, such as ubiquitin or proteasomal proteins, they might also require some domain-specific resources, such as E2 and E3 proteins. This might explain the higher degree of coupling we see for the EGFP-PEST and mKATE-PEST combination, as besides competing from the pool of degradation resources shared by all tags, they might also face shortages of cellular resources specifically involved in the degradation of PEST-tagged proteins.

5.2.1.4. Inducible degradation domains enable complete decoupling of competition for gene expression and protein degradation resources

Although the test constructs used in previous experiments differ only for the sequence of the degradation domain – and therefore reasonably have the same footprint on transcriptional and translational resources - slight modifications to the N-terminal or C-terminal protein sequence might result in different expression levels. To consider the possible impact of the different degradation domains on the gene expression of the test constructs, we decided to consider two systems enabling inducible modulation of protein turnover: the auxin-inducible degradation and the TMP-inducible stabilisation technologies. Importantly, in these systems, the degradation/stabilisation domain is always appended to the protein of interest and modulation of protein turnover only happens in response to an externally supplied small-molecule inducer. This ensures that the demand for transcriptional and translational resources is the same before and after inducible modulation of protein turnover. While these two technologies enable inducible control over the stability of a protein of interest, they carry significant differences.

The auxin-degradation technology relies on three components: the *Oryza sativa* TIR1 protein (OsTIR1), a 7-KDa degron termed mAID fused to the protein of interest, and the small-molecule auxin[182]. When expressed in mammalian cells, OsTIR1 forms an Skp1-Cul1-Fbox (SCF) E3 ligase complex with endogenous components that, in the presence of auxin, binds to mAID leading to ubiquitination and proteasomal degradation of the protein it is fused to (Figure 5.5a). This technology allows inducible protein degradation in OsTIR1-expressing human cells with a half-life of 60-65 minutes[182]. We were kindly given a plasmid encoding for the expression of an OsTIR1-mAID-EGFP fusion protein by Prof. Tom Ellis. When co-

transfected with our UbVR-tagged capacity monitor design, we observed decreased capacity monitor accumulation levels upon the addition of 0.1uM of the inducer molecule 5-Phenyl-1H-indole-3-acetic acid (5-Ph-IAA) (Figure 5.5b-c). This is in accordance with our previous findings.

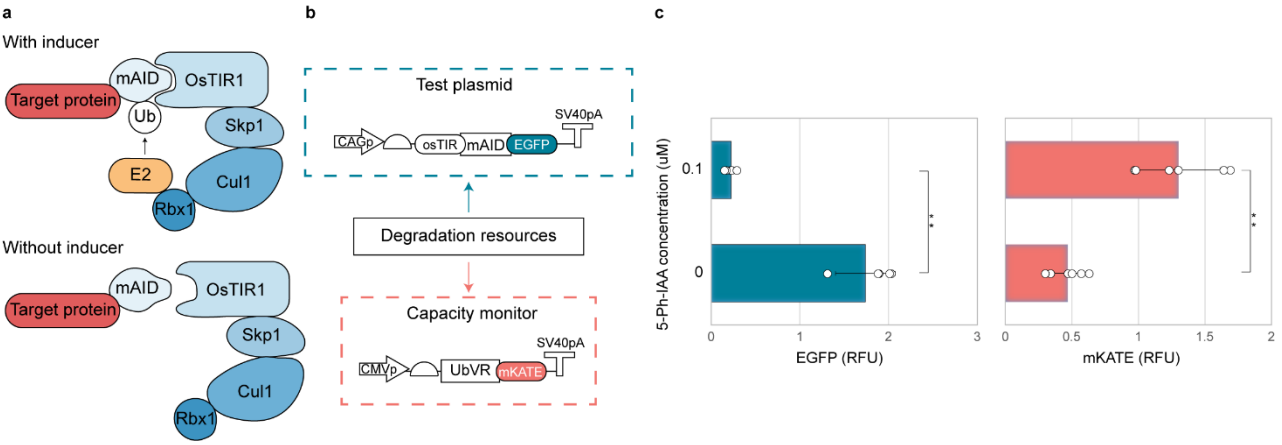

Figure 5.5 Resource footprint of the auxin-inducible degradation system in HEK293T. (a) OsTIR1 forms an Skp1-Cul1-Fbox (SCF) E3 ligase complex with mammalian endogenous components that, in the presence of auxin, binds to mAID leading to ubiquitination and proteasomal degradation of the mAID-appended protein. **(b)** A test plasmid encoding for an OsTIR1-mAID-EGFP fusion protein was co-transfected with the UbVR-tagged capacity monitor in HEK293T. **(c)** Test plasmid (EGFP) and capacity monitor (mKATE) intracellular protein levels in the absence (5-Ph-IAA 0uM) or presence (5-Ph-IAA 0.1uM) of the inducer. Test plasmid and capacity monitor intracellular protein levels are reported as mean RFU ± std. Number of independent experiments = 3; two technical repeats per experiment. Two-sided Mann-Whitney test P value: ****<0.0001, ***<0.0005, **<0.005, *<0.05. Data analysis is described in Materials and methods (Chapter 7).

As further proof, we considered TMP-inducible protein stabilisation[183]. In this system, a degradation-prone mutant of the *E. coli* dihydrofolate reductase protein (DHFR) is stabilised by the addition of the small-molecule trimethoprim (TMP) (Figure 5.6a). We cloned a plasmid encoding for a DHFR-EGFP fusion protein and co-transfected it with our UbVR-mKATE capacity monitor design. We observed decreased mKATE intracellular levels upon the addition of 0.1uM of TMP and stabilisation of the competing EGFP protein (Figure 5.6b-c).

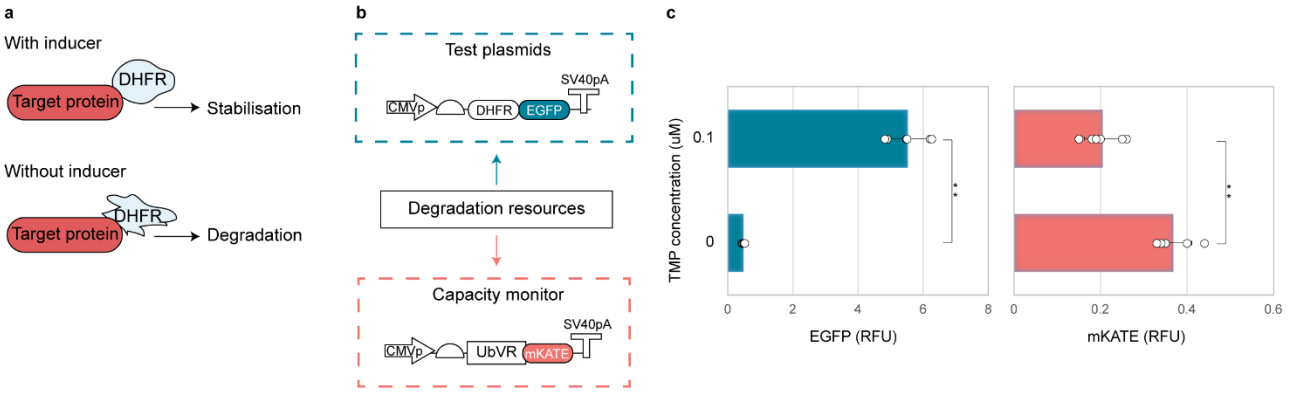

Figure 5.6 Resource footprint of the TMP-inducible degradation system in HEK293T. (a) A degradation-prone mutant of the *E. coli* dihydrofolate reductase protein (DHFR) is stabilised by the addition of the small-molecule trimethoprim (TMP) **(b)** A test plasmid encoding for a DHFP-EGFP fusion protein was co-transfected with the UbVR-tagged capacity monitor in HEK293T. **(c)** Test plasmid (EGFP) and capacity monitor (mKATE) intracellular protein levels in the absence (TMP 0uM) or presence (TMP 0.1uM) of the inducer. Test plasmid and capacity monitor intracellular protein levels are reported as mean RFU ± std. Number of independent experiments = 3. Two-sided Mann-Whitney test P value: ****<0.0001, ***<0.0005, **<0.005, *<0.05. Data analysis is described in Materials and methods (Chapter 7).

5.2.1.5. Competition for degradation resource amongst integrated and transient systems

Understanding whether competition for degradation resources also affects the behaviour of integrated genes is important, given the widespread use of integrated genetic circuits in mammalian cell engineering. It might also inspire new therapeutic avenues since dysregulated protein degradation is a hallmark of disease[184–186]. To do this, we used our inducible capacity monitor cell line described in Section 4.2.5.1, as it encoded for the expression of an actively degraded mCherry fusion protein by the appendment of two degradation domains (an N-terminal UbVR and a C-terminal PEST). We reasoned that this would make the capacity monitor cell line a good candidate to sense shortages in the degradation resources as it actively engaged such resources. We transfected the capacity monitor cell line with a library of test plasmids encoding for either an untagged EGFP or an EGFP appended with different degradation domains (Figure 5.7a). Upon transfection of tagged test plasmids, we noticed increased capacity monitor accumulation levels (Figure 5.7b). Interestingly, while the UbVR-tagged EGFP was more degraded than the PEST-tagged EGFP, the latter led to higher accumulation levels of the integrated capacity monitor (Figure 5.7b). This is in line with previous observations on the competition of two PEST-tagged proteins (Figure 5.4b-c), suggesting that some key PEST-specific cellular resources might become limiting upon the expression of two PEST-tagged proteins.

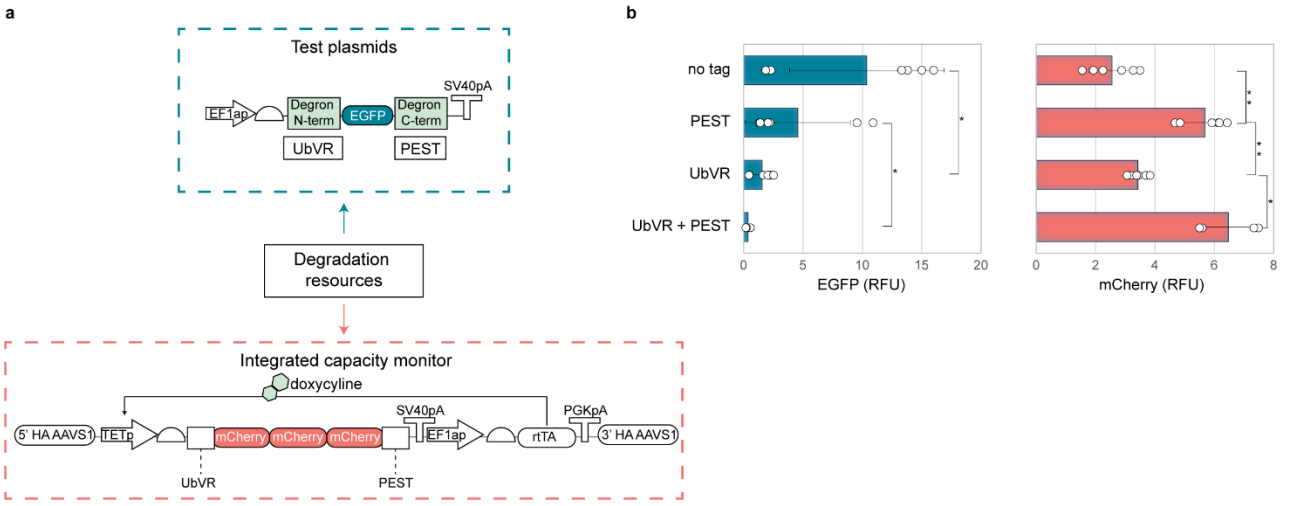

Figure 5.7 Competition for degradation resources impacts integrated systems in HEK293T. (a) We transfected our capacity monitor cell line (HEK293T) with a library of test plasmids bearing different degradation domains. **(b)** Test plasmid (EGFP) and capacity monitor (mCherry) intracellular protein levels. Test plasmid and capacity monitor intracellular protein levels are reported as mean RFU ± std. Number of independent experiments ≥ 2 (depending on the condition); two technical repeats per experiment. Two-sided Mann-Whitney test P value: ****<0.0001, ***<0.0005, **<0.005, *<0.05. Data analysis is described in Materials and methods (Chapter 7).

5.2.2. Competition for secretion resources

5.2.2.1. Protein secretion in mammalian cells

Cells generate a wide variety of secreted and membrane proteins to interact with their surroundings. In mammals, this encompasses a minimum of 2641 secreted proteins (such as enzymes, hormones, antibodies, and extracellular matrix proteins) and over 5500 membrane proteins[113]. Most of these proteins are processed within the conventional secretory pathway, a sequence of events primarily occurring within the endoplasmic reticulum (ER), Golgi apparatus, and endomembrane system. In the conventional secretory pathway, proteins targeted for secretion present a signal peptide that mediates co-translational translocation inside the ER. They are then transported into the Golgi apparatus through COPII-coated vesicles and subsequently to the plasma membrane through constitutive or controlled secretion mechanisms[187]. Each step of the secretion pathway relies on crucial cellular resources, some of which are believed to be limiting even in non-engineered cells[188], and might therefore face shortages upon heterologous expression of secreted proteins. Given the importance of protein secretion for the bioproduction industry, understanding to what extent competition for secretion resources impacts the behaviour of engineered systems is fundamental.

Recently, Gutierrez *et al.* developed a genome-scale metabolic model allowing calculation of the energy requirements of secreted proteins. Using the model, the authors effectively predicted increased monoclonal antibody production following the knock-down of unnecessary selection marker genes[113]. In addition, Kol *et al.* performed multiplexed deletion of non-essential secreted host proteins, resulting in

enhanced productivity of heterologous secreted proteins and improved cellular growth profiles[66]. While these works provide useful insights, they do not conclusively demonstrate the presence of resource bottlenecks in the secretion pathway. In other words, the increase in the secretion of heterologous proteins following the knock-down/knock-out of integrated genes, observed in these two studies, might be due to the increased availability of gene expression resources. A clear way of assessing resource limitations in the secretion pathway would benefit the community.

5.2.2.2. Resource-aware framework to investigate competition for protein degradation resources

We modified our resource-aware framework to investigate the competition for secretion resources in engineered mammalian cells. We cloned an inducible capacity monitor where the mKATE was fused to the N-terminal leader sequence from mouse immunoglobulin κ light chain (IgKL), which targets the protein to the conventional secretion pathway (Figure 5.8a). We thus generated a transient capacity monitor design that actively engaged secretion resources, representing an appropriate tool to visualise resource bottlenecks in the secretion pathway caused by the secretion of competing heterologous proteins. In the absence of competition for secretion resources, the capacity monitor mKATE protein is efficiently secreted, and its abundance in the culture media can be estimated by secreted fluorescence measurements performed with a microplate reader (Figure 5.8a). In a competition scenario where test constructs sequester secretion resources, fewer resources are available for the secretion of the capacity monitor mKATE protein, leading to a decrease in its abundance in the culture media.

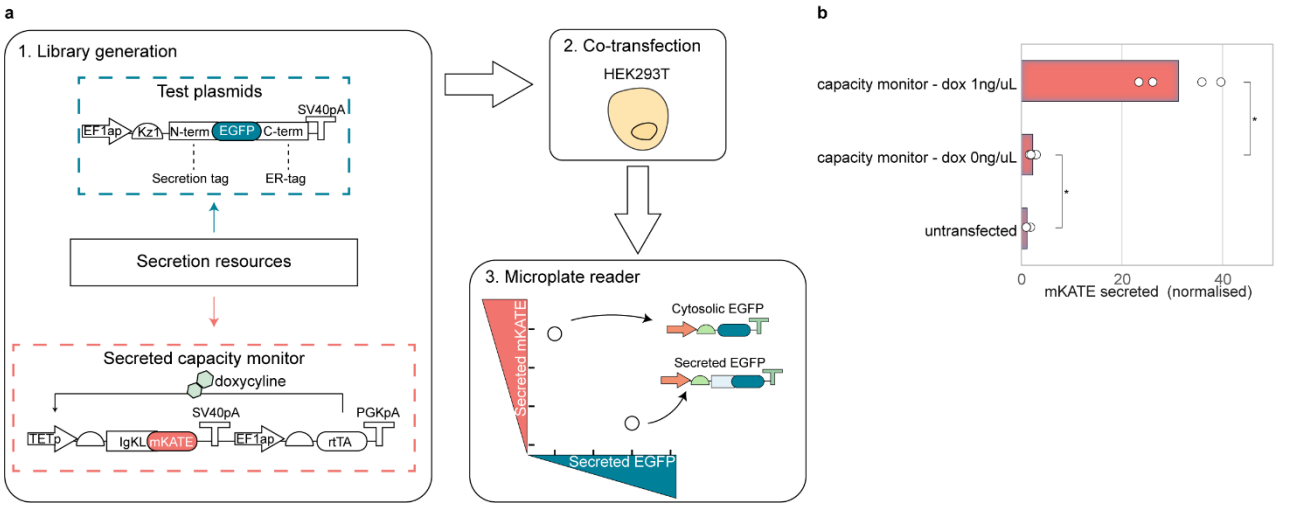

Figure 5.8 Workflow to study the competition for secretion resources in engineered mammalian cells. (a) Schematics of the experimental workflow. We cloned a dox-inducible secreted capacity monitor and co-transfected it with a library of constructs targeted to the secretion pathway in HEK293T. Secreted protein levels are quantified two days post-transfection using a microplate reader. The secreted fluorescence of the capacity monitor can be used as a proxy measure of the availability of secretion resources. **(b)** Capacity monitor's secreted protein abundance at uninduced (dox 0ng/uL) and induced (dox 1ng/uL) conditions expressed as normalised secreted fluorescence. Normalised values were calculated by dividing the RFU of each data point by the averaged RFU of the untransfected

condition run on the same day. Number of independent experiments ≥ 2 (depending on the condition); two technical repeats per experiment. Two-sided Mann-Whitney test P value: ****<0.0001, ***<0.0005, **<0.005, *<0.05. Data analysis is described in Materials and methods (Chapter 7).

As a first experiment, we confirmed that the capacity monitor displayed the intended behaviour, with induction of the capacity monitor expression leading to increased secreted fluorescence (Figure 5.8b). We then built a library of test plasmids encoding a cytosolic EGFP, a secreted EGFP, and an ER-retained EGFP (Figure 5.9a). We reasoned that this library would give us enough information for an unambiguous proof-of-concept experiment, allowing us to distinguish the different contributions of gene expression and secretion to the overall resource competition landscape. The test constructs all showed the expected behaviour, with secreted fluorescence observed only for the constructs encoding the secreted EGFP. The capacity monitor secreted fluorescence reported a first decrease upon competition with the cytosolic EGFP relative to the pEmpty condition (Figure 5.9b). This initial decrease is due to the competition for gene expression resources between the cytosolic EGFP and the secreted mKATE. A further reduction is observed upon competition with the secreted EGFP, indicating the presence of an additional resource bottleneck on the secretion pathway (Figure 5.9b). Interestingly, the ER-retained EGFP and the secreted EGFP had similar resource footprints on the secreted capacity monitor (Figure 5.9b). This gives us information on the location of the resource bottleneck within the secretory pathway, suggesting that most of the competition for secretion resources happens at the ER step of the secretory pathway.

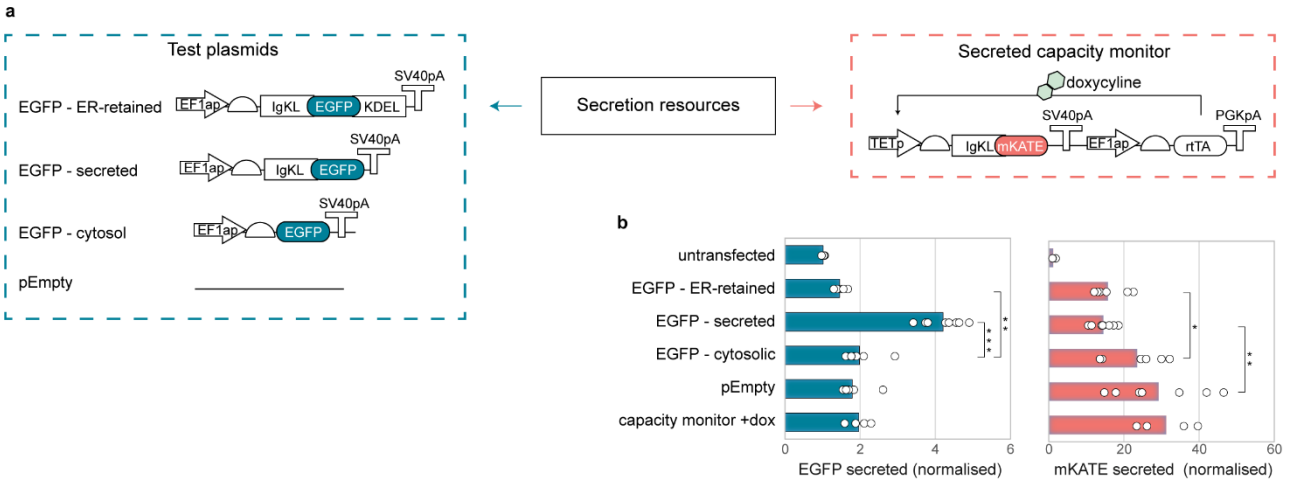

Figure 5.9 Competition for secretion resources can be visualised using a secreted transient capacity monitor in HEK293T. (a) We co-transfected a secreted capacity monitor with a library of test constructs encoding for a cytosolic EGFP, a secreted EGFP, or an ER-retained EGFP. The three test constructs have been cloned starting from the same backbone. They share the same promoter, Kozak, and polyA signal and differ only for the C-terminal or N-terminal signal peptides. **(b)** Test plasmid (EGFP) and capacity monitor (mKATE) secreted protein abundance at induced conditions (dox 1ng/uL) expressed as normalised secreted fluorescence in HEK293T. Normalised values were calculated by dividing the RFU of each data point by the averaged RFU of the untransfected condition run on the same day. Number of independent experiments ≥ 2 (depending on the condition; two technical repeats per experiment).

Two-sided Mann-Whitney test P value: ****<0.0001, ***<0.0005, **<0.005, *<0.05. Data analysis is described in Materials and methods (Chapter 7).

Of note, in these experiments, we used a very simple capacity monitor design (i.e. a secreted mKATE protein) that is probably not an optimal proxy of general secretion resources availability in mammalian cells. More complex secreted proteins (i.e. antibodies) might require more specific types of resources - such as folding and glycosylation resources - which might undergo limitations. Likely, our current secreted capacity monitor design cannot inform on the availability of these specific resource pools. Nevertheless, we provided the first direct evidence of competition for secretion resources in engineered mammalian cells, and we developed a prototype sensing tool that offers quantitative and spatial insights on such competition.

5.3. Conclusions

In this Chapter, we used resource-aware frameworks to visualise resource bottlenecks in cellular pathways other than transcription and translation. We demonstrated that competition for degradation resources leads to the coupling of the intracellular protein levels of actively degraded proteins in engineered mammalian cells. We characterised the resource footprint of a library of N-terminal and C-terminal degradation domains and observed bottlenecks in common degradation and tag-specific resources. We then recapitulated our findings using inducible degradation systems where gene expression and protein degradation are decoupled and provided the first evidence of competition for degradation resources between integrated and transient systems. Finally, we showed clear evidence of competition for secretion resources in engineered mammalian cells and suggested that resource bottlenecks localise at the ER step of the conventional secretion pathway.

Chapter 6 - Discussion

6.1. Summary

In this thesis we used a synthetic biology approach to address resource competition in engineered mammalian cells. We expanded the understanding of the phenomenon, explored resource-aware design as a new mitigation strategy, and developed tools to report on the cellular stress caused by resource shortages in key cellular pathways. In this Chapter, we will discuss our findings, including how they compare to previous literature and their limitations. We also outline a roadmap for further work, highlighting what has yet to be answered and crucial points that need addressing to develop a unified, integrated framework to understand resource competition in mammalian systems.

6.2. Our contribution to resource competition research

Since the first reports of resource competition in engineered *E. coli*[61,68], our understanding of resource limitations in engineered microbes has progressed to a point where we have many strategies to mitigate the drawbacks of shortages in key cellular pathways. Recent research pioneered similar investigations in mammalian cells, proving resource competition to be a problem even for engineered mammalian systems and proposing the engineering of adaptive circuits and the deletion of unnecessary host genes as two strategies to address the issue[48,66,72,132]. However, our understanding of the problem in mammalian systems is currently limited and key outstanding questions remain, which we also outlined in Section 1.5. First, we have limited knowledge of the cellular pathways affected in mammalian cells and the location of the resource bottlenecks within these pathways. Second, we do not know whether resource competition is limited to co-transfected systems – as shown so far – or also interests integrated genes. Third, we do not have a clear picture of the effect of resource limitation on the physiology of mammalian cells. While in microbes resource limitations lead to altered gene expression profiles and reduced cellular growth, in mammalian cells, clear evidence is still missing. Finally, the currently available mitigation strategies for use in mammalian cells have their limitations. For instance, while the implementation of adaptive circuits is effective in buffering the output of a regulated gene to disturbances in the systems, it often entails the implementation of complex genetic architectures, involves the expression of one or more resource-depleting genes, and leads to output suppression of the regulated genes. On the other end, the depletion of unnecessary host genes involves complicated engineering experiments that must be repeated every time a different host is used. In this work, we delivered on all these four outstanding questions (Figure 6.1).

In Chapter 2, we developed a resource-aware framework to quantify the resource footprint of libraries of genetic constructs for mammalian cell expression. In our framework a transient-capacity monitor was used to quantify resource coupling in engineered mammalian cells, similar to previously published work[72]. By screening a library of promoters, Kozaks, polyA signals, and parts for multi-gene expression, we showed that combinations of these biological parts have different resource footprints. While previous research used transient capacity monitors to visualise resource competition in mammalian cells, a direct connection between the design of genetic constructs and their resource footprints was never established. More importantly, using our framework we individuated the location of the resource bottlenecks within the gene expression pathway. We demonstrated that while transcriptional resources often face shortages in engineered mammalian cells, translational resources undergo saturation only in very specific scenarios, which correspond to the ones with high intracellular heterologous mRNA concentration. This finding is in striking contrast with what has been previously reported in bacteria, where shortages in ribosome availability contribute majorly to the overall competition for gene expression resources. Our research also emphasises the importance of specific biological parts in shaping the resource footprint of genetic constructs and provides higher-order design rules for generating genetic constructs with improved gene

expression performance and minimal resource footprint. Although comparable improvements in the gene expression performance of genetic constructs can be achieved by changing each of the tested biological parts – promoter, Kozak, or polyA signal – the type of biological part changed results in very different resource footprint profiles. For instance, substituting a weak promoter with a strong promoter increased the resource footprint of the resulting design more than changing a low-performance Kozak or polyA signal to high-performance configurations. These considerations support the implementation of resource-informed genetic constructs design in mammalian cells as a further strategy to mitigate the shortcomings of resource competition in engineered mammalian cells. We applied resource-informed design to predictably generate genetic constructs with reduced resource footprint for bioproduction and biotherapy applications, including EBNA1-expressing constructs and a widely-used multi-gene switch for inducible expression. Informed design of genetic constructs has obvious advantages over other available mitigation strategies in mammalian cells, as it does not require extensive engineering and does not cause output suppression of the transiently-expressed genes. Indeed, by acting at the design stage of the engineering pipeline, co-expression systems can be optimised in a relatively simple way. This makes informed design the most suited mitigation strategy for all those applications where output maximisation of transfected genes is sought, such as for bioproduction applications.

Arguably, our strategy comes with limitations, the more important being context dependencies. We observed variability in the performance and resource footprints of equal parts when assembled with different components and when tested in different cell lines. These are well-reported effects in the synthetic biology literature and threaten informed design[167]. To understand to what extent context dependencies flawed our approach, we tested our framework to optimise EBNA1-expressing constructs and noticed that although the gene expression performance of the EBNA1 designs did not correlate with the performance of EGFP-expressing designs, the resource footprints of the selected designs in the two cases showed a good linear correlation. This indicates that informed design can at least be used to predictably generate genetic constructs with reduced resource footprint independently of the gene in use, which is still very valuable and impossible with current methods. Nevertheless, the differences in the behaviour of equal parts and constructs across cell lines imply that an initial characterisation must be undertaken when working with a new cell line. This is something we expected as cell lines likely differ in the composition and allocation of their resource pool.

In Chapter 3 we extended our considerations on informed design by analysing the robustness to resource load of a library of genetic constructs. While previous research analysed the robustness of a panel of constitutive promoters to the expression of transactivator domains[48], we characterised how biological parts essential to any mammalian gene expression system – promoters, Kozaks, and polyA signals – shape the robustness of genetic constructs. Arguably, our characterisation can inform the design of genetic constructs that are not restricted to specific applications. With our work we highlighted three key findings. First, the

robustness to resource load for a specific design does not depend on its gene expression performance. Second, TET-ON inducible genetic constructs display perfect robustness profiles, suggesting that these constructs are particularly efficient in hijacking cellular resources even in a condition of resource loading from a competing module. It would be interesting to understand whether this is a characteristic of inducible systems in general or it is only specific to the TET-ON system that we tested. Finally, polyA selection dramatically impacts the robustness to resource load of genetic constructs. Although we do not have an explanation for this behaviour, it is certainly very interesting and adds complexity to the overall contribution of polyA signals to the resource competition landscape in engineered mammalian cells. We believe our findings are important to any professional working with transfected systems in mammalian cells and not only to cell engineers and synthetic biologists. To support this, we showed the biases introduced in experimental results when using a low-robust genetic construct as a transfection control for transient transfections experiments. We hope this will raise awareness of the unintended effects of resource competition within the life science community.

In Chapter 4, we developed an integrated capacity monitor in mammalian cells to report on the impact of transfected genetic constructs on mammalian cellular physiology. To identify an optimal integrated capacity monitor design, we had to go through four capacity monitor iterations to increase the dynamics of the reporting system. In doing so, we highlighted inducibility and increased protein turnover as two crucial parameters to improve the dynamics of biosensing systems integrated into the mammalian genome. We used our final capacity monitor design to measure the effects on cellular physiology of a library of genetic constructs, proving that it functioned as a proxy of cellular growth. With these findings, we considerably advanced our current understanding of resource competition in mammalian cells. First, we proved that single-copy integrated genes are affected by resource competition initiated by transfected plasmids. While initial evidence that endogenous genes could face coupling effects when in the presence of a resource-demanding plasmid was reported by Frei *et al*[72], this was only demonstrated by qPCR experiments. Here we went a step further and proved that resource competition decreases the protein concentration of a single-copy integrated gene and has, therefore, the potential to alter cellular functions. More importantly, we showed initial evidence that the capacity monitor fluorescence could inform on mammalian cellular growth profiles, making it the first tool of its kind. Unfortunately, we obtained the final capacity monitor design at a late stage of the PhD and managed to do only limited testing. With more experiments, this tool could represent a great asset for the high-throughput screening of genetic designs for the bioproduction industry.

A key limitation of our integrated capacity monitor comes from its improved dynamics. In other words, to observe any change in the capacity monitor protein levels, we had to combine inducible expression and increased protein turnover. It could be argued that these modifications make the capacity monitor very different from endogenous genes and that the latter could not experience a similar decrease in protein levels in a condition of resource shortage. This would also be in accordance with the lack of response we

observed with our first capacity monitor design, which as a single-copy constitutive gene with normal protein turnover, would more closely resemble endogenous genes. While this is a valid point, previous evidence showed suppression of transcript levels from housekeeping endogenous genes in resource-depleted mammalian cells[72], suggesting that a change in the endogenous proteome will likely happen within a longer timeline.

In Chapter 5 we set out to characterise the competition for resources within the protein degradation and secretion pathway. First, we proved for the first time that protein degradation resources are limited in mammalian cells. This is in accordance with previous work in bacteria showing similar findings[69] and introduces complexity to the resource competition landscape in mammalian cells. By testing a library of degradation domains against a panel of actively degraded capacity monitors, we concluded that bottlenecks likely involve resources shared by every degradation domain, such as ubiquitin or proteasomal proteins. Despite this, in one case (EGFP-PEST vs mKATE-PEST) we observed a higher degree of coupling, possibly indicating the presence of additional bottlenecks involving the saturation of domain-specific resources. We also proved that competition for degradation resources can arise between integrated and transient systems, expanding the implications of our findings for bioproduction and biotherapy. For instance, overloading the protein degradation pathway could be an effective strategy to reduce unwanted degradation of therapeutic proteins in bioproduction applications or to stabilise those cellular proteins whose dysregulated degradation causes human disease. Our findings show that to understand intracellular protein levels, the combined effect of competition for gene expression and protein degradation resources must be considered. Finally, we provided the first proof of competition for secretion resources in engineered mammalian cells and highlighted the ER as the step undergoing resource limitation within the conventional secretory pathway.

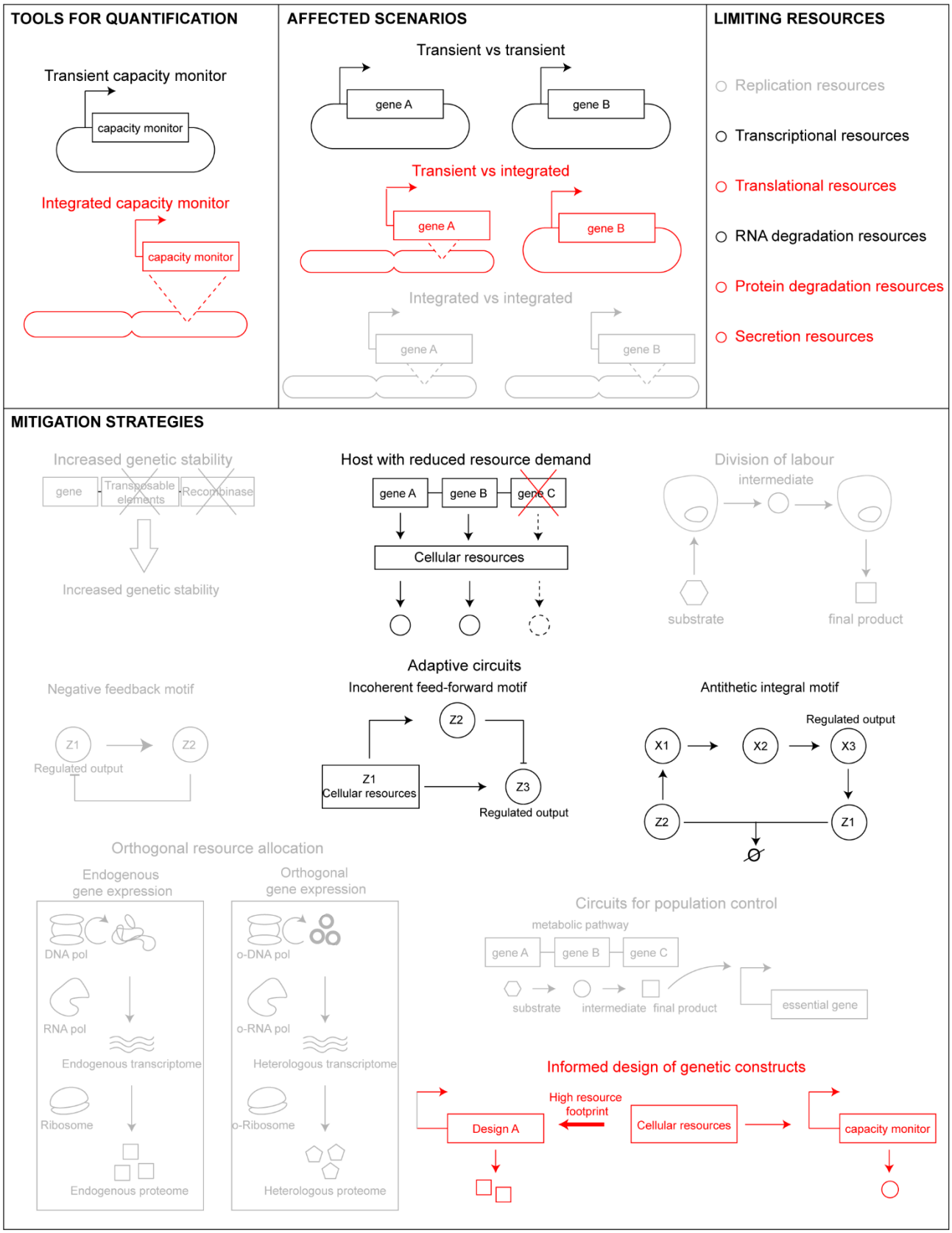

Figure 6.1 Our contribution to resource competition research in mammalian cells. Published evidence of resource competition in mammalian cells before this work is displayed in black in the figure. Our contributions outlined in this thesis are displayed in red. Awaited advancements are displayed in grey.

6.3. Future work and open challenges

Although we contributed to the current understanding of resource competition in mammalian cells, further research is needed to address unanswered questions and add depth and relevance to our findings.

First, with our work, we uncovered context dependencies that limit the translatability of our results to other mammalian cell lines. Although this was expected, as different cell lines likely have different resource pools, a comprehensive characterisation of resource allocation profiles in a wider panel of mammalian cell lines is needed to expand the applicability of our findings. Recent development in AI-informed high-throughput approaches can be extremely useful in providing a framework to understand this complexity. Similar characterisations could also inform which cell types are most robust to resource overloading in certain cellular pathways, enabling the selection of cell types based on their resource pool for specific applications.

Second, with our work, we contributed to emerging evidence showing limitations in resource pathways other than transcription and translation. Including the work presented here, the available literature now reports evidence of resource limitations in RNA degradation, protein degradation, and secretion pathways[66,71]. Resource shortages may arise elsewhere. For instance, nuclear import/export, folding and post-translational resources could all reasonably face limitations in engineered cells. These types of resources have the potential to alter intracellular protein levels and, therefore, to alter cellular functions. A thorough understanding of the cellular pathways facing competition is awaited.

Third, although we provided initial information on the location of resource bottlenecks, further work is needed to pinpoint the specific cellular resources undergoing limitations. Achieving this milestone would have an enormous impact on our ability to deal with resource competition. It would refine resource-aware design considerations while supporting the development of new mitigation strategies. For instance, cell lines engineered to overexpress limiting resources could completely prevent resource shortages in the first place. It could also open to considerations on orthogonal resource allocation to avoid overloading of key cellular resources.

Fourth, further work should be directed at understanding whether competition effects can arise between integrated genes. This would be particularly relevant as genomic integration is one of the most sought-after methods to deliver genetic payloads for bioproduction and biotherapy applications. Complementary to this, a more thorough understanding of resource competition's effects on endogenous genes would benefit the community. While in this work we proved that integrated capacity monitors are affected by resource competition from a transient competing module, the effect of resource competition on endogenous genes might depend on the resources required for their expression and their location in the genome. Transcriptomics and proteomic approaches can help the identification of endogenous genes that are more affected or more robust to resource competition.

Finally, all the points above will be crucial to develop a unified, integrated framework to fully understand resource competition in engineered mammalian cells, supporting the generation of more accurate mathematical models, refined mitigation strategies, and renewed awareness of potential experimental bias. This will pave the way to the development of novel resource-aware technologies that will benefit bioproduction, biotherapy, and – more in general - cell engineering.

Chapter 7 - Materials and methods

7.1. Bacterial manipulation

7.1.1. Bacterial culture

E. coli DH5a (*F– φ80lacZΔM15 Δ(lacZYA-argF) U169 recA1 endA1 hsdR17(rK–, mK+) phoA supE44 λ–thi-1 gyrA96 relA1*) were used for all routine cloning. A glycerol stock was made at the beginning of the project, and competent cells were prepared starting from it.

Bacterial growth was performed in LB Broth (Formedium, LMM0102) at 37°C in shaking conditions. The media was supplemented with appropriate antibiotics, except for competent cell preparation (section 7.1.2). For all experiments, either Ampicillin at 100ug/mL (Sigma Aldrich, A0166-25G) or Kanamycin at 50ug/mL (Sigma Aldrich, B5264-1G) were used.

7.1.2. Competent cells preparation and transformation

Competent DH5a were prepared following the protocol developed by Chung *et al*[189]. Briefly, a bacterial overnight culture was diluted 1:100 in fresh antibiotic-free LB media and grown to OD 0.3-0.4. At this point, the culture was incubated for 10 minutes in ice, centrifuged at 4°C (10 minutes, 3000rpm), and resuspended in ice-cold transformation and storage solution (TSS), which was previously prepared (Table 8.1). The cell suspension was then aliquoted and stored at -70°C.

Table 8.1 TSS preparation

Component	Volume/Weight
PEG 8000	5g
$MgCl_2$ 1M	1.5uL
DMSO (100%)	2.5uL
LB media	to 50mL
Total	**50mL**

For transformations, 50uL of competent cells were thawed in ice, mixed with DNA, and incubated in ice for 20 minutes. The cell suspension was then heat-shocked at 42°C for 30 seconds and moved back in ice for 2 minutes. After that, 900uL of antibiotic-free LB was added, followed by a phase of recovery growth at 37°C in shaking conditions. The transformed cells were centrifuged (3 minutes, 3500rpm) to remove excess media and plated on LB agar plates (Formedium, LMM0202) with appropriate antibiotics.

7.2. Molecular biology work

7.2.1. Cloning

7.2.1.1. PCR

PCR reactions were performed using Phusion® High-Fidelity DNA Polymerase (NEB, M0530L) kit with the following protocol (Table 8.2).

Table 8.2 Protocol for PCR reactions

Component	Volume
MgCl2 (50mM)	0.5uL
DMSO (100%)	1.5uL
dNTPs (10mM)	1uL
GC buffer (5X)	10uL
DNA template (1-10ng/uL)	1uL
Primer fw (10uM)	2.5uL
Primer rv (10uM)	2.5uL
Phusion polymerase (2000 units/mL)	0.5uL
Nuclease-free H_2O	30.5uL
Total	**50uL**

Primer annealing temperatures were calculated using the NEB Tm calculator online tool version 1.16.5. The cycling conditions were set according to the manufacturer's protocol. DNA sequences of the parts used in this study can be found in section 8.3.

Completed PCR reactions were mixed with loading dye (NEB, B7024S) and loaded in 1% agarose gels (CarlRoth, 2267.5) stained with SybrSafe (ThermoFisher, S33102). Gel extractions were performed using a commercial kit (Qiagen, 28706).

7.2.1.2. Restriction digest

Digestion reactions were performed using the following protocol (Table 8.3). Reaction buffer, enzymes, reaction temperature and incubation time, and inactivation temperature and incubation time have been determined using the NEBcloner online tool version 1.13.8.

Table 8.3 Protocol for digestion reactions

Component	Volume/weight
DNA	1ug
Reaction buffer (10X)	1.5uL
Enzyme #1	1uL
Enzyme #2	1uL
Nuclease-free H_2O	to 50uL
Total	**50uL**

7.2.1.3. Ligation

Ligation reactions were performed using T4 ligase (NEB, M0202L) according to the following protocol (Table 8.4).

Table 8.4 Protocol for ligation reactions

Component	Volume/moles
T4 DNA ligase buffer (10X)	2uL
Backbone #1	3x pmol
Insert #2	x pmol
T4 DNA ligase (400000units/mL)	1uL
Nuclease-free H_2O	to 25uL
Total	**25uL**

Backbone and insert(s) were added in a 3:1 molar ratio. Sticky ends ligation reactions were incubated at room temperature for at least 30 minutes. Blunt end ligation reactions were incubated at room temperature for 2 hours.

7.2.1.4. Golden Gate reactions

For Golden Gate reactions BsaI-HF®v2 (NEB, R3733L), Esp3I (NEB, R0734L), BbsI-HF® (NEB, R3539L), or SapI (NEB, R0569L) were used according to the following protocol (Table 8.5).

Table 8.5 Protocol for PCR reactions

Component	Volume/moles
Backbone	13fmol
Insert #1	13fmol
Insert #2	13fmol
Insert #n	13fmol
T4 DNA ligase buffer (10X)	2uL
T4 DNA ligase (400000units/mL)	0.5uL
Enzyme (BsaI, Esp3I, BbSI, or SapI)	0.5uL
Nuclease-free H_2O	to 20uL
Total	**20uL**

The reaction was performed in a thermocycler, using cycling conditions outlined in Martella *et al*[190].

7.2.2. Western blot

Transfected HEK293T were detached, centrifuged, and washed in ice-cold DPBS. Cell pellets were lysed in RIPA buffer (Sigma Aldrich, 89901) with 1% v/v protease inhibitor cocktail (Sigma Aldrich). A BCA assay was used to quantify the protein concentration of the lysates. The lysates were denatured, and 20µg were loaded into SDS PAGE gels. Proteins were transferred onto nitrocellulose membranes and probed for EBNA1 (Merck, clone 1H4, monoclonal, cat MABF2800, dilution 1:1000) and a loading control, Vinculin (Sigma, clone hVIN-1, monoclonal, cat V9264, dilution 1:5000). Secondary antibodies (Life Technologies, Goat anti-Rat IgG A18868, polyclonal, and Abcam Rabbit Anti-Mouse IgG H&L (Alkaline Phosphatase) ab6729, dilutions 1:10000) were detected using the Novex™ AP Chromogenic Substrate (Invitrogen). Image J was used for densitometry analysis. To obtain a normalisation factor specific to each lane, the Vinculin signal for each lane was normalised against the lane with the strongest signal. This was then multiplied by the EBNA1 signal of the same lane to get a normalised EBNA1 experimental signal. Finally, to control for

inter-blot variability, normalised EBNA1 values were divided once more against the average normalised signal of each lane present within the blot (except the negative control).

7.2.3. *In vitro* mRNA production

mRNAs were transcribed, capped and polyadenylated in vitro using the HiScribe® T7 ARCA mRNA Kit (NEB, E2060S) following the manufacturer's instructions. RNAs were then purified from the in vitro reaction using RNA Clean & Concentrator-25 (Zymo Research, R1017) as per the manufacturer's instructions.

7.2.4. Design of 5'UTR structures

5'UTR structures were designed using RNAfold (http://rna.tbi.univie.ac.at/cgi-bin/RNAWebSuite/RNAfold.cgi) and NUPACK (http://www.nupack.org/partition/new). Both online tools predict the minimum free energy (MFE), the frequency of the MFE structure in the ensemble and the melting temperature of each structure.

7.3. Mammalian cells manipulation

7.3.1. Cell culture

HEK293T were purchased at passage 16 (ATCC-CRL-3216) and tested for mycoplasma multiple times throughout the PhD. CHO-K1 were kindly given to us by Prof. Cleo Kontoravdi at passage 2 and were never tested for mycoplasma. The HEK293T AAVS1-TETp-BFP cell line was kindly gifted to us by Fowler's lab[152] and has never been tested for mycoplasma.

HEK293T were cultured in DMEM Glutamax (ThermoFisher 31966047) supplemented with 10% FBS (Life Technologies, 16000-044). CHO-K1 were cultured in MEMα (Sigma Aldrich, M4526) supplemented with 10% FBS, 1% non-essential amino acids (ThermoFisher, 11140050), and 1% L-glutamine (ThermoFisher, 25030081). HEK293T and CHO-K1 were split every three to four days and cultured in a humidified incubator set at 37°C and 5% CO2.

For inducible experiments doxycycline hyclate (Sigma Aldrich, D9891-1G) was used at a concentration of 1ng/uL for the induction of transiently-transfected plasmids and at 2ng/uL for the induction of integrated cell lines unless otherwise specified.

7.3.2. DNA transfections

Transfections were performed in a 24-well plate format. Cells were seeded approximately 24h before transfection in complete media. HEK293T were transfected using xTREME GENE HP DNA (Merck, 6366236001) or FugeneHD (Promega, E2311) as transfection reagent, depending on the experiment. CHO-K1 were transfected using TransIT-X2 (Mirus, MIR 6003). In all cases, the transfected plasmids were

normalised by molar ratio. Two days after transfection, cells were detached and prepared for flow cytometry analysis.

For transfection using xTREME GENE HP DNA, 10^5 HEK293T cells were seeded in 24-well plates one day before transfection. The day after, the transfection mix was prepared using 0.08-0.12pmol of total DNA (0.04 each plasmid) – depending on the experiment - and xTREME GENE HP at a ratio of 0.72uL per 0.08pmol diluted in a final volume of 50uL in Optimem Serum Free media (ThermoFisher, 31985062). The transfection mix was incubated for 30 minutes at room temperature and was then added to the cells.

For transfection using FugeneHD, 1-2*10^5 HEK293T cells were seeded in 24-well plates one day before transfection. The day after, the transfection mix was prepared using 0.08-0.16pmol of total DNA – depending on the experiment - and FugeneHD at a ratio of 0.72uL per 0.08pmol diluted in a final volume of 25uL in Optimem Serum Free media (ThermoFisher, 31985062). The transfection mix was incubated for 7 minutes at room temperature and added to the cells.

For transfection using TransIT-X2, 8*10^4 CHO-K1 cells were seeded in 24-well plates one day before transfection. The day after, the transfection mix was prepared using 0.08-0.12pmol of total DNA (0.04 each plasmid) – depending on the experiment - and TransIT-X2 at a ratio of 0.968uL per 0.08pmol diluted in a final volume of 50uL in Optimem Serum Free media (ThermoFisher, 31985062). The transfection mix was incubated for 25 minutes at room temperature and added to the cells.

7.3.3. RNA transfections

For RNA transfection, 10^5 HEK293T cells were seeded in 24-well plates one day before transfection. The day after, the transfection mix was prepared using 0.5ug of total purified RNA (0.25ug each plasmid), 1uL of TransIT®-mRNA transfection reagent, and 1uL of TransIT®-mRNA transfection boost (Mirus, MIR 2225) in a final volume of 50uL in Optimem Serum Free media. The transfection mix was incubated for 5 minutes at room temperature and added to the cells.

7.3.4. Fluorescence measurements

7.3.4.1. Flow cytometry and sorting

For flow cytometry analysis, cells were detached, centrifuged for 5 minutes at 3000rpm, washed in DPBS (ThermoFisher, 14190250) with 10% FBS, centrifuged again, and resuspended in a final volume of 500uL of DPBS. Cells were then filtered to disrupt any cell clumps and analysed at either the Attune NxT flow cytometer (ThermoFisher) – for experiments in Chapters 2, 3, and 4 – or the MACS Quant VYB flow cytometer (Miltenyi Biotec) – for experiments in Chapter 5. For experiments performed at the Attune NxT, EGFP and mNeonGreen were measured using a 488nm laser and a 530/30 bandpass filter; mKATE and mCherry were measured using a 561nm laser and a 620/15 bandpass filter; mTagBFP2 was measured using a 405nm laser and a 450/40 bandpass filter. For experiments performed at the MACS Quant VYB, EGFP was

measured using a 488nm laser and a 525/50 bandpass filter; mKATE was measured using a 561nm laser and a 615/20 bandpass filter. In all cases, 10000-50000 viable cells were recorded for each sample.

For cell sorting experiments, cells were detached, centrifuged for 5 minutes at 3000rpm, washed in DPBS, centrifuged again, and resuspended in DPBS at a final concentration of 10^7 cells/mL. Cells were kept in ice and filtered before sorting to disrupt cell clumps. Sorting was carried out a 4°C using a BD FACSAria Fusion (BD Biosciences) by specialised facility staff. Purity checks on the sorted population were carried out whenever possible.

7.3.4.2. Microplate reader

For secreted fluorescence experiments, cells were transfected in DMEM FluoroBrite (ThermoFisher, A1896701) to minimise the background noise of fluorescent measurements. One day after transfection, dox (1ng/uL) was added to the media to induce expression of the capacity monitor. The day after induction, the spent media was collected, centrifuged for 10 minutes at 5000rpm, and 100uL was transferred on a black flat-bottom 48-well plate. To record secreted fluorescence, a Tecan Spark (Tecan) or a TECAN Infinite® M Nano (Tecan) were used. This is because we faced several malfunctions of the Tecan Spark throughout the project; therefore – although not ideal – another microplate reader had to be used. Secreted EGFP was measured with excitation of 485/9, emission of 535/20, and gain set to optimal; secreted mCherry was measured with excitation of 561/9, emission of 630/20, and gain set to optimal.

7.3.5. Stable cell line generation

7.3.5.1. CRISPR-based integration

This study's CRISPR integrations involved integrative payloads expressing a fluorescence gene. This facilitated the selection of the integrated population and allowed easy assessment of the stability of the selected clones. All the integrative payloads in this study have been integrated into AAVS1 in HEK293T, using the following gRNA sequence, which was already experimentally validated[191]: GGGGCCACTAGGGACAGGAT.

For CRISPR-based integrations, 2*10^5 HEK293T cells were seeded in 24-well plates one day before transfection. The day after, 0.06pmol of donor backbone and 0.04pmol of All-in-One backbone were co-transfected in HEK293T using FugeneHD as transfection reagent at a ratio of 3:1 - FugeneHD (uL):transfected DNA (ug). One day after transfection, the spent media was changed with fresh culture media and cells were cultured for at least ten days. Cells were then detached, and the positive fluorescent population was single-cell sorted into individual wells of a 96-well plate. The resulting single-cell clones were cultured for three-to-four weeks in conditioned medium (i.e. filtered spent media collected from a confluent flask) at the ratio of 1:3 to fresh medium with 25% FBS and 1% penicillin-streptomycin (Sigma Aldrich, P4333). Confluent wells were then assessed with the fluorescence microscopes, and the ones displaying homogeneous fluorescence from the integrations were expanded. As soon as they reached

confluency in a 24-well plate format, the genomic DNA of the expanded clones was extracted to perform PCR genotyping. Correct clones were expanded further, aliquoted, and frozen in culture media with 10% DMSO.

7.3.5.2. Landing pad-mediated recombination

For landing pad-mediated recombination, 1.5*10^5 landing pad cell lines (HEK293T) were seeded in 24-well plates one day before transfection. The day after, 0.08pmol of integrative backbone and 0.08pmol of integrase-expressing backbone were co-transfected using 3.6uL of FugeneHD as transfection reagent. One day after transfection, the spent media was changed with fresh culture media. Transfected cells were gradually expanded to a T75 flask, and the cellular population with correct recombination events was sorted.

7.4. Data analysis

7.4.1. Flow cytometry data analysis

Flow cytometry data were initially analysed using Flowjo. Single cells were selected based on FSC/SSC parameters, and a compensation matrix was applied for multi-fluorescence experiments. Gating of transfected cells was performed in three different ways across chapters using either Flowjo or custom Python scripts. In Chapter 2, we used an automatic OR gate on the fluorescent markers implemented by Simone Furini. In Chapter 3, we used an automatic gate on the EGFP marker implemented by Simone Furini. In Chapters 4 and 5, we set up manual OR gates on the fluorescent markers using Flowjo. The different gating strategies were justified by the type of experiments and the collaborators involved. Please refer to section 8.2 for an in-depth discussion of each gating strategy and justification for their use. After the gating was applied, the geometric mean of the area measurement of each fluorescent marker was calculated and expressed as relative fluorescent units (RFU). The data were then exported as an Excel file for further analysis.

7.4.2. Further data analysis and plotting

All data were imported in Excel or R for further analysis. All data points lower than $Q1 - 1.5 * IQR$ and higher than $Q3 + 1.5 * IQR$ - where IQR stands for the interquartile range - were considered outliers and excluded from further analysis. For most cases, the arithmetic mean and standard deviation of the raw data points have been plotted.

When the variability was high or when pooling data acquired with different machines, normalised values were reported instead. These have been calculated by dividing the raw data value of each data point by the averaged data values of two repeats of a control sample run on the same day. For normalised values, standard deviation has not been calculated, but individual data points have been plotted in all cases.

7.5. Statistics and reproducibility

Each experiment was repeated at least twice on different days to assess variability, except in a few instances. Statistical significance was calculated using a two-sided Mann-Whitney test.

Chapter 8 - Supplementary information

8.1. Supplementary figures

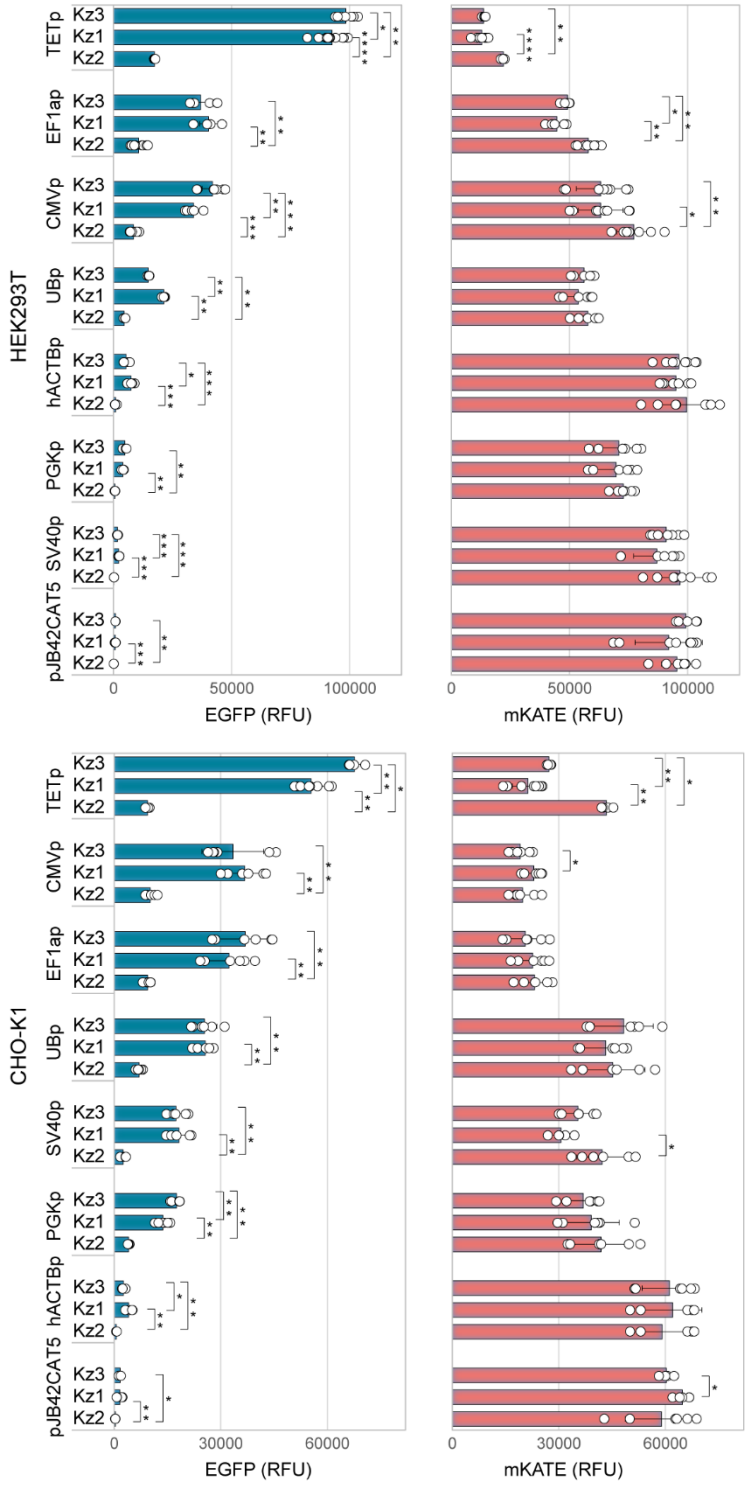

Figure 8.1 Resource competition imposed by different Kozak sequences. Resource footprint of different promoter-Kozak configurations on the expression of the capacity monitor in HEK293T (top panels) and CHO-K1 (bottom panels). Test plasmid and capacity monitor expression levels are reported as mean RFU ± std. Number of independent experiments ≥ 2. The exact number of repeats for each sample is reported in Di Blasi et al.[151]. Two-sided Mann-

Whitney test P value: ****<0.0001, ***<0.0005, **<0.005, *<0.05. Data analysis is described in Materials and methods (Chapter 7).

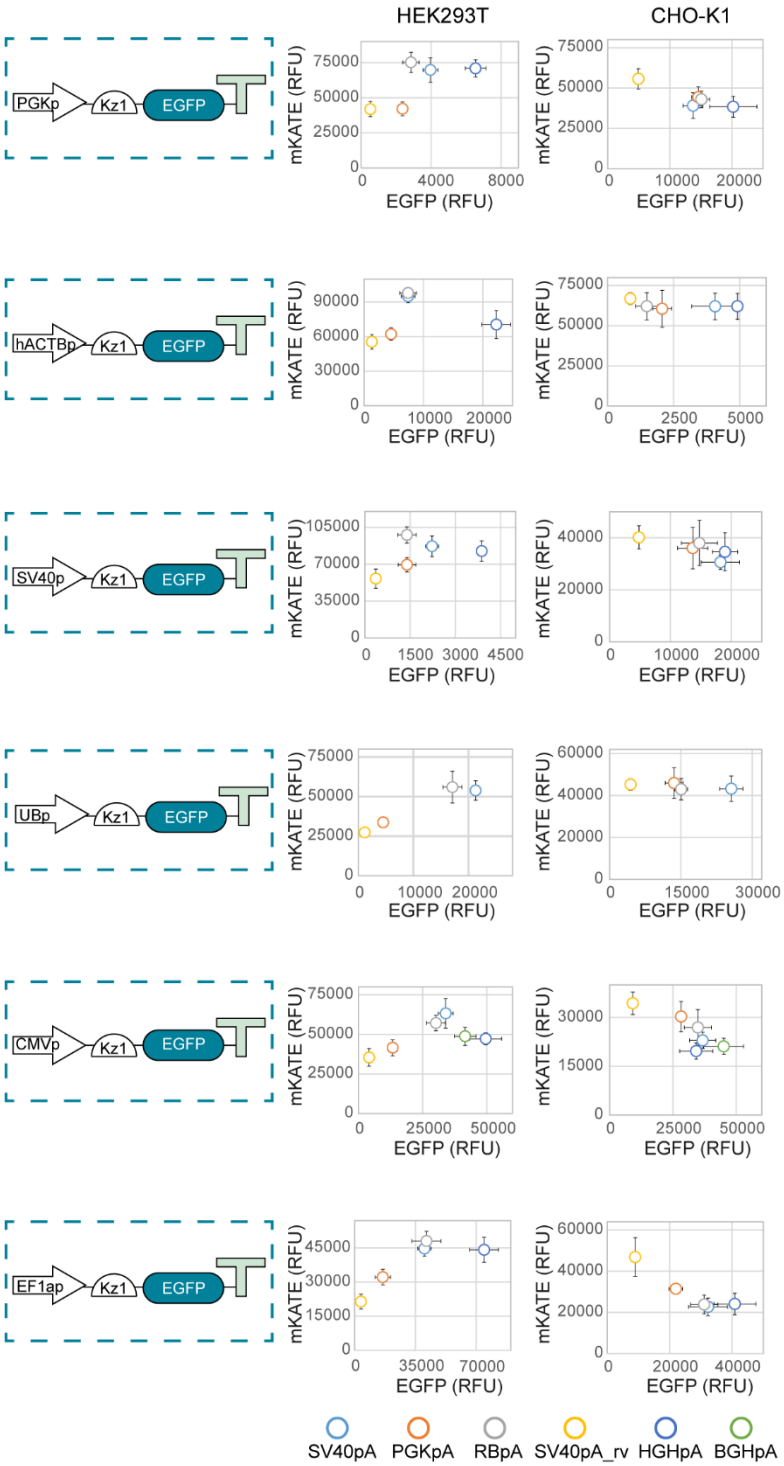

Figure 8.2 Resource footprint imposed by different promoter-polyA combinations. Impact of different promoter-polyA configurations on the capacity monitor expression in HEK293T and CHO-K1. Test plasmid and capacity monitor expression levels are reported as mean RFU ± std. Number of independent experiments ≥ 2. The exact number of repeats for each sample is reported in Di Blasi et al.[151]. Data analysis is described in Materials and methods (Chapter 7).

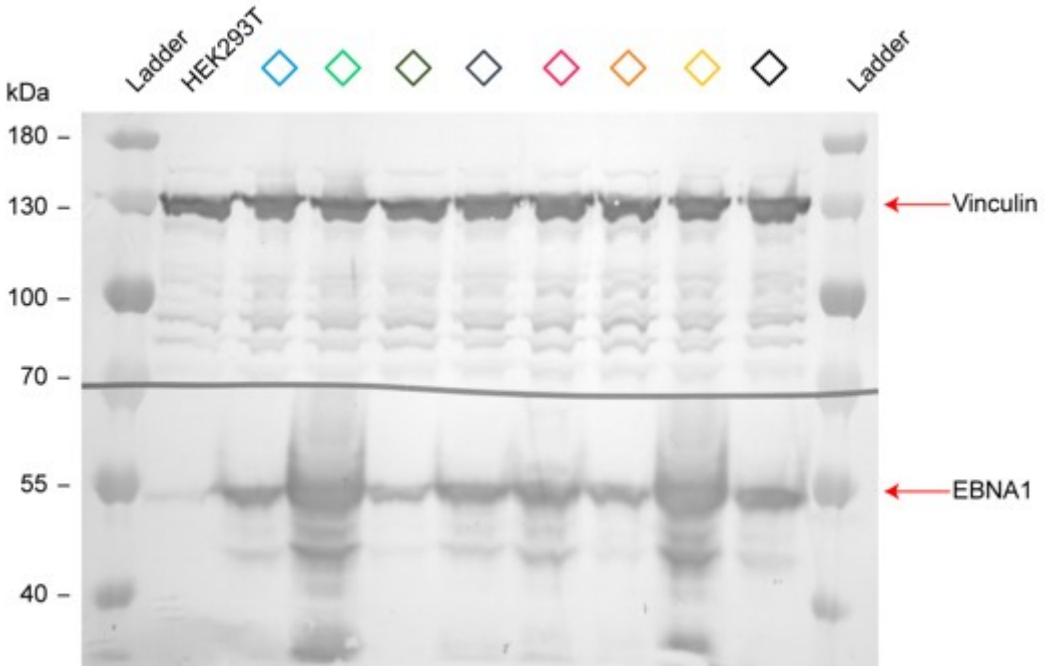

Figure 8.3 Expression output of EBNA1-expressing constructs in HEK293T. EBNA1 (test plasmid) and Vinculin (internal control) expression for the selected EBNA1-expressing designs in Figure 2.14 detected by Western blot. Number of independent experiments ≥ 2. The exact number of repeats for each sample is reported in Di Blasi et al.[151]. Data analysis is described in Materials and methods (Chapter 7)

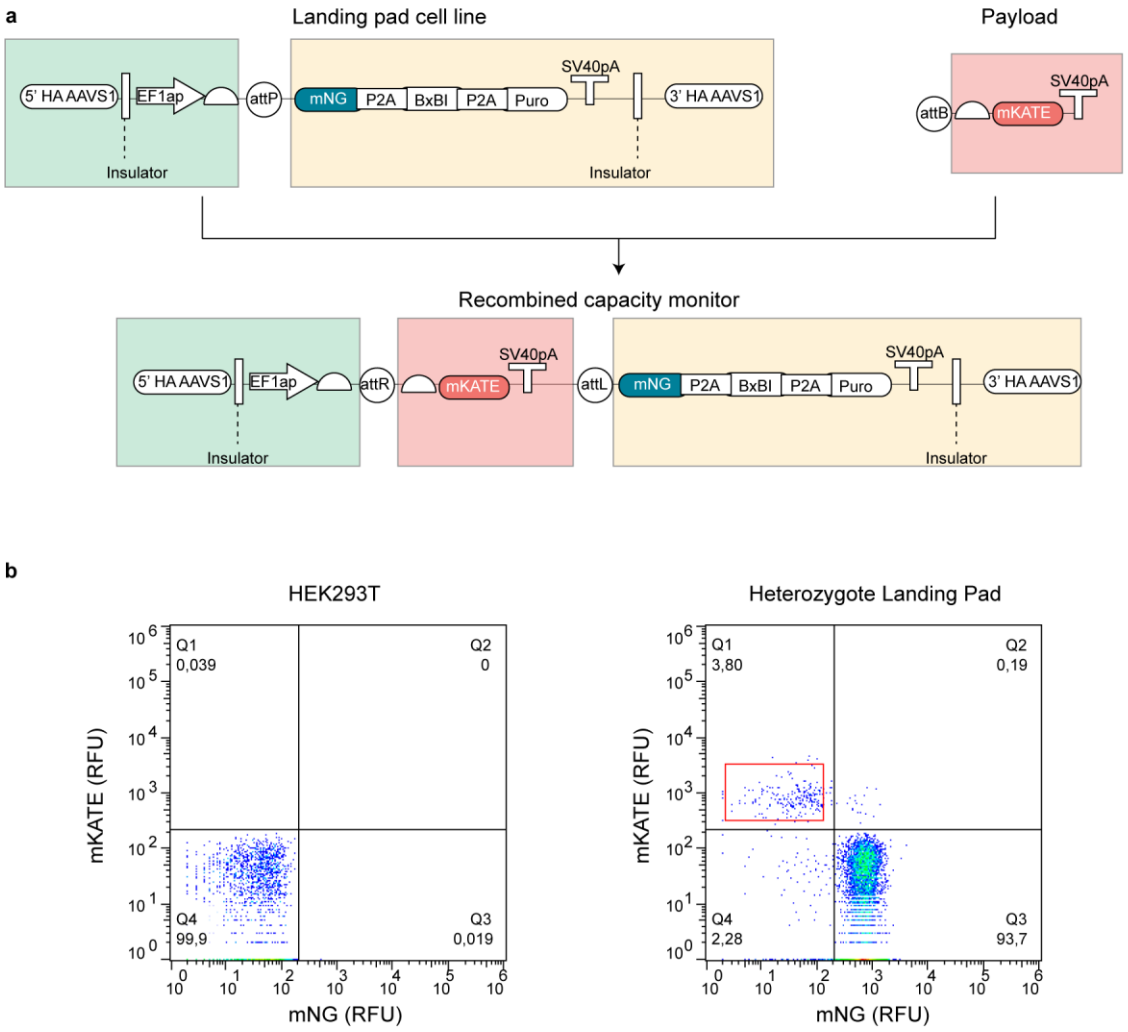

Figure 8.4 Landing pad recombination of a simple payload. We transfected the landing pad cell line with a payload for recombination and identified correct recombination events (red square). No hybrid population in Q2 (i.e. mKATE+/mNG+) was detected in this case.

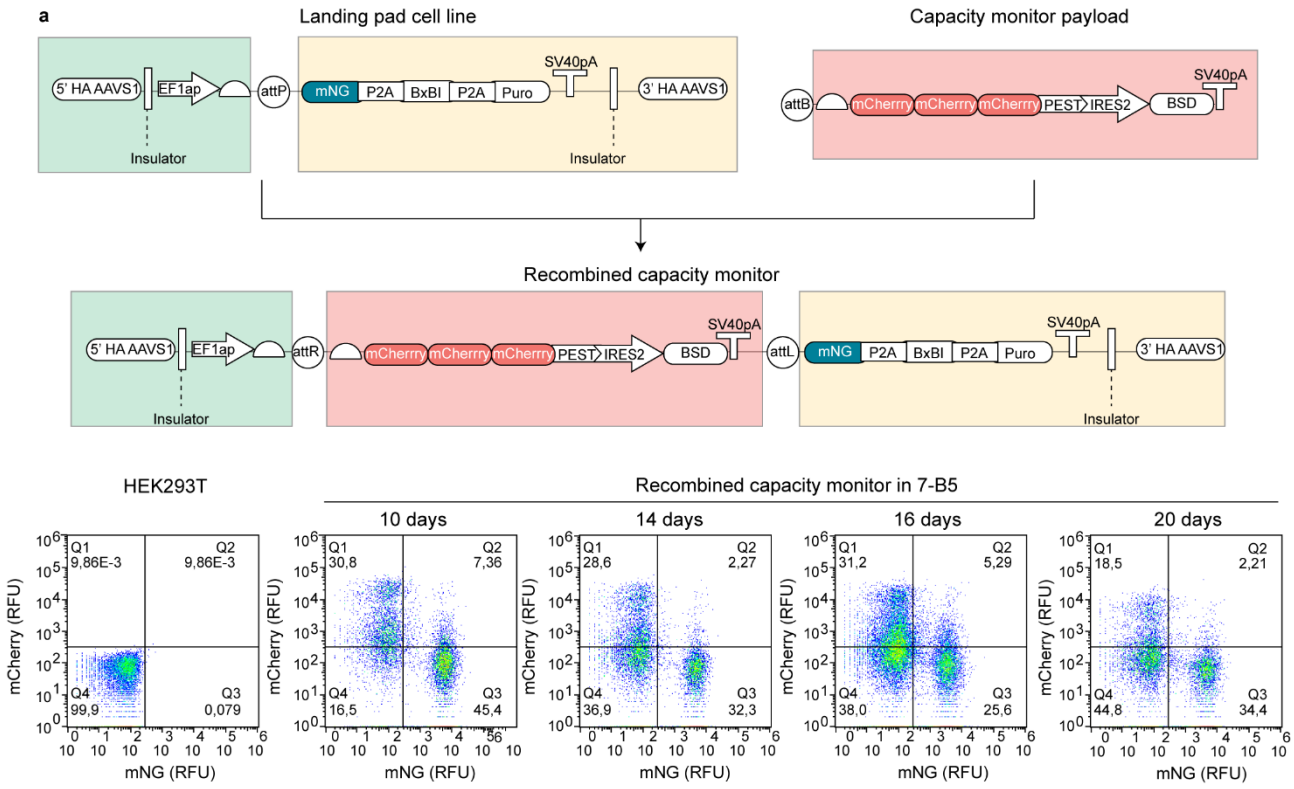

Figure 8.5 Landing pad post-recombination instability. We transfected the landing pad cell line with a payload for recombination and selected correct recombination events by adding blasticidin in the culture media. Even in the presence of positive antibiotic selection, the cellular population in Q1 – which corresponds to cells with correct recombination – decreased in percentage over time, suggesting silencing of the expression of the recombined landing pad.

8.2. Flow cytometry data analysis

Single cells were manually gated in all cases, and a compensation matrix was applied to multi-colour data using Flowjo. At this point, we realised the importance of the fluorescence gating strategy to analyse flow cytometry data for our experiments. Since most of our experiments involved the comparison of libraries of constructs of different expression levels, we had to be sure not to introduce experimental artefacts by improper gating. The most common gating strategy in multi-fluorescence experiments is to set up a double-positive gate, only selecting the positive cells for both fluorescence markers (EGFP and mKATE). This is a very stringent gating strategy and, when applied to our experiments, introduces overestimation of the capacity monitor output (mKATE) when co-transfected with low EGFP-expressing plasmids (Figure 8.5). Alternative gating strategies, such as OR gates, have been used in published work to avoid similar artefacts[48] (Figure 8.5)

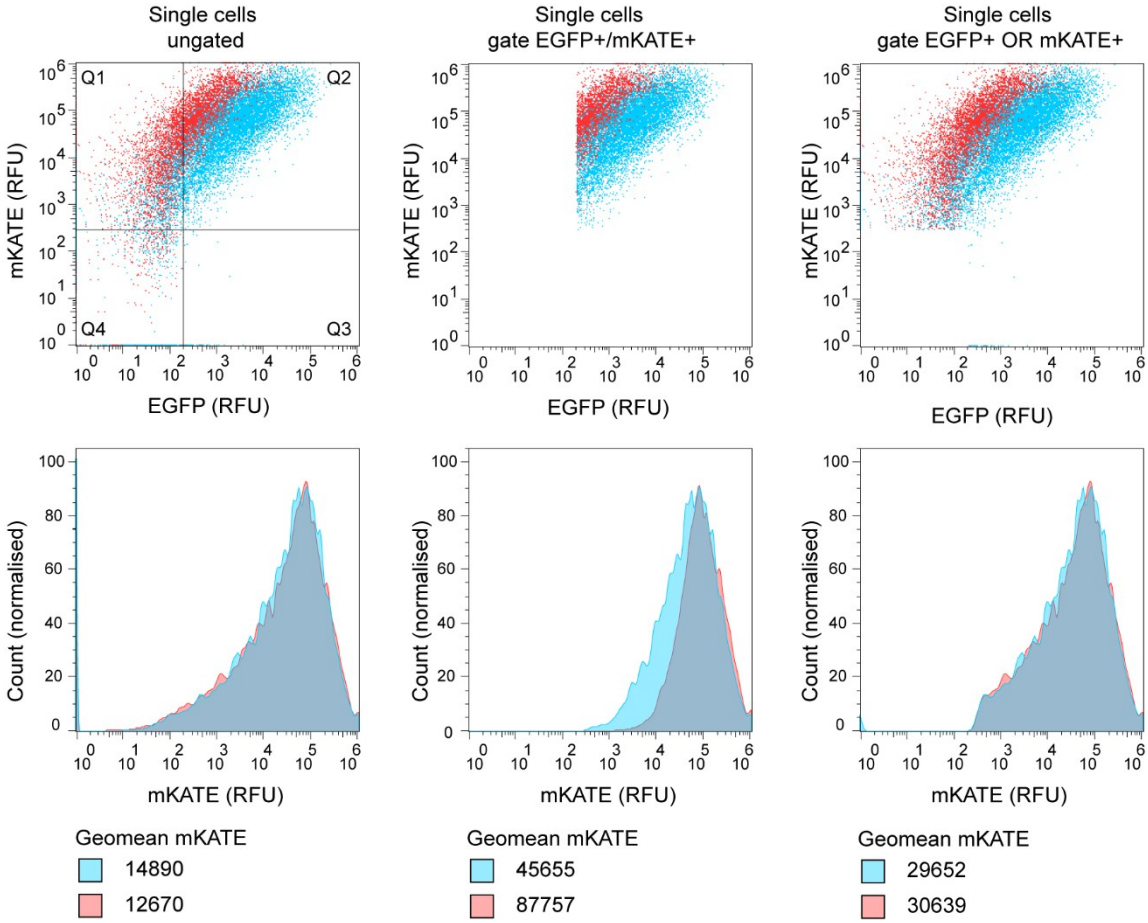

Figure 8.5 Comparisons of two gating strategies. The capacity monitor (mKATE) was co-transfected either with a low EGFP-expressing plasmid (red) or a high EGFP-expressing plasmid (blue) in HEK293T. Overlaid fluorescence dot plots (top panels) and histograms (bottom panel) of the single cell population of the two co-transfected conditions. When comparing the ungated single cell population (left panels) with either a double positive gate (EGFP+/mKATE+, middle panels) or an OR gate on the two fluorescence markers (EGFP+ OR mKATE+, right panels), the double positive gate gives an incorrect estimation of the mKATE fluorescence, as evidenced by the reported geomean values. While the red and blue samples have similar geomean mKATE levels in the ungated and OR-gated populations, this is not the case when a double positive gate is applied. This indicates that double positive gates are unsuitable for our dataset and should not be used to avoid experimental artefacts.

For the data in Chapter 2, Simone Furini performed an automatic OR gate on the fluorescent markers (EGFP and mKATE) using a custom-made Python script he developed. For the data in Chapter 3, Simone Furini performed an automatic gate on the EGFP marker using a custom-made Python script he developed. For the data in Chapter 4 and 5, I performed a manual OR gate on the fluorescent markers (EGFP and mKATE). The three methods and the reasons for their choosing will be discussed below.

8.2.1.1. Automatic OR gate

The data in Chapter 2 have been analysed using an automatic OR gate on the two fluorescent markers. We chose to use an OR gate as we thought it was the method introducing less bias in our results (Figure 8.5).

Compared to standard analysis methods using OR gates with manual thresholds, the procedure implemented here has two advantages. Firstly, the thresholds between cells expressing or not expressing each marker are automatically defined, reducing the risk of a possible bias introduced by the operator. Secondly, in cases where a blunt separation between expressing/not-expressing cells does not exist, the results of a manual OR-gate depend on intrinsically ill-defined thresholds. The automatic procedure adopted with this type of gating better reflects the uncertain classification of cells in these transition regions. For more information on this gating strategy, please refer to our published work[151].

8.2.1.2. Manual OR gate

The data in Chapters 4 and 5 have been analysed using a manual OR gate on the two fluorescent markers. The reasons for using manual OR gates over automatic ones are twofold. First, these data were acquired over many years and before our collaboration with Simone Furini, meaning that a consistent chunk of data was analysed using manual gating in Flowjo before the automatic gating strategy was developed. Second, we noticed that the automatic gating strategy had problems with samples with ill-defined fluorescence populations or where more than two populations were present. This represented a problem, particularly for the data in Chapters 4 and 5, where epigenetic silencing and the addition of degradation domains resulted in fluorescence distributions that were much less defined than the ones in Chapter 2. For these reasons, we used manual OR gates to analyse the data in these chapters. Nevertheless, we confirmed that for samples with well-defined fluorescent populations, there is almost no difference in the results generated with the two gating strategies (Figure 8.6).

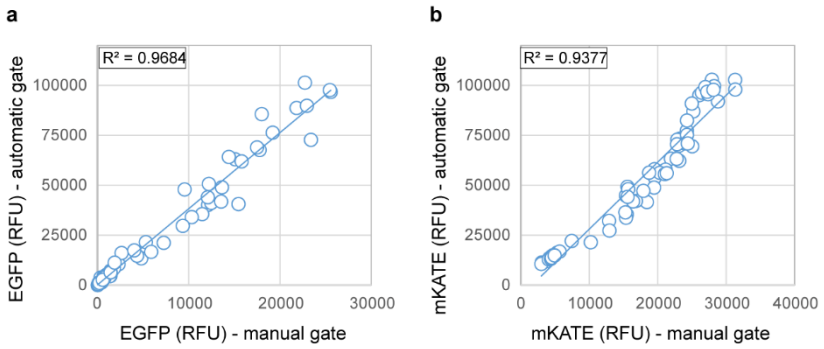

Figure 8.6 Comparison of manual and automatic OR gates. (a) A subset of the samples presented in Chapter 2 was analysed by applying an automatic or a manual OR gate. Test plasmid (EGFP) expression levels for the two gating strategies are reported as RFU (Pearson coefficient 0.9841). **(b)** Capacity monitor (mKATE) expression levels for the two gating strategies are reported as RFU (Pearson coefficient 0.9684).

8.2.1.3. Automatic EGFP+ gate

The data in Chapter 3 were analysed using an automatic gate on the EGFP marker. These data include samples with two fluorescence markers – EGFP and mKATE, when the loading module was co-transfected with the library plasmids – and samples with only one fluorescence marker – EGFP, when the pEmpty was

co-transfected with the library plasmids. Using an OR gate to analyse all the samples would lead to an underestimation of the EGFP fluorescence when the library plasmids are co-transfected with the loading module (Figure 8.7).

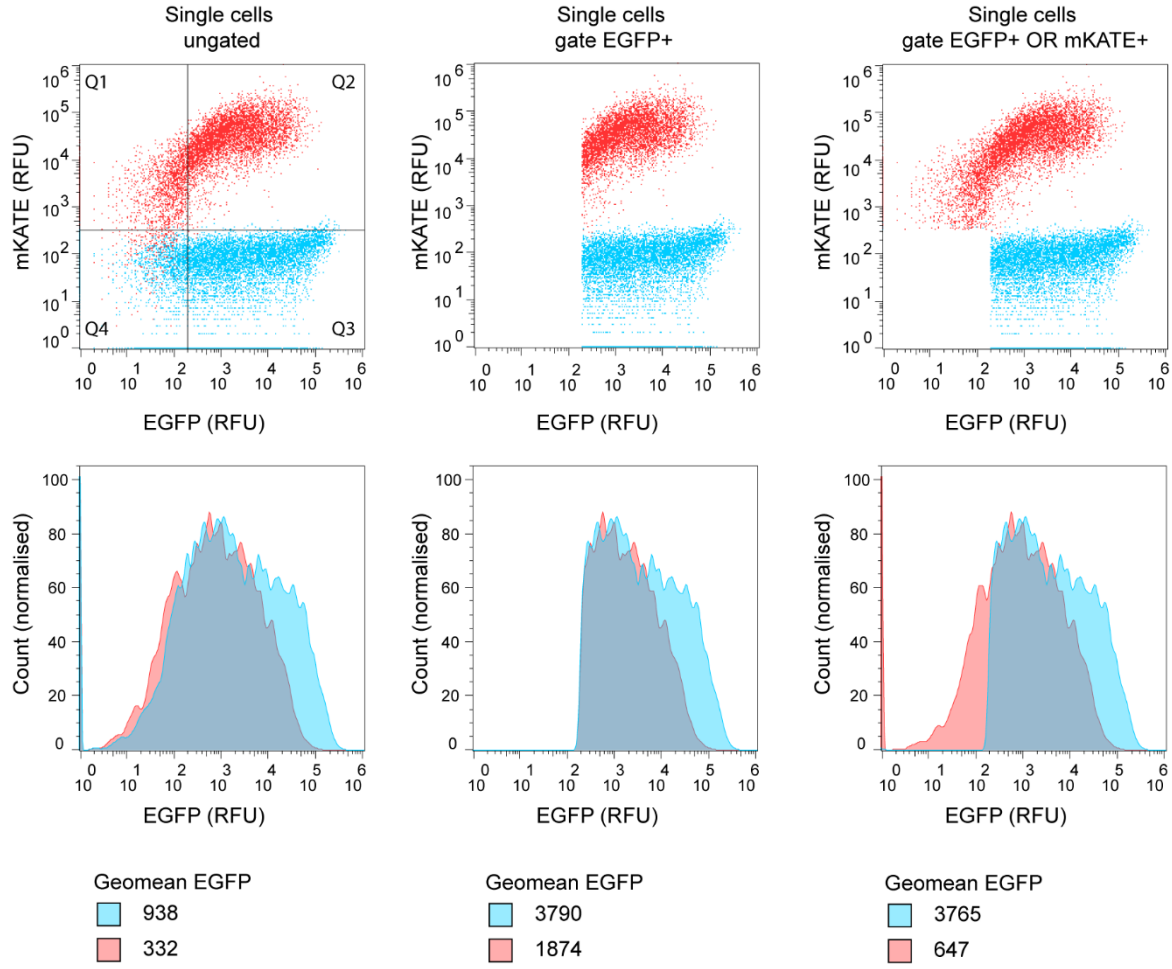

Figure 8.7 Justification for the use of an EGFP+ gate. An EGFP-expressing plasmid (EF1ap-Kz1-EGFP-SV40pA_rv) was co-transfected either with an mKATE-expressing loading module (red) or a pEmpty (blue) in HEK293T. Overlaid fluorescence dot plots (top panels) and histograms (bottom panel) of the single cell population of the two co-transfected samples. When comparing the ungated single cell population (left panels) with either an EGFP positive gate (EGFP+, middle panels) or an OR gate on the two fluorescence markers (EGFP+ OR mKATE+, right panels), the OR gate gives an incorrect estimation of the mKATE fluorescence, as evidenced by the reported geomean values. While the red and blue samples have similar geomean mKATE levels in the ungated and in the EGFP+-gated populations, this is not the case when an OR gate is applied. This indicates that OR gates are unsuitable for this specific data*set* and should not be used to avoid experimental artefacts.

8.3. Parts used in this study

Table 8.6 Parts used in this study

Promoters	
pJB42CAT5	CTGACAAATTCAGTATAAAAGCTTGGGGCTGGGGCCGAGCACTGGGGACTTTGAGGGTGGCCAGGCCAGCGTA GGAGGCCAGCGTAGGATCCTGCTGGGAGCGGGGAACTGAGGGAAGCGACGCCGAGAAAGCAGGCGTACCACG GAGGGAGAGAAAAGCTCCGGAAGCCCAGCAGCG
hACTBp	ACTGCCTGGCCACTCCATGCCCTCCAAGAGCTCCTTCTGCAGGAGCGTACAGAACCCAGGGCCCTGGCACCCGTG CAGACCCTGGCCCACCCCACCTGGGCGCTCAGTGCCCAAGAGATGTCCACACCTAGGATGTCCCGCGGTGGGTG GGGGGCCCGAGAGACGGGCAGGCCGGGGGCAGGCCTGGCCATGCGGGGCCGAACCGGGCACTGCCCAGCGTG GGGCGCGGGGGCCACGGCGCGCGCCCCCAGCCCCCGGGCCCAGCACCCCAAGGCGGCCAACGCCAAAACTCTCC CTCCTCCTCTTCCTCAATCTCGCTCTCGCTCTTTTTTTTTTTCGCAAAAGGAGGGGAGAGGGGGTAAAAAAATGCT GCACTGTGCGGCGAAGCCGGTGAGTGAGCGGCGCGGGGCCAATCAGCGTGCGCCGTTCCGAAAGTTGCCTTTTA TGGCTCGAGCGGCCGCGGCGGCGCCCTATAAAACCCAGCGGCGCGACGCGCCACCACCGCCGAGACCGCGTCCG CCCCGCGAGCACAGAGCCTCGCCTTTGCCGA
SV40p	GTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTA GTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTC AGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATG GCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCTGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGG CTTTTTTGGAGGCCTAGGCTTTTGCAAA
PGKp	GGGGTTGGGGTTGCGCCTTTTCCAAGGCAGCCCTGGGTTTGCGCAGGGACGCGGCTGCTCTGGGCGTGGTTCCG GGAAACGCAGCGGCGCCGACCCTGGGTCTCGCACATTCTTCACGTCCGTTCGCAGCGTCACCCGGATCTTCGCCG CTACCCTTGTGGGCCCCCCGGCGACGCTTCCTGCTCCGCCCCTAAGTCGGGAAGGTTCCTTGCGGTTCGCGGCGT GCCGGACGTGACAAACGGAAGCCGCACGTCTCACTAGTACCCTCGCAGACGGACAGCGCCAGGGAGCAATGGC AGCGCGCCGACCGCGATGGGCTGTGGCCAATAGCGGCTGCTCAGCAGGGCGCGCCGAGAGCAGCGGCCGGGA AGGGGCGGTGCGGGAGGCGGGGTGTGGGGCGGTAGTGTGGGCCCTGTTCCTGCCCGCGCGGTGTTCCGCATTC TGCAAGCCTCCGGAGCGCACGTCGGCAGTCGGCTCCCTCGTTGACCGAATCACCGACCTCTCTCCCCAG
UBp	GGCCTCCGCGCCGGGTTTTGGCGCCTCCCGCGGGCGCCCCCCTCCTCACGGCGAGCGCTGCCACGTCAGACGAA GGGCGCAGCGAGCGTCCTGATCCTTCCGCCCGGACGCTCAGGACAGCGGCCCGCTGCTCATAAGACTCGGCCTTA GAACCCCAGTATCAGCAGAAGGACATTTTAGGACGGGACTTGGGTGACTCTAGGGCACTGGTTTTCTTTCCAGAG AGCGGAACAGGCGAGGAAAAGTAGTCCCTTCTCGGCGATTCTGCGGAGGGATCTCCGTGGGGCGGTGAACGCC GATGATTATATAAGGACGCGCCGGGTGTGGCACAGCTAGTTCCGTCGCAGCCGGGATTTGGGTCGCGGTTCTTGT TTGTGGATCGCTGTGATCGTCACTTGGTGAGTAGCGGGCTGCTGGGCTGGCCGGGGCTTTCGTGGCCGCCGGGC CGCTCGGTGGGACGGAAGCGTGTGGAGAGACCGCCAAGGGCTGTAGTCTGGGTCCGCGAGCAAGGTTGCCCTG AACTGGGGGTTGGGGGGAGCGCAGCAAAATGGCGGCTGTTCCCGAGTCTTGAATGGAAGACGCTTGTGAGGCG GGCTGTGAGGTCGTTGAAACAAGGTGGGGGGCATGGTGGGCGGCAAGAACCCAAGGTCTTGAGGCCTTCGCTA ATGCGGGAAAGCTCTTATTCGGGTGAGATGGGCTGGGGCACCATCTGGGGACCCTGACGTGAAGTTTGTCACTG ACTGGAGAACTCGGTTTGTCGTCTGTTGCGGGGGCGGCAGTTATGGCGGTGCCGTTGGGCAGTGCACCCGTACC TTTGGGAGCGCGCGCCCTCGTCGTGTCGTGACGTCACCCGTTCTGTTGGCTTATAATGCAGGGTGGGGCCACCTG CCGGTAGGTGTGCGGTAGGCTTTTCTCCGTCGCAGGACGCAGGGTTCGGGCCTAGGGTAGGCTCTCCTGAATCG ACAGGCGCCGGACCTCTGGTGAGGGGAGGGATAAGTGAGGCGTCAGTTTCTTTGGTCGGTTTTATGTACCTATCT TCTTAAGTAGCTGAAGCTCCGGTTTTGAACTATGCGCTCGGGGTTGGCGAGTGTGTTTTGTGAAGTTTTTTAGGCA CCTTTTGAAATGTAATCATTTGGGTCAATATGTAATTTTCAGTGTTAGACTAGTAAATTGTCCGCTAAATTCTGGCC GTTTTTGGCTTTTTTGTTAGAC
CMVp	CGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGAC GTATGTTCCCATAGTAACGCCAATAGGGACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCAC TTGGCAGTACATCAAGTGTATCATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGG CATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATCGCTATTACCAT GGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGGGGATTTCCAAGTCTCCACCC CATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACAACTCCGCCCCA TTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTGGTTTAGTGAACCGTCAGAT CGAATTCGACCCAAGTTTGTACAAAAAAGCAGGCT

EF1ap	GGCTCCGGTGCCCGTCAGTGGGCAGAGCGCACATCGCCCACAGTCCCCGAGAAGTTGGGGGGAGGGGTCGGCA ATTGAACCGGTGCCTAGAGAAGGTGGCGCGGGGTAAACTGGGAAAGTGATGTCGTGTACTGGCTCCGCCTTTTT CCCGAGGGTGGGGGAGAACCGTATATAAGTGCAGTAGTCGCCGTGAACGTTCTTTTTCGCAACGGGTTTGCCGCC AGAACACAGGTAAGTGCCGTGTGTGGTTCCCGCGGGCCTGGCCTCTTTACGGGTTATGGCCCTTGCGTGCCTTGA ATTACTTCCACCTGGCTGCAGTACGTGATTCTTGATCCCGAGCTTCGGGTTGGAAGTGGGTGGGAGAGTTCGAGG CCTTGCGCTTAAGGAGCCCCTTCGCCTCGTGCTTGAGTTGAGGCCTGGCCTGGGCGCTGGGGCCGCCGCGTGCGA ATCTGGTGGCACCTTCGCGCCTGTCTCGCTGCTTTCGATAAGTCTCTAGCCATTTAAAATTTTTGATGACCTGCTGC GACGCTTTTTTTCTGGCAAGATAGTCTTGTAAATGCGGGCCAAGATCTGCACACTGGTATTTCGGTTTTTGGGGCC GCGGGCGGCGACGGGGCCCGTGCGTCCCAGCGCACATGTTCGGCGAGGCGGGGCCTGCGAGCGCGGCCACCGA GAATCGGACGGGGGTAGTCTCAAGCTGGCCGGCCTGCTCTGGTGCCTGGCCTCGCGCCGCCGTGTATCGCCCCG CCCTGGGCGGCAAGGCTGGCCCGGTCGGCACCAGTTGCGTGAGCGGAAAGATGGCCGCTTCCCGGCCCTGCTGC AGGGAGCTCAAAATGGAGGACGCGGCGCTCGGGAGAGCGGGCGGGTGAGTCACCCACACAAAGGAAAAGGGC CTTTCCGTCCTCAGCCGTCGCTTCATGTGACTCCACGGAGTACCGGGCGCCGTCCAGGCACCTCGATTAGTTCTCG AGCTTTTGGAGTACGTCGTCTTTAGGTTGGGGGGAGGGGTTTTATGCGATGGAGTTTCCCCACACTGAGTGGGTG GAGACTGAAGTTAGGCCAGCTTGGCACTTGATGTAATTCTCCTTGGAATTTGCCCTTTTTGAGTTTGGATCTTGGT TCATTCTCAAGCCTCAGACAGTGGTTCAAAGTTTTTTTCTTCCATTTCAGGTGTCGTGACGCTAGCGCTACCGGACT CAGATCTCGAGCTCAAGCTTCGAATTC
TRE3Gp (TETp)	GACTTTATACGAAGTTATCTCGAGTTTACTCCCTATCAGTGATAGAGAACGTATGAAGAGTTTACTCCCTATCAGT GATAGAGAACGTATGCAGACTTTACTCCCTATCAGTGATAGAGAACGTATAAGGAGTTTACTCCCTATCAGTGATA GAGAACGTATGACCAGTTTACTCCCTATCAGTGATAGAGAACGTATCTACAGTTTACTCCCTATCAGTGATAGAGA ACGTATATCCAGTTTACTCCCTATCAGTGATAGAGAACGTATAAGCTTTAGGCGTGTACGGTGGGCGCCTATAAA AGCAGAGCTCGTTTAGTGAACCGTCAGATCGCCTGGAGCAATTCCACAACACTTTTGTCTTATACCAACTTTCCGT ACCACTTCCTACCCTCGTAAAGGA
Coding sequences	
EGFP	ATGGGAGTGAGCAAGGGCGAGGAGCTGTTCACCGGGGTGGTGCCCATCCTGGTCGAGCTGGACGGCGACGTAA ACGGCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATC TGCACCACCGGCAAGCTGCCCGTGCCCTGGCCCACCCTCGTGACCACCCTGACCTACGGCGTGCAGTGCTTCAGC CGCTACCCCGACCACATGAAGCAGCACGACTTCTTCAAGTCCGCCATGCCCGAAGGCTACGTCCAGGAGCGCACC ATCTTCTTCAAGGACGACGGCAACTACAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCG CATCGAGCTGAAGGGCATCGACTTCAAGGAGGACGGCAACATCCTGGGGCACAAGCTGGAGTACAACTACAACA GCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGGCATCAAGGTGAACTTCAAGATCCGCCACAACATCG AGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCCATCGGCGACGGCCCCGTGCTGCTGCCC GACAACCACTACCTGAGCACCCAGTCCGCCCTGAGCAAAGACCCCAACGAGAAGCGCGATCACATGGTCCTGCTG GAGTTCGTGACCGCCGCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTAA
mKATE	ATGGTGTCTAAGGGCGAAGAGCTGATTAAGGAGAACATGCACATGAAGCTGTACATGGAGGGCACCGTGAACA ACCACCACTTCAAGTGCACATCCGAGGGCGAAGGCAAGCCCTACGAGGGCACCCAGACCATGAGAATCAAGGTG GTCGAGGGCGGCCCTCTCCCCTTCGCCTTCGACATCCTGGCTACCAGCTTCATGTACGGCAGCAAAACCTTCATCA ACCACACCCAGGGCATCCCCGACTTCTTTAAGCAGTCCTTCCCTGAGGGCTTCACATGGGAGAGAGTCACCACATA CGAAGACGGGGGCGTGCTGACCGCTACCCAGGACACCAGCCTCCAGGACGGCTGCCTCATCTACAACGTCAAGA TCAGAGGGGTGAACTTCCCATCCAACGGCCCTGTGATGCAGAAGAAAACACTCGGCTGGGAGGCCTCCACCGAG ATGCTGTACCCCGCTGACGGCGGCCTGGAAGGCAGAAGCGACATGGCCCTGAAGCTCGTGGGCGGGGGCCACC TGATCTGCAACTTGAAGACCACATACAGATCCAAGAAACCCGCTAAGAACCTCAAGATGCCCGGCGTCTACTATG TGGACAGAAGACTGGAAAGAATCAAGGAGGCCGACAAAGAGACCTACGTCGAGCAGCACGAGGTGGCTGTGGC CAGATACTGCGACCTCCCTAGCAAACTGGGGCACAAACTTAATTGA
mCherry	ATGGTGAGCAAGGGCGAGGAGGATAACATGGCCATCATCAAGGAGTTCATGCGCTTCAAGGTGCACATGGAGG GCTCCGTGAACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCCAGACCGC CAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCTGTCCCCTCAGTTCATGTACGGCTC CAAGGCCTACGTGAAGCACCCCGCCGACATCCCCGACTACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGA GCGCGTGATGAACTTCGAGGACGGCGGCGTGGTGACCGTGACCCAGGACTCCTCACTGCAGGACGGCGAGTTCA TCTACAAGGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCATGGGCTGG GAGGCCTCCTCCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGAGATCAAGCAGAGGCTGAAGCTGA AGGACGGCGGCCACTACGACGCTGAGGTCAAGACCACCTACAAGGCCAAGAAGCCCGTGCAGCTGCCCGGCGCC TACAACGTCAACATCAAGTTGGACATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAACGCGCT GAGGGCCGCCACTCCACCGGCGGCATGGACGAGCTGTACAAGTAG

mTagBPF2	ATGGGAATGGTGTCTAAGGGCGAAGAGCTGATTAAGGAGAACATGCACATGAAGCTGTACATGGAGGGCACCG TGGACAACCATCACTTCAAGTGCACATCCGAGGGCGAAGGCAAGCCCTACGAGGGCACCCAGACCATGAGAATC AAGGTGGTCGAGGGCGGCCCTCTCCCCTTCGCCTTCGACATCCTGGCTACTAGCTTCCTCTACGGCAGCAAGACCT TCATCAACCACACCCAGGGCATCCCCGACTTCTTCAAGCAGTCCTTCCCTGAGGGCTTCACATGGGAGAGAGTCAC CACATACGAAGACGGGGGCGTGCTGACCGCTACCCAGGACACCAGCCTCCAGGACGGCTGCCTCATCTACAACG TCAAGATCAGAGGGGTGAACTTCACATCCAACGGCCCTGTGATGCAGAAGAAAACACTCGGCTGGGAGGCCTTC ACCGAAACGCTGTACCCCGCTGACGGCGGCCTGGAAGGCAGAAACGACATGGCCCTGAAGCTCGTGGGCGGGA GCCATCTGATCGCAAACGCCAAGACCACATATAGATCCAAGAAACCCGCTAAGAACCTCAAGATGCCTGGCGTCT ACTATGTGGACTACAGACTGGAAAGAATCAAGGAGGCCAACAACGAAACCTACGTCGAGCAGCACGAGGTGGC AGTGGCCAGATACTGCGACCTCCCTAGCAAACTGGGGCACAAGCTTAACTAA
TETON3G	ATGTCTAGACTGGACAAGAGCAAAGTCATAAACTCTGCTCTGGAATTACTCAATGGAGTCGGTATCGAAGGCCTG ACGACAAGGAAACTCGCTCAAAAGCTGGGAGTTGAGCAGCCTACCCTGTACTGGCACGTGAAGAACAAGCGGGC CCTGCTCGATGCCCTGCCAATCGAGATGCTGGACAGGCATCATACCCACTCCTGCCCCCTGGAAGGCGAGTCATG GCAAGACTTTCTGCGGAACAACGCCAAGTCATACCGCTGTGCTCTCCTCTCACATCGCGACGGGGCTAAAGTGCA TCTCGGCACCCGCCCAACAGAGAAACAGTACGAAACCCTGGAAAATCAGCTCGCGTTCCTGTGTCAGCAAGGCTT CTCCCTGGAGAACGCACTGTACGCTCTGTCCGCCGTGGGCCACTTTACACTGGGCTGCGTATTGGAGGAACAGGA GCATCAAGTAGCAAAAGAGGAAAGAGAGACACCTACCACCGATTCTATGCCCCCACTTCTGAAACAAGCAATTGA GCTGTTCGACCGGCAGGGAGCCGAACCTGCCTTCCTTTTCGGCCTGGAACTAATCATATGTGGCCTGGAGAAACA GCTAAAGTGCGAAAGCGGCGGGCCGACCGACGCCCTTGACGATTTTGACTTAGACATGCTCCCAGCCGATGCCCT TGACGACTTTGACCTTGATATGCTGCCTGCTGACGCTCTTGACGATTTTGACCTTGACATGCTCCCCGGGTCACTAA GTAGCGGCTAA
EBNA1	ATGGGATCTGACGAGGGGCCAGGTACAGGACCTGGAAATGGCCTAGGAGAGAAGGGAGACACATCTGGACCAG AAGGCTCCGGCGGCAGTGGACCTCAAAGAAGAGGGGGTGATAACCATGGACGAGGACGGGGAAGAGGACGAG GACGAGGAGGCGGAAGACCAGGAGCCCCGGGCGGCTCAGGATCAGGGCCAAGACATAGAGATGGTGTCCGGA GACCCCAAAAACGTCCAAGTTGCATTGGCTGCAAAGGGACCCACGGTGGAACAGGAGCAGGAGCAGGAGCGGG AGGGGCAGGAGCAGGAGGTGGAGGCCGGGGTCGAGGAGGCAGTGGAGGCCGGGGTCGAGGAGGTAGTGGA GGCCGGGGTCGAGGAGGTAGTGGAGGCCGCCGGGGTAGAGGACGTGAAAGAGCCAGGGGGGGAAGTCGTGA AAGAGCCAGGGGGAGAGGTCGTGGACGTGGAGAAAAGAGGCCCAGGAGTCCCAGTAGTCAGTCATCATCATCC GGGTCTCCACCGCGCAGGCCCCCTCCAGGTAGAAGGCCATTTTTCCACCCTGTAGGGGAAGCCGATTATTTTGAA TACCACCAAGAAGGTGGCCCAGATGGTGAGCCTGACGTGCCCCCGGGAGCGATAGAGCAGGGCCCCGCAGATG ACCCAGGAGAAGGCCCAAGCACTGGACCCCGGGGTCAGGGTGATGGAGGCAGGCGCAAAAAAGGAGGGTGGT TTGGAAAGCATCGTGGTCAAGGAGGTTCCAACCCGAAATTTGAGAACATTGCAGAAGGTTTAAGAGCTCTCCTGG CTAGGAGTCACGTAGAAAGGACTACCGACGAAGGAACTTGGGTCGCCGGTGTGTTCGTATATGGAGGTAGTAAG ACCTCCCTTTACAACCTAAGGCGAGGAACTGCCCTTGCTATTCCACAATGTCGTCTTACACCATTGAGTCGTCTCCC CTTTGGAATGGCCCCTGGACCCGGCCCACAACCTGGCCCGCTAAGGGAGTCCATTGTCTGTTATTTCATGGTCTTT TTACAAACTCATATATTTGCTGAGGTTTTGAAGGATGCGATTAAGGACCTTGTTATGACAAAGCCCGCTCCTACCT GCAATATCAGGGTGACTGTGTGCAGCTTTGACGATGGAGTAGATTTGCCTCCCTGGTTTCCACCTATGGTGGAAG GGGCTGCCGCGGAGGGTGATGACGGAGATGACGGAGATGAAGGAGGTGATGGAGATGAGGGTGAGGAAGGG CAGGAGTAA
BxBI integrase	ATGCCAAAAAAGAAAAGAAAAGTGTATCCCTATGATGTCCCCGATTATGCCGGTTCAAGAGCCCTGGTCGTGATT AGACTGAGCCGAGTGACAGACGCCACCACAAGTCCCGAGAGACAGCTGGAATCATGCCAGCAGCTCTGTGCTCA GCGGGGTTGGGATGTGGTCGGCGTGGCAGAGGATCTGGACGTGAGCGGGGCCGTCGATCCATTCGACAGAAAG AGGAGGCCCAACCTGGCAAGATGGCTCGCTTTCGAGGAACAGCCCTTTGATGTGATCGTCGCCTACAGAGTGGA CCGGCTGACCCGCTCAATTCGACATCTCCAGCAGCTGGTGCATTGGGCTGAGGACCACAAGAAACTGGTGGTCA GCGCAACAGAAGCCCACTTCGATACTACCACACCTTTTGCCGCTGTGGTCATCGCACTGATGGGCACTGTGGCCC AGATGGAGCTCGAAGCTATCAAGGAGCGAAACAGGAGCGCAGCCCATTTCAATATTAGGGCCGGTAAATACAGA GGCTCCCTGCCCCCTTGGGGATATCTCCCTACCAGGGTGGATGGGGAGTGGAGACTGGTGCCAGACCCCGTCCA GAGAGAGCGGATTCTGGAAGTGTACCACAGAGTGGTCGATAACCACGAACCACTCCATCTGGTGGCACACGACC TGAATAGACGCGGCGTGCTCTCTCCAAAGGATTATTTTGCTCAGCTGCAGGGAAGAGAGCCACAGGGAAGAGAA TGGAGTGCTACTGCACTGAAGAGATCTATGATCAGTGAGGCTATGCTGGGTTACGCAACACTCAATGGCAAAACT GTCCGGGACGATGACGGAGCCCCTCTGGTGAGGGCTGAGCCTATTCTCACCAGAGAGCAGCTCGAAGCTCTGCG GGCAGAACTGGTCAAGACTAGTCGCGCCAAACCTGCCGTGAGCACCCCAAGCCTGCTCCTGAGGGTGCTGTTCTG CGCCGTCTGTGGAGAGCCAGCATACAAGTTTGCCGGCGGAGGGCGCAAACATCCCCGCTATCGATGCAGGAGCA TGGGGTTCCCTAAGCACTGTGGAAACGGGACAGTGGCCATGGCTGAGTGGGACGCCTTTTGCGAGGAACAGGT GCTGGATCTCCTGGGTGACGCTGAGCGGCTGGAAAAAGTGTGGGTGGCAGGATCTGACTCCGCTGTGGAGCTG GCAGAAGTCAATGCCGAGCTCGTGGATCTGACTTCCCTCATCGGATCTCCTGCATATAGAGCTGGGTCCCCACAG AGAGAAGCTCTGGACGCACGAATTGCTGCACTCGCTGCTAGACAGGAGGAACTGGAGGGCCTGGAGGCCAGGC CCTCTGGATGGGAGTGGCGAGAAACCGGACAGAGGTTTGGGGATTGGTGGAGGGAGCAGGACACCGCAGCCA

	AGAACACATGGCTGAGATCCATGAATGTCCGGCTCACATTCGACGTGCGCGGTGGCCTGACTCGAACCATCGATT TTGGCGACCTGCAGGAGTATGAACAGCACCTGAGACTGGGGTCCGTGGTCGAAAGACTGCACACTGGGATGTCC TAG
OsTIR1	ATGACATACTTTCCTGAAGAGGTCGTCGAACACATTTTTAGCTTCCTGCCTGCACAGAGAGATAGAAACACAGTG AGCCTGGTCTGCAAAGTGTGGTACGAGATCGAACGCCTGAGCCGGAGAGGAGTGTTCGTCGGCAACTGCTATGC TGTGAGAGCAGGCAGGGTCGCCGCTAGGTTTCCAAATGTGCGCGCACTGACCGTCAAGGGGAAACCCCACGGCG CCGACTTTAACCTGGTGCCCCCTGATTGGGGAGGATACGCCGGCCCTTGGATCGAGGCAGCCGCTCGCGGCTGTC ATGGACTGGAGGAACTGCGCATGAAGCGAATGGTGGTCTCTGACGAAAGTCTGGAGCTGCTGGCTCGGAGCTTC CCTAGGTTTCGCGCACTGGTGCTGATTTCTTGCGAAGGCTTCAGCACCGATGGACTGGCAGCCGTGGCCTCCCAC TGTAAGCTGCTGCGGGAGCTGGACCTCCAGGAGAATGAAGTGGAGGATAGAGGCCCCAGATGGCTGTCTTGCTT CCCAGACTCATGTACCAGCCTGGTGTCCCTGAACTTTGCCTGCATCAAAGGCGAAGTGAATGCTGGGTCCCTGGA GCGGCTGGTCTCAAGAAGCCCCAACCTGAGGTCTCTGCGGCTGAACCGGAGCGTGAGCGTGGACACTCTGGCTA AGATTCTGCTGAGAACCCCTAACCTGGAGGATCTGGGAACCGGCAATCTGACAGACGATTTCCAGACAGAATCCT ACTTTAAACTGACTTCTGCCCTGGAGAAGTGTAAAATGCTGAGGAGTCTGTCAGGATTCTGGGATGCTTCACCCG TGTGCCTGAGCTTTATCTACCCTCTGTGTGCACAGCTGACAGGCCTGAACCTGAGCTATGCACCAACCCTGGACGC CAGTGATCTGACAAAGATGATCTCACGCTGCGTGAAACTCCAGCGACTGTGGGTGCTGGACTGTATTTCCGATAA GGGGCTCCAGGTGGTCGCCAGCTCCTGCAAGGACCTCCAGGAGCTGAGAGTGTTCCCATCTGATTTTTACGTGGC CGGATATAGTGCTGTCACTGAGGAAGGCCTGGTGGCAGTCTCACTGGGATGCCCAAAGCTGAACAGCCTGCTGT ATTTCTGTCATCAGATGACTAATGCTGCACTGGTGACCGTCGCCAAGAACTGCCCTAATTTCACCCGATTTCGGCT GTGTATTCTGGAACCAGGCAAACCCGACGTGGTCACATCCCAGCCACTGGATGAAGGGTTTGGAGCTATCGTGA GAGAGTGCAAGGGACTCCAGAGGCTGAGCATTTCCGGCCTGCTGACAGACAAAGTGTTCATGTACATCGGCAAG TATGCTAAGCAGCTGGAGATGCTGAGCATTGCATTTGCCGGAGACTCCGATAAGGGCATGATGCACGTGATGAA CGGGTGTAAGAATCTGCGAAAACTGGAAATCCGGGACAGCCCTTTCGGGGATGCCGCTCTGCTGGGAAACTTTG CCAGATACGAGACAATGAGGAGCCTGTGGATGTCTAGTTGCAATGTGACTCTGAAGGGCTGTCAGGTCCTGGCT AGTAAAATGCCTATGCTGAACGTGGAAGTCATTAATGAGCGGGACGGGTCTAACGAAATGGAGGAAAATCATGG CGACCTGCCAAAGGTGGAGAAACTGTATGTGTATCGGACCACCGCAGGGGCAAGAGATGATGCTCCCAACTTTG TGAAGATTCTG
DHFR	ATGATCAGTCTGATTGCGGCGTTAGCGGTAGATTACGTTATCGGCATGGAAAACGCCATGCCGTGGAACCTGCCT GCCGATCTCGCCTGGTTTAAACGCAACACCTTAAATAAACCCGTGATTATGGGCCGCCATACCTGGGAATCAATCG GTCGTCCGTTGCCAGGACGCAAAAATATTATCCTCAGCAGTCAACCGAGTACGGACGATCGCGTAACGTGGGTGA AGTCGGTGGATGAAGCCATCGCGGCGTGTGGTGACGTACCAGAAATCATGGTGATTGGCGGCGGTCGCGTTATT GAACAGTTCTTGCCAAAAGCGCAAAAACTGTATCTGACGCATATCGACGCAGAAGTGGAAGGCGACACCCATTTC CCGGATTACGAGCCGGATGACTGGGAATCGGTATTCAGCGAATTCCACGATGCTGATGCGCAGAACTCTCACAGC TATTGCTTTGAGATTCTGGAGCGGCGA
Blastidicin resistance	ATGGCGTCCGCCAAGCCTCTGAGCCAGGAGGAGAGCACCCTGATCGAGCGCGCCACCGCCACCATCAACAGCAT CCCTATCAGCGAAGACTACAGCGTGGCCAGCGCCGCTCTGAGCAGCGACGGCCGCATCTTCACAGGAGTGAACG TGTACCACTTCACCGGAGGACCTTGCGCCGAACTGGTGGTGCTGGGCACCGCTGCCGCTGCTGCCGCTGGAAACC TGACCTGCATCGTGGCCATCGGCAACGAGAACCGCGGAATCCTGAGCCCTTGCGGCCGCTGCCGCCAGGTGCTG CTGGACCTGCACCCCGGCATCAAGGCCATCGTGAAGGACAGCGACGGCCAGCCCACCGCCGTGGGCATCCGCGA GCTGCTGCCCAGCGGCTACGTGTGGGAAGGCTAA
Puromycin resistance	ATGTCCACCGAGTACAAGCCCACGGTGCGCCTCGCCACCCGCGACGACGTCCCCAGGGCCGTACGCACCCTCGCC GCCGCGTTCGCCGACTACCCCGCCACGCGCCACACCGTCGATCCGGACCGCCACATCGAGCGGGTCACCGAGCTG CAAGAACTCTTCCTCACGCGCGTCGGGCTCGACATCGGCAAGGTGTGGGTCGCGGACGACGGCGCCGCGGTGGC GGTCTGGACCACGCCGGAGAGCGTCGAAGCGGGGGCGGTGTTCGCCGAGATCGGCCCGCGCATGGCCGAGTTG AGCGGTTCCCGGCTGGCCGCGCAGCAACAGATGGAAGGCCTCCTGGCGCCGCACCGGCCCAAGGAGCCCGCGT GGTTCCTGGCCACCGTCGGCGTGTCGCCCGACCACCAGGGCAAGGGTCTGGGCAGCGCCGTCGTGCTCCCCGGA GTGGAGGCGGCCGAGCGCGCCGGGGTGCCCGCCTTCCTGGAAACCTCCGCGCCCCGCAACCTCCCCTTCTACGA GCGGCTCGGCTTCACCGTCACCGCCGACGTCGAGTGCCCGAAGGACCGCGCGACCTGGTGCATGACCCGCAAGC CCGGTGCCTAA
Kozak sequences	
Kz1	CGCCACCATGG
Kz2	CGTGCGCATGC
Kz3	CGGCGACATGC

PolyA signals	
SV40pA	ACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCA CTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGT
SV40pA_rv	ACAGACATGATAAGATACATTGATGAGTTTGGACAAACCACAACTAGAATGCAGTGAAAAAAATGCTTTATTTGT GAAATTTGTGATGCTATTGCTTTATTTGTAACCATTATAAGCTGCAATAAACAAGT
PGKpA	CCTTGAGCATCTGACTTCTGGCTAAATTGATGATCTATTAAACAATAAAGATGTCCACATGGAAGTTTTTTCCTGTC ATACTTTGTTAAGAAGGGTGAGAACAGAGTACCTACATTTTGAATGGAAGGATTGGAGCTACGGGGGTGGGGGT GGGGTGGGATTAGATAAATGCCTGCTCTTTACTGAAGGCTCTTTACTATTGCTTTATGATAATGTTTCATAGTTGG AT
RBpA	ACTCCTCAGGTGCAGGCTGCCTATCAGAAGGTGGTGGCTGGTGTGGCCAATGCCCTGGCTCACAAATACCACTGA GATCTTTTTCCCTCTGCCAAAAATTATGGGGACATCATGAAGCCCCTTGAGCATCTGACTTCTGGCTAATAAAGGA AATTTATTTTCATTGCAATAGTGTGTTGGAATTTTTTGTGTCTCTCACTCGGAAGGACATATGGGAGGGCAAATCA TTTAAAACATCAGAATGAGTATTTGGTTTAGAGTTTGGCAACATATGCCCATATGCTGGCTGCCATGAACAAAGGT TGGCTATAAAGAGGTCATCAGTATATGAAACAGCCCCCTGCTGTCCATTCCTTATTCCATAGAAAAGCCTTGACTT GAGGTTAGATTTTTTTTATATTTTGTTTTGTGTTATTTTTTTCTTTAACATCCCTAAAATTTTCCTTACATGTTTTACTA GCCAGATTTTTCCTCCTCTCCTGACTACTCCCAGTCATAGCTGTCCCTCTTCTCTTATGAAGATCCCTCGACCTGCAG
HGHpA	GGGTGGCATCCCTGTGACCCCTCCCCAGTGCCTCTCCTGGCCCTGGAAGTTGCCACTCCAGTGCCCACCAGCCTTG TCCTAATAAAATTAAGTTGCATCATTTTGTCTGACTAGGTGTCCTTCTATAATATTATGGGGTGGAGGGGGGTGGT ATGGAGCAAGGGGCAAGTTGGGAAGACAACCTGTAGGGCCTGCGGGGTCTATTGGGAACCAAGCTGGAGTGCA GTGGCACAATCTTGGCTCACTGCAATCTCCGCCTCCTGGGTTCAAGCGATTCTCCTGCCTCAGCCTCCCGAGTTGTT GGGATTCCAGGCATGCATGACCAGGCTCAGCTAATTTTTGTTTTTTTGGTAGAGACGGGGTTTCACCATATTGGCC AGGCTGGTCTCCAACTCCTAATCTCAGGTGATCTACCCACCTTGGCCTCCCAAATTGCTGGGATTACAGGCGTGAA CCACTGCTCCCTTCCCTGTCCTT
BGHpA	CTCTGGGGTTCGAAATGACCGACCAAGCGACGCTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCC GTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCT GAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAG CAGGCATGCTGGGGATGCGGTGGGCTCTATGGG
Degradation domains	
UbM	ATGACTAGTCAGATTTTCGTGAAGACCCTGACCGGCAAGACCATCACCCTAGAGGTGGAGCCCAGTGACACCATC GAGAACGTGAAGGCCAAGATCCAGGATAAAGAGGGCATCCCCCCTGACCAGCAGAGGCTGATCTTTGCCGGCAA GCAGCTGGAAGATGGCCGCACCCTCTCTGATTACAACATCCAGAAGGAGTCAACCCTGCACCTGGTCCTTCGCCT GAGAGGTGGC
UbR	ATGACTAGTCAGATTTTCGTGAAGACCCTGACCGGCAAGACCATCACCCTAGAGGTGGAGCCCAGTGACACCATC GAGAACGTGAAGGCCAAGATCCAGGATAAAGAGGGCATCCCCCCTGACCAGCAGAGGCTGATCTTTGCCGGCAA GCAGCTGGAAGATGGCCGCACCCTCTCTGATTACAACATCCAGAAGGAGTCAACCCTGCACCTGGTCCTTCGCCT GAGAGGTGGCCGG
UbVR	ATGACTAGTCAGATTTTCGTGAAGACCCTGACCGGCAAGACCATCACCCTAGAGGTGGAGCCCAGTGACACCATC GAGAACGTGAAGGCCAAGATCCAGGATAAAGAGGGCATCCCCCCTGACCAGCAGAGGCTGATCTTTGCCGGCAA GCAGCTGGAAGATGGCCGCACCCTCTCTGATTACAACATCCAGAAGGAGTCAACCCTGCACCTGGTCCTTCGCCT GAGAGGTGTCAGG
2xUbAV	ATGACTAGTCAGATTTTCGTGAAGACCCTGACCGGCAAGACCATCACCCTAGAGGTGGAGCCCAGTGACACCATC GAGAACGTGAAGGCCAAGATCCAGGATAAAGAGGGCATCCCCCCTGACCAGCAGAGGCTGATCTTTGCCGGCAA GCAGCTGGAAGATGGCCGCACCCTCTCTGATTACAACATCCAGAAGGAGTCAACCCTGCACCTGGTCCTTCGCCT GAGAGGTGCCGTGGCTAGTCAGATTTTCGTGAAGACCCTGACCGGCAAGACCATCACCCTAGAGGTGGAGCCCA GTGACACCATCGAGAACGTGAAGGCCAAGATCCAGGATAAAGAGGGCATCCCCCCTGACCAGCAGAGGCTGATC TTTGCCGGCAAGCAGCTGGAAGATGGCCGCACCCTCTCTGATTACAACATCCAGAAGGAGTCAACCCTGCACCTG GTCCTTCGCCTGAGAGGTGCCGTG
PEST	TCTCACGGCTTCCCTCCCGAGGTGGAGGAGCAGGCCGCCGGCACCCTGCCCATGAGCTGCGCCCAGGAGAGCGG CATGGATAGACACCCTGCTGCTTGCGCCAGCGCCAGGATCAACGTC
mAID	AAGGAGAAGAGTGCTTGTCCTAAAGATCCAGCCAAACCTCCGGCCAAGGCACAAGTTGTGGGATGGCCACCGGT GAGATCATACCGGAAGAACGTGATGGTTTCCTGCCAAAAATCAAGCGGTGGCCCGGAGGCGGCGGCGTTCGTGA AGGTATCAATGGACGGAGCACCGTACTTGAGGAAAATCGATTTGAGGATGTATAAA

Miscellaneous	
oriP	ACTGAACCCTAAACGGGTAGCATATGCTTCCCGGGTAGTAGTATATACTATCCAGACTAACCCTAATTCAATAGCA TATGTTACCCAACGGGAAGCATATGCTATCGAATTAGGGTTAGTAAAAGGGTCCTAAGGAACAGCGATGTAGGT GGGCGGGCCAAGATAGGGGCGCGATTGCTGCGATCTGGAGGACAAATTACACACACTTGCGCCTGAGCGCCAA GCACAGGGTTGTTGGTCCTCATATTCACGAGGTCGCTGAGAGCACGGTGGGCTAATGTTGCCATGGGTAGCATAT ACTACCCAAATATCTGGATAGCATATGCTATCCTAATCTATATCTGGGTAGCATAGGCTATCCTAATCTATATCTGG GTAGCATATGCTATCCTAATCTATATCTGGGTAGTATATGCTATCCTAATTTATATCTGGGTAGCATAGGCTATCCT AATCTATATCTGGGTAGCATATGCTATCCTAATCTATATCTGGGTAGTATATGCTATCCTAATCTGTATCCGGGTAG CATATGCTATCCTAATAGAGATTAGGGTAGTATATGCTATCCTAATTTATATCTGGGTAGCATATACTACCCAAATA TCTGGATAGCATATGCTATCCTAATCTATATCTGGGTAGCATATGCTATCCTAATCTATATCTGGGTAGCATAGGCT ATCCTAATCTATATCTGGGTAGCATATGCTATCCTAATCTATATCTGGGTAGTATATGCTATCCTAATTTATATCTG GGTAGCATAGGCTATCCTAATCTATATCTGGGTAGCATATGCTATCCTAATCTATATCTGGGTAGTATATGCTATCC TAATCTGTATCCGGGTAGCATATGCTATCCTCATGAT
IRES2	CCCCTCTCCCTCCCCCCCCCCTAACGTTACTGGCCGAAGCCGCTTGGAATAAGGCCGGTGTGCGTTTGTCTATATG TTATTTTCCACCATATTGCCGTCTTTTGGCAATGTGAGGGCCCGGAAACCTGGCCCTGTCTTCTTGACGAGCATTCC TAGGGGTCTTTCCCCTCTCGCCAAAGGAATGCAAGGTCTGTTGAATGTCGTGAAGGAAGCAGTTCCTCTGGAAGC TTCTTGAAGACAAACAACGTCTGTAGCGACCCTTTGCAGGCAGCGGAACCCCCCACCTGGCGACAGGTGCCTCTG CGGCCAAAAGCCACGTGTATAAGATACACCTGCAAAGGCGGCACAACCCCAGTGCCACGTTGTGAGTTGGATAG TTGTGGAAAGAGTCAAATGGCTCTCCTCAAGCGTATTCAACAAGGGGCTGAAGGATGCCCAGAAGGTACCCCATT GTATGGGATCTGATCTGGGGCCTCGGTGCACATGCTTTACATGTGTTTAGTCGAGGTTAAAAAAACGTCTAGGCC CCCCGAACCACGGGGACGTGGTTTTCCTTTGAAAAACACGATGATAATATGGCCACAACC
IgKL	GAAACAGACACACTGCTGCTATGGGTACTGCTGCTCTGGGTTCCAGGTTCCACTGGTGACAGCAGCACGCTAGCC GCCACC
KDEL	AAGGACGAGCTG
P2A	GGAAGCGGAGCGACCAACTTTAGCCTGCTGAAACAGGCGGGCGATGTGGAAGAAAACCCTGGACCT
T2A	GGAAGCGGAGAGGGCAGAGGAAGTCTGCTAACATGCGGTGACGTCGAGGAGAATCCTGGCCCA
Hairpin 1 (H1)	GGTTTGGCTTATTCGCCACC
Hairpin 2 (H2)	CATGGTGGCTTCGCCACCATG
Hairpin 3 (H3)	CTTGCTCACTCCCATGGTGGCTTCGCCACCATGGGAGTGAGCAAG
attP (BxBI)	GTGGTTTGTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCCA
attB (BxBI)	TCCGGGCTTGTCGACGACGGCGGTCTCCGTCGTCAGGATCAT
5' HA - AAVS1	ACTAGGGACAGGATTGGTGACAGAAAAGCCCCATCCTTAGGCCTCCTCCTTCCTAGTCTCCTGATATTGGGTCTAA CCCCCACCTCCTGTTAGGCAGATTCCTTATCTGGTGACACACCCCCATTTCCTGGAGCCATCTCTCTCCTTGCCAGA ACCTCTAAGGTTTGCTTACGATGGAGCCAGAGAGGATCCTGGGAGGGAGAGCTTGGCAGGGGGTGGGAGGGAA GGGGGGGATGCGTGACCTGCCCGGTTCTCAGTGGCCACCCTGCGCTACCCTCTCCCAGAACCTGAGCTGCTCTGA CGCGGCTGTCTGGTGCGTTTCACTGATCCTGGTGCTGCAGCTTCCTTACACTTCCCAAGAGGAGAAGCAGTTTGG AAAAACAAAATCAGAATAAGTTGGTCCTGAGTTCTAACTTTGGCTCTTCACCTTTCTAGTCCCCAATTTATATTGTT CCTCCGTGCGTCAGTTTTACCTGTGAGATAAGGCCAGTAGCCAGCCCCGTCCTGGCAGGGCTGTGGTGAGGAGG GGGGTGTCCGTGTGGAAAACTCCCTTTGTGAGAATGGTGCGTCCTAGGTGTTCACCAGGTCGTGGCCGCCTCTAC TCCCTTTCTCTTTCTCCATCCTTCTTTCCTTAAAGAGTCCCCAGTGCTATCTGGGACATATTCCTCCGCCCAGAGCAG GGTCCCGCTTCCCTAAGGCCCTGCTCTGGGCTTCTGGGTTTGAGTCCTTGGCAAGCCCAGGAGAGGCGCTCAGGC TTCCCTGTCCCCCTTCCTCGTCCACCATCTCATGCCCCTGGCTCTCCTGCCCCTTCCCTACAGGGGTTCCTGGCTCTG CTCT
3' HA - AAVS1	TGCTTTCTCTGACCAGCATTCTCTCCCCTGGGCCTGTGCCGCTTTCTGTCTGCAGCTTGTGGCCTGGGTCACCTCTA CGGCTGGCCCAGATCCTTCCCTGCCGCCTCCTTCAGGTTCCGTCTTCCTCCACTCCCTCTTCCCCTTGCTCTCTGCTG TGTTGCTGCCCAAGGATGCTCTTTCCGGAGCACTTCCTTCTCGGCGCTGCACCACGTGATGTCCTCTGAGCGGATC CTCCCCGTGTCTGGGTCCTCTCCGGGCATCTCTCCTCCCTCACCCAACCCCATGCCGTCTTCACTCGCTGGGTTCCC TTTTCCTTCTCCTTCTGGGGCCTGTGCCATCTCTCGTTTCTTAGGATGGCCTTCTCCGACGGATGTCTCCCTTGCGTC CCGCCTCCCCTTCTTGTAGGCCTGCATCATCACCGTTTTTCTGGACAACCCCAAAGTACCCCGTCTTCCTGGCTTTA GCCACCTCTCCATCCTCTTGCTTTCTTTGCCTGGACACCCCGTTCTCCTGTGGATTCGGGTCACCTCTCACTCCTTTC ATTTGGGCAGCTCCCCTACCCCCCTTACCTCTCTAGTCTGTGCTAGCTCTTCCAGCCCCCTGTCATGGCATCTTCCAG

	GGGTCCGAGAGCTCAGCTAGTCTTCTTCCTCCAACCCGGGCCCCTATGTCCACTTCAGGACAGCATGTTTGCTGCC TCCAGGGATCCTGTGTCCCCGAGCTGGGACCACCTTATATTCCCAGGGCCGGTTAATGTGGCTCTGGTTCTGGGT ACTTTTATCTGTCCCCTCCACCCCACAGTGGGGC
gRNA AAVS1	GGGGCCACTAGGGACAGGAT
Insulator	GGCCGCGAATTCTGAAAGACCCCACCTGTAGGTTTGGCAAGCCCAGGGATGTACGTCCCTAACCCGCTAGGGGG CAGCAACTAGTCCCAGGCCTGCACTGCCGCCTGCCGGCAGGGGTCCAGTCGCTAGCGCATGCCTGCA

Chapter 9 - Bibliography

f1561. Kitada, T., DiAndreth, B., Teague, B. & Weiss, R. Programming gene and engineered-cell therapies with synthetic biology. *Science (80-.).* **359**, (2018).

2. Morton, S., Pencheon, D. & Squires, N. Sustainable Development Goals (SDGs), and their implementation. *Br. Med. Bull.* **124**, 81–90 (2017).

3. Bashor, C. J. & Hilton, I. B. Engineering the next generation of cell-based therapeutics. **21**, 655–675 (2022).

4. Fischbach, M. A., Bluestone, J. A. & Lim, W. A. Cell-based therapeutics: The next pillar of medicine. *Science Translational Medicine* **5**, 1–7 (2013).

5. Scheller, L. & Fussenegger, M. From synthetic biology to human therapy : engineered mammalian cells. *Curr. Opin. Biotechnol.* **58**, 108–116 (2019).

6. June, C. H., O'Connor, R. S., Kawalekar, O. U., Ghassemi, S. & Milone, M. C. CAR T cell immunotherapy for human cancer. *Science (80-.).* **359**, 1361–1365 (2018).

7. Tomasik, J., Jasiński, M. & Basak, G. W. Next generations of CAR-T cells - new therapeutic opportunities in hematology? *Front. Immunol.* **13**, 1–13 (2022).

8. European Clinical Trials RegisterAbout, C. Clinical Trials register - Search for CAR T Cells. (2020).

9. Irvine, D. J., Maus, M. V, Mooney, D. J. & Wong, W. W. The future of engineered immune cell therapies. *Science* **378**, 853–858 (2022).

10. Ellebrecht, C. T. *et al.* Reengineering chimeric antigen receptor T cells for targeted therapy of autoimmune disease. **353**, (2016).

11. Reighard, S. D. *et al.* Therapeutic Targeting of Follicular T Cells with Chimeric Antigen Receptor-Expressing Natural Killer Cells ll Therapeutic Targeting of Follicular T Cells with Chimeric Antigen Receptor-Expressing Natural Killer Cells. *Cell Reports Med.* **1**, 100003 (2020).

12. Arya, Y. & Sahi, P. K. Cell-Based Gene Therapy for b-Thalassemia. *Indian Pediatr.* **60**, 313–316 (2023).

13. Yamanaka, S. Pluripotent Stem Cell-Based Cell Therapy—Promise and Challenges. *Cell Stem Cell* **27**, 523–531 (2020).

14. Kim, J. Y., Nam, Y., Rim, Y. A. & Ju, J. H. Review of the Current Trends in Clinical Trials Involving Induced Pluripotent Stem Cells. *Stem Cell Rev. Reports* **18**, 142–154 (2022).

15. McNerney, M. P., Doiron, K. E., Ng, T. L., Chang, T. Z. & Silver, P. A. Theranostic cells: emerging clinical applications of synthetic biology. *Nature Reviews Genetics* **22**, 730–746 (2021).

16. Choi, Y. R. *et al.* A genome-engineered bioartificial implant for autoregulated anticytokine drug delivery. *Sci. Adv.* **7**, 1–13 (2021).

17. Smole, A., Lainšček, D., Bezeljak, U., Horvat, S. & Jerala, R. A Synthetic Mammalian Therapeutic Gene Circuit for Sensing and Suppressing Inflammation. *Mol. Ther.* **25**, 102–119 (2017).

18. Chassin, H. *et al.* Sensing and responding to allergic response cytokines through a genetically encoded circuit. *Nat. Commun.* **8**, 1–13 (2017).
19. Schukur, L., Geering, B., Charpin-El Hamri, G. & Fussenegger, M. Implantable synthetic cytokine converter cells with AND-gate logic treat experimental psoriasis. *Sci. Transl. Med.* **7**, (2015).
20. Ye, H. *et al.* Self-adjusting synthetic gene circuit for correcting insulin resistance. *Nat. Biomed. Eng.* **1**, 1–9 (2017).
21. Qudrat, A., Mosabbir, A. Al & Truong, K. Engineered Proteins Program Mammalian Cells to Target Inflammatory Disease Sites. *Cell Chem. Biol.* **24**, 703-711.e2 (2017).
22. Smole, A., Lainšček, D., Bezeljak, U., Horvat, S. & Jerala, R. A Synthetic Mammalian Therapeutic Gene Circuit for Sensing and Suppressing Inflammation. *Mol. Ther.* **25**, 102–119 (2017).
23. Kim, J. Y., Ahn, Y. J., Lee, J. A. & Lee, S. Y. Recent advances in the production of platform chemicals using metabolically engineered microorganisms. *Current Opinion in Green and Sustainable Chemistry* **40**, 100777 (2023).
24. Ro, D. K. *et al.* Production of the antimalarial drug precursor artemisinic acid in engineered yeast. *Nature* **440**, 940–943 (2006).
25. Li, Y. *et al.* Complete biosynthesis of noscapine and halogenated alkaloids in yeast. *Proc. Natl. Acad. Sci. U. S. A.* **115**, E3922–E3931 (2018).
26. Luo, X. *et al.* Complete biosynthesis of cannabinoids and their unnatural analogues in yeast. *Nature* **567**, 123–126 (2019).
27. Thodey, K., Galanie, S. & Smolke, C. D. A microbial biomanufacturing platform for natural and semisynthetic opioids. *Nat. Chem. Biol.* **10**, 837–844 (2014).
28. Srinivasan, P. & Smolke, C. D. Biosynthesis of medicinal tropane alkaloids in yeast. *Nature* **585**, 614–619 (2020).
29. Zhang, J. *et al.* A microbial supply chain for production of the anti-cancer drug vinblastine. *Nature* **609**, 341–347 (2022).
30. O'Flaherty, R. *et al.* Mammalian cell culture for production of recombinant proteins: A review of the critical steps in their biomanufacturing. *Biotechnology Advances* **43**, 107552 (2020).
31. Tripathi, N. K. & Shrivastava, A. Recent Developments in Bioprocessing of Recombinant Proteins: Expression Hosts and Process Development. *Frontiers in Bioengineering and Biotechnology* **7**, (2019).
32. El Abd, Y., Tabll, A., Smolic, R. & Smolic, M. Mini-review: The market growth of diagnostic and therapeutic monoclonal antibodies - SARS CoV-2 as an example. *Hum. Antibodies* **30**, 15–24 (2022).
33. Cameron, D. E., Bashor, C. J. & Collins, J. J. A brief history of synthetic biology. *Nat. Rev. Microbiol.* **12**, 381–390 (2014).
34. Andrianantoandro, E., Basu, S., Karig, D. K. & Weiss, R. Synthetic biology: New engineering rules for an emerging discipline. *Mol. Syst. Biol.* **2**, 1–14 (2006).

35. Elowitz. A synthetic oscillatory network repressilator. *Nature* **403**, 335–338 (2000).
36. Gardner, T. S., Cantor, C. R. & Collins, J. J. Construction of a genetic toggle switch in Escherichia coli. *Nature* **403**, 339–342 (2000).
37. Meng, F. & Ellis, T. The second decade of synthetic biology: 2010–2020. *Nature Communications* **11**, 1–4 (2020).
38. Kwok, R. Five hard truths for synthetic biology. *Nature* **463**, (2010).
39. Decoene, T. *et al.* Standardization in synthetic biology: an engineering discipline coming of age. *Crit. Rev. Biotechnol.* **38**, 647–656 (2018).
40. Elowitz, M. B., Levine, A. J., Siggia, E. D. & Swain, P. S. Stochastic gene expression in a single cell. *Science (80-.).* **297**, 1183–1186 (2002).
41. Balázsi, G., Van Oudenaarden, A. & Collins, J. J. Cellular decision making and biological noise: From microbes to mammals. *Cell* **144**, 910–925 (2011).
42. Borkowski, O., Ceroni, F., Stan, G. B. & Ellis, T. Overloaded and stressed: whole-cell considerations for bacterial synthetic biology. *Curr. Opin. Microbiol.* **33**, 123–130 (2016).
43. Kelly, J. R. *et al.* Measuring the activity of BioBrick promoters using an in vivo reference standard. *J. Biol. Eng.* **3**, 1–13 (2009).
44. de Paiva, D. P. *et al.* A study on the use of strain-specific and homologous promoters for heterologous expression in industrial Saccharomyces cerevisiae strains. *AMB Express* **8**, 1–11 (2018).
45. Qin, J. Y. *et al.* Systematic comparison of constitutive promoters and the doxycycline-inducible promoter. *PLoS One* **5**, 3–6 (2010).
46. Ede, C., Chen, X., Lin, M. Y. & Chen, Y. Y. Quantitative Analyses of Core Promoters Enable Precise Engineering of Regulated Gene Expression in Mammalian Cells. *ACS Synth. Biol.* **5**, 395–404 (2016).
47. Fonseca, J. P. *et al.* A Toolkit for Rapid Modular Construction of Biological Circuits in Mammalian Cells. *ACS Synth. Biol.* **8**, 2593–2606 (2019).
48. Jones, R. D. *et al.* An endoribonuclease-based feedforward controller for decoupling resource-limited genetic modules in mammalian cells. *Nat. Commun.* **11**, 1–16 (2020).
49. Keren, L. *et al.* Promoters maintain their relative activity levels under different growth conditions. *Mol. Syst. Biol.* **9**, 1–17 (2013).
50. Rudge, T. J. *et al.* Characterization of Intrinsic Properties of Promoters. *ACS Synth. Biol.* **5**, 89–98 (2016).
51. Sauer, C. *et al.* Effect of Genome Position on Heterologous Gene Expression in Bacillus subtilis: An Unbiased Analysis. *ACS Synth. Biol.* **5**, 942–947 (2016).
52. Englaender, J. A. *et al.* Effect of Genomic Integration Location on Heterologous Protein Expression and Metabolic Engineering in E. coli. *ACS Synth. Biol.* **6**, 710–720 (2017).
53. Sadelain, M., Papapetrou, E. P. & Bushman, F. D. Safe harbours for the integration of new DNA in the

human genome. *Nat. Rev. Cancer* **12**, 51–58 (2012).

54. Shin, S. *et al.* Comprehensive Analysis of Genomic Safe Harbors as Target Sites for Stable Expression of the Heterologous Gene in HEK293 Cells. *ACS Synth. Biol.* **9**, 1263–1269 (2020).
55. Cabrera, A. *et al.* The sound of silence: Transgene silencing in mammalian cell engineering. *Cell Syst.* **13**, 950–973 (2022).
56. Wang, B., Kitney, R. I., Joly, N. & Buck, M. Engineering modular and orthogonal logic gates for robust digital-like synthetic biology. *Nat. Commun.* **2**, 508–509 (2011).
57. Cardinale, S., Joachimiak, M. P. & Arkin, A. P. Effects of genetic variation on the e. coli host-circuit interface. *Cell Rep.* **4**, 231–237 (2013).
58. Moser, F. *et al.* Genetic circuit performance under conditions relevant for industrial bioreactors. *ACS Synth. Biol.* **1**, 555–564 (2012).
59. Tas, H., Grozinger, L., Stoof, R., de Lorenzo, V. & Goñi-Moreno, Á. Contextual dependencies expand the re-usability of genetic inverters. *Nat. Commun.* **12**, 1–9 (2021).
60. Klumpp, S., Zhang, Z. & Hwa, T. Growth Rate-Dependent Global Effects on Gene Expression in Bacteria. *Cell* **139**, 1366–1375 (2009).
61. 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).
62. Tan, C., Marguet, P. & You, L. Emergent bistability by a growth-modulating positive feedback circuit. *Nat. Chem. Biol.* **5**, 842–848 (2009).
63. Kintaka, R. *et al.* Genetic profiling of protein burden and nuclear export overload. *Elife* **9**, 1–22 (2020).
64. Bragdon, M. D. J. *et al.* Cooperative assembly confers regulatory specificity and long-term genetic circuit stability. *bioRxiv* 2022.05.22.492993 (2022).
65. Kallehauge, T. B. *et al.* Ribosome profiling-guided depletion of an mRNA increases cell growth rate and protein secretion. *Sci. Rep.* **7**, 1–12 (2017).
66. Kol, S. *et al.* Multiplex secretome engineering enhances recombinant protein production and purity. *Nat. Commun.* **11**, (2020).
67. Rugbjerg, P. & Sommer, M. O. A. Overcoming genetic heterogeneity in industrial fermentations. *Nat. Biotechnol.* **37**, 869–876 (2019).
68. Gyorgy, A. *et al.* Isocost Lines Describe the Cellular Economy of Genetic Circuits. *Biophys. J.* **109**, 639–646 (2015).
69. Cookson, N. A. *et al.* Queueing up for enzymatic processing: Correlated signaling through coupled degradation. *Mol. Syst. Biol.* **7**, (2011).
70. Carrillo, S. S., Galez, H., Napolitano, S., Bertaux, F. & Batt, G. Maximizing protein production by keeping cells at optimal secretory stress levels using real - time control approaches. *Nat. Commun.*

(2023). doi:10.1038/s41467-023-38807-9

71. Cella, F. *et al.* MIRELLA: a mathematical model explains the effect of microRNA-mediated synthetic genes regulation on intracellular resource allocation. *Nucleic Acids Res.* **51**, 3452–3464 (2023).
72. Frei, T. *et al.* Characterization and mitigation of gene expression burden in mammalian cells. *Nat. Commun.* **11**, 1–14 (2020).
73. Deris, J. B. *et al.* The innate growth bistability and fitness landscapes of antibiotic-resistant bacteria. *Science (80-.).* **342**, (2013).
74. Nevozhay, D., Adams, R. M., van Itallie, E., Bennett, M. R. & Balázsi, G. Mapping the environmental fitness landscape of a synthetic gene circuit. *PLoS Comput. Biol.* **8**, (2012).
75. Zhang, R. *et al.* Topology-dependent interference of synthetic gene circuit function by growth feedback. *Nat. Chem. Biol.* **16**, (2020).
76. Kong, L., Shi, W., Tian, X. & Lai, Y. Effects of growth feedback on gene circuits : A dynamical understanding. *bioRxiv* 1–35 (2023).
77. Zhang, R. *et al.* Winner-takes-all resource competition redirects cascading cell fate transitions. *Nat. Commun.* **12**, 1–9 (2021).
78. Prindle, A. *et al.* Rapid and tunable post-translational coupling of genetic circuits. *Nature* **508**, 387–391 (2014).
79. Di Blasi, R., Marbiah, M. M., Siciliano, V., Polizzi, K. & Ceroni, F. A call for caution in analysing mammalian co-transfection experiments and implications of resource competition in data misinterpretation. *Nat. Commun.* **12**, 1–6 (2021).
80. Chong, Z. X., Yeap, S. K. & Ho, W. Y. Transfection types, methods and strategies: A technical review. *PeerJ* **9**, 1–37 (2021).
81. Farr, A. & Roman, A. A pitfall of using a second plasmid to determine transfection efficiency. *Nucleic Acids Research* **20**, 920 (1992).
82. Qin, C. *et al.* Precise programming of multigene expression stoichiometry in mammalian cells by a modular and programmable transcriptional system. *Nat. Commun.* **14**, 1–10 (2023).
83. Mehta, D., Chirmade, T., Tungekar, A. A., Gani, K. & Bhambure, R. Cloning and expression of antibody fragment (Fab) I: Effect of expression construct and induction strategies on light and heavy chain gene expression. *Biochem. Eng. J.* **176**, 108189 (2021).
84. Srivastava, A., Mallela, K. M. G., Deorkar, N. & Brophy, G. Manufacturing Challenges and Rational Formulation Development for AAV Viral Vectors. *J. Pharm. Sci.* **110**, 2609–2624 (2021).
85. Shachrai, I., Zaslaver, A., Alon, U. & Dekel, E. Cost of Unneeded Proteins in E. coli Is Reduced after Several Generations in Exponential Growth. *Mol. Cell* **38**, 758–767 (2010).
86. Cardinale, S., Joachimiak, M. P. & Arkin, A. P. Effects of genetic variation on the e. coli host-circuit interface. *Cell Rep.* **4**, 231–237 (2013).

87. Gorochowski, T. E., Van Den Berg, E., Kerkman, R., Roubos, J. A. & Bovenberg, R. A. L. Using synthetic biological parts and microbioreactors to explore the protein expression characteristics of escherichia coli. *ACS Synth. Biol.* **3**, 129–139 (2014).

88. Scott, M., Gunderson, C. W., Mateescu, E. M., Zhang, Z. & Hwa, T. Interdependence of Cell Growth and Gene Expression: Origin and Consequences. *Science (80-.).* **330**, 1099–1102 (2010).

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

90. Du, Z. *et al.* Non-Invasive UPR Monitoring System and Its Applications in CHO Production Cultures. *Biotechnol. Bioeng.* **xxx**, 1–11 (2013).

91. Roy, G. *et al.* Development of a fluorescent reporter system for monitoring ER stress in Chinese hamster ovary cells and its application for therapeutic protein production. *PLoS One* 1–20 (2017).

92. Cedras, G., Kroukamp, H., Van Zyl, W. H. & Den Haan, R. The in vivo detection and measurement of the unfolded protein response in recombinant cellulase producing Saccharomyces cerevisiae strains. *Biotechnol. Appl. Biochem.* **67**, 82–94 (2020).

93. Merksamer, P. I., Trusina, A. & Papa, F. R. Real-Time Redox Measurements during Endoplasmic Reticulum Stress Reveal Interlinked Protein Folding Functions. *Cell* **135**, 933–947 (2008).

94. Hetz, C., Zhang, K. & Kaufman, R. J. Mechanisms, regulation and functions of the unfolded protein response. *Nat. Rev. Mol. Cell Biol.* **21**, 421–438 (2020).

95. Tabara, K., Iwata, Y. & Koizumi, N. The unfolded protein response: from stress pathway to homeostatic regulation. *Methods Mol. Biol.* **1691**, 223–230 (2018).

96. Cartwright, J. F., Anderson, K., Longworth, J., Lobb, P. & James, D. C. Highly sensitive detection of mutations in CHO cell recombinant DNA using multi-parallel single molecule real-time DNA sequencing. *Biotechnol. Bioeng.* **115**, 1485–1498 (2018).

97. Newbert, R. W., Barton, B., Greaves, P., Harper, J. & Turner, G. Analysis of a commercially improved Penicillium chrysogenum strain series: Involvement of recombinogenic regions in amplification and deletion of the penicillin biosynthesis gene cluster. *J. Ind. Microbiol. Biotechnol.* **19**, 18–27 (1997).

98. Yoshikawa, T. *et al.* Amplified gene location in chromosomal DNA affected recombinant protein production and stability of amplified genes. *Biotechnol. Prog.* **16**, 710–715 (2000).

99. Schlegel, S., Genevaux, P. & de Gier, J. W. De-convoluting the Genetic Adaptations of E.coli C41(DE3) in Real Time Reveals How Alleviating Protein Production Stress Improves Yields. *Cell Rep.* **10**, 1758–1766 (2015).

100. Rugbjerg, P., Myling-petersen, N., Porse, A., Sarup-lytzen, K. & Sommer, M. O. A. Diverse genetic error modes constrain large-scale bio-based production. *Nat. Commun.* doi:10.1038/s41467-018-03232-w

101. Choi, J. W., Yim, S. S., Kim, M. J. & Jeong, K. J. Enhanced production of recombinant proteins with

Corynebacterium glutamicum by deletion of insertion sequences (IS elements). *Microb. Cell Fact.* **14**, 1–12 (2015).

102. Burgard, A., Burk, M. J., Osterhout, R., Van Dien, S. & Yim, H. Development of a commercial scale process for production of 1,4-butanediol from sugar. *Curr. Opin. Biotechnol.* **42**, 118–125 (2016).
103. Pósfai, G., Umenhoffer, K., Kolisnychenko, V., Stahl, B. & Sharma, S. S. Emergent Properties of Reduced-Genome Escherichia coli. *Science (80-.).* **1044**, 1044–1047 (2013).
104. Richardson, S. M. *et al.* Design of a synthetic yeast genome. *Science (80-.).* **355**, 1040–1044 (2017).
105. Kurnit, D. M. Escherichia coli recA deletion strains that are highly competent for transformation and for in vivo phage packaging. *Gene* **82**, 313–315 (1989).
106. Csörgo, B., Fehér, T., Tímár, E., Blattner, F. R. & Pósfai, G. Low-mutation-rate, reduced-genome Escherichia coli: An improved host for faithful maintenance of engineered genetic constructs. *Microb. Cell Fact.* **11**, 1–13 (2012).
107. Grob, A., Di Blasi, R. & Ceroni, F. Experimental tools to reduce the burden of bacterial synthetic biology. *Curr. Opin. Syst. Biol.* **28**, 100393 (2021).
108. Xu, X. *et al.* Trimming the genomic fat: minimising and re-functionalising genomes using synthetic biology. *Nat. Commun.* **14**, 1984 (2023).
109. Mizoguchi, H., Mori, H. & Fujio, T. Escherichia coli minimum genome factory. *Biotechnol. Appl. Biochem.* **46**, 157 (2007).
110. Komatsu, M., Uchiyama, T., Omura, S., Cane, D. E. & Ikeda, H. Genome-minimized Streptomyces host for the heterologous expression of secondary metabolism. *Proc. Natl. Acad. Sci. U. S. A.* **107**, 2646–2651 (2010).
111. Suárez, R. A., Stülke, J. & Van Dijl, J. M. Less is More: Toward a Genome-Reduced Bacillus Cell Factory for 'difficult Proteins'. *ACS Synth. Biol.* **8**, 99–108 (2019).
112. Lastiri-Pancardo, G., Mercado-Hernández, J. S., Kim, J., Jiménez, J. I. & Utrilla, J. A quantitative method for proteome reallocation using minimal regulatory interventions. *Nat. Chem. Biol.* **16**, 1026–1033 (2020).
113. Gutierrez, J. M. *et al.* Genome-scale reconstructions of the mammalian secretory pathway predict metabolic costs and limitations of protein secretion. *Nat. Commun.* **11**, 1–10 (2020).
114. Strain, B., Morrissey, J., Antonakoudis, A. & Kontoravdi, C. Genome-scale models as a vehicle for knowledge transfer from microbial to mammalian cell systems. *Comput. Struct. Biotechnol. J.* **21**, 1543–1549 (2023).
115. Atkinson, E., Tuza, Z., Perrino, G., Stan, G. B. & Ledesma-Amaro, R. Resource-aware whole-cell model of division of labour in a microbial consortium for complex-substrate degradation. *Microb. Cell Fact.* **21**, 1–12 (2022).
116. Zhang, H., Pereira, B., Li, Z., Stephanopoulos, G. & Demain, A. L. Engineering Escherichia coli

coculture systems for the production of biochemical products. *Proc. Natl. Acad. Sci. U. S. A.* **112**, 8266–8271 (2015).

117. Zhou, K., Qiao, K., Edgar, S. & Stephanopoulos, G. Distributing a metabolic pathway among a microbial consortium enhances production of natural products. *Nat. Biotechnol.* **33**, 377–383 (2015).
118. Jones, J. A. *et al.* Experimental and computational optimization of an Escherichia coli co-culture for the efficient production of flavonoids. *Metab. Eng.* **35**, 55–63 (2016).
119. Fang, Z., Jones, J. A., Zhou, J. & Koffas, M. A. G. Engineering Escherichia coli Co-Cultures for Production of Curcuminoids From Glucose. *Biotechnol. J.* **13**, (2018).
120. Jones, J. A. *et al.* Complete Biosynthesis of Anthocyanins Using E- coli Polycultures. *MBio* **8**, 1–9 (2017).
121. Aulakh, S. K., Vidal, L. S., South, E. J., Peng, H. & Varma, S. J. Spontaneously establishing syntrophic yeast communities improve bioproduction. (2023). doi:10.1038/s41589-023-01341-2
122. Tsoi, R. *et al.* Metabolic division of labor in microbial systems. *Proc. Natl. Acad. Sci. U. S. A.* **115**, 2526–2531 (2018).
123. Roell, G. W. *et al.* Engineering microbial consortia by division of labor. *Microb. Cell Fact.* **18**, 1–11 (2019).
124. Barajas, C. & Del Vecchio, D. Synthetic biology by controller design. *Curr. Opin. Biotechnol.* **78**, 102837 (2022).
125. El-Samad, H., Goff, J. P. & Khammash, M. Calcium homeostasis and parturient hypocalcemia: An integral feedback perspective. *J. Theor. Biol.* **214**, 17–29 (2002).
126. Muzzey, D., Gómez-Uribe, C. A., Mettetal, J. T. & van Oudenaarden, A. A Systems-Level Analysis of Perfect Adaptation in Yeast Osmoregulation. *Cell* **138**, 160–171 (2009).
127. Shopera, T., He, L., Oyetunde, T., Tang, Y. J. & Moon, T. S. Decoupling Resource-Coupled Gene Expression in Living Cells. *ACS Synth. Biol.* **6**, 1596–1604 (2017).
128. Briat, C., Zechner, C. & Khammash, M. Design of a Synthetic Integral Feedback Circuit: Dynamic Analysis and DNA Implementation. *ACS Synth. Biol.* **5**, 1108–1116 (2016).
129. Khammash, M. H. Perfect adaptation in biology. *Cell Syst.* **12**, 509–521 (2021).
130. Aoki, S. K. *et al.* A universal biomolecular integral feedback controller for robust perfect adaptation. *Nature* **570**, 533–537 (2019).
131. Huang, H. H., Qian, Y. & Del Vecchio, D. A quasi-integral controller for adaptation of genetic modules to variable ribosome demand. *Nat. Commun.* **9**, 1–12 (2018).
132. Frei, T., Chang, C. H., Filo, M., Arampatzis, A. & Khammash, M. A genetic mammalian proportional-integral feedback control circuit for robust and precise gene regulation. *Proc. Natl. Acad. Sci. U. S. A.* **119**, 1–12 (2022).
133. Lillacci, G., Benenson, Y. & Khammash, M. Synthetic control systems for high performance gene

expression in mammalian cells. *Nucleic Acids Res.* **46**, 9855–9863 (2018).

134. Liu, C. C., Jewett, M. C., Chin, J. W. & Voigt, C. A. Toward an orthogonal central dogma. *Nat. Chem. Biol.* **14**, 103–106 (2018).
135. Ravikumar, A., Arrieta, A. & Liu, C. C. An orthogonal DNA replication system in yeast. *Nat. Chem. Biol.* **10**, 175–177 (2014).
136. Becskei Attila & Serrano Luis. Engineering stability in gene networks by autoregulation. *Nature* **405**, 590–593 (2000).
137. Aleksashin, N. A. *et al.* A fully orthogonal system for protein synthesis in bacterial cells. *Nat. Commun.* **11**, 1–11 (2020).
138. Orelle, C. *et al.* Protein synthesis by ribosomes with tethered subunits. *Nature* **524**, 119–124 (2015).
139. Carlson, E. D. *et al.* Engineered ribosomes with tethered subunits for expanding biological function. *Nat. Commun.* **10**, 1–13 (2019).
140. Aleksashin, N. A. *et al.* Assembly and functionality of the ribosome with tethered subunits. *Nat. Commun.* **10**, (2019).
141. Schmied, W. H. *et al.* Controlling orthogonal ribosome subunit interactions enables evolution of new function. *Nature* **564**, 444–448 (2018).
142. Cameron, D. E. & Collins, J. J. Tunable protein degradation in bacteria. *Nat. Biotechnol.* **32**, 1276–1281 (2014).
143. Darlington, A. P. S., Kim, J., Jiménez, J. I. & Bates, D. G. Dynamic allocation of orthogonal ribosomes facilitates uncoupling of co-expressed genes. *Nat. Commun.* **9**, (2018).
144. An, W. & Chin, J. W. Synthesis of orthogonal transcription-translation networks. *Proc. Natl. Acad. Sci. U. S. A.* **106**, 8477–8482 (2009).
145. Darlington, A. P. S. & Bates, D. G. Architectures for Combined Transcriptional and Translational Resource Allocation Controllers. *Cell Syst.* **11**, 382-392.e9 (2020).
146. Rugbjerg, P., Sarup-Lytzen, K., Nagy, M. & Sommer, M. O. A. Synthetic addiction extends the productive life time of engineered Escherichia coli populations . *Proc. Natl. Acad. Sci.* **115**, 201718622 (2018).
147. D'Ambrosio, V. *et al.* Regulatory control circuits for stabilizing long-term anabolic product formation in yeast. *Metab. Eng.* **61**, 369–380 (2020).
148. Lee, S. W., Rugbjerg, P. & Sommer, M. O. A. Exploring Selective Pressure Trade-Offs for Synthetic Addiction to Extend Metabolite Productive Lifetimes in Yeast. *ACS Synth. Biol.* **10**, 2842–2849 (2021).
149. Nielsen, A. A. K. *et al.* Genetic circuit design automation. *Science (80-.).* **352**, (2016).
150. O'Connel, R. W. *et al.* Ultra-high throughput mapping of genetic design space. *bioRxiv* (2023).
151. Di Blasi, R. *et al.* Resource-aware construct design in mammalian cells. *Nat. Commun.* https://doi.org/10.1038/s41467-023-39252-4 (2023). doi:10.1038/s41467-023-39252-4

152. Matreyek, K. A., Stephany, J. J. & Fowler, D. M. A platform for functional assessment of large variant libraries in mammalian cells. *Nucleic Acids Res.* **45**, 1–12 (2017).

153. Haberle, V. & Stark, A. Eukaryotic core promoters and the functional basis of transcription initiation. *Nat. Rev. Mol. Cell Biol.* **19**, 621–637 (2018).

154. Kozak, M. Point mutations close to the AUG initiator codon affect the efficiency of translation of rat preproinsulin in vivo. *Nature* **308**, 241–246 (1984).

155. Proudfoot, N. J. Ending the message: Poly(A) signals then and now. *Genes Dev.* **25**, 1770–1782 (2011).

156. Jalkanen, A. L., Coleman, S. J. & Wilusz, J. Determinants and Implications of mRNA Poly(A) Tail Size - Does this. *Semin Cell Dev Biol.* **0**, 24–32 (2014).

157. Wang, X. Y. *et al.* Enhanced Transgene Expression by Optimization of Poly A in Transfected CHO Cells. *Front. Bioeng. Biotechnol.* **10**, 1–9 (2022).

158. Noderer, W. L. *et al.* Quantitative analysis of mammalian translation initiation sites by FACS-seq. 1–14 (2014).

159. Blanco, N. *et al.* Tailoring translational strength using Kozak sequence variants improves bispecific antibody assembly and reduces product-related impurities in CHO cells. *Biotechnol. Bioeng.* **117**, 1946–1960 (2020).

160. Eisenhut, P. *et al.* Systematic use of synthetic 5'-UTR RNA structures to tune protein translation improves yield and quality of complex proteins in mammalian cell factories. *Nucleic Acids Res.* **48**, 1–16 (2020).

161. Black, J. & Vos, J. M. Establishment of an oriP/EBNA1-based episomal vector transcribing human genomic β-globin in cultured murine fibroblasts. *Gene Ther.* **9**, 1447–1454 (2002).

162. Budge, J. D., Young, R. J. & Smales, C. M. Engineering of Chinese Hamster Ovary Cells With NDPK-A to Enhance DNA Nuclear Delivery Combined With EBNA1 Plasmid Maintenance Gives Improved Exogenous Transient Reporter, mAb and SARS-CoV-2 Spike Protein Expression. *Front. Bioeng. Biotechnol.* **9**, 1–14 (2021).

163. Tom, R., Bisson, L. & Durocher, Y. Culture of HEK293-EBNA1 cells for production of recombinant proteins. *Cold Spring Harb. Protoc.* **3**, 1–5 (2008).

164. Kiyosue, K. & Miwa, Y. Epstein-Barr virus-derived vector suitable for long-term expression in neurons. *Heliyon* **6**, e03504 (2020).

165. Ren, C. *et al.* Establishment and Applications of Epstein-Barr Virus-Based Episomal Vectors in Human Embryonic Stem Cells. *Stem Cells* **24**, 1338–1347 (2006).

166. Nassor, F. *et al.* Long Term Gene Expression in Human Induced Pluripotent Stem Cells and Cerebral Organoids to Model a Neurodegenerative Disease. *Front. Cell. Neurosci.* **14**, 1–7 (2020).

167. Cardinale, S. & Arkin, A. P. Contextualizing context for synthetic biology - identifying causes of failure

of synthetic biological systems. *Biotechnol. J.* **7**, 856–866 (2012).

168. Wang, X. & Marchisio, M. A. Synthetic polycistronic sequences in eukaryotes. *Synth. Syst. Biotechnol.* **6**, 254–261 (2021).
169. Liu, Z. *et al.* Systematic comparison of 2A peptides for cloning multi-genes in a polycistronic vector. *Sci. Rep.* **7**, 1–9 (2017).
170. Koh, E. Y. C. *et al.* An internal ribosome entry site (IRES) mutant library for tuning expression level of multiple genes in mammalian cells. *PLoS One* **8**, (2013).
171. Mizuguchi, H., Xu, Z., Ishii-Watabe, A., Uchida, E. & Hayakawa, T. IRES-Dependent Second Gene Expression Is Significantly Lower Than Cap-Dependent First Gene Expression in a Bicistronic Vector. *Mol. Ther.* **1**, 376–382 (2000).
172. Patel, Y. D. *et al.* Control of Multigene Expression Stoichiometry in Mammalian Cells Using Synthetic Promoters. *ACS Synth. Biol.* **10**, 1155–1165 (2021).
173. Johnstone, C. P. & Galloway, K. E. Supercoiling-mediated feedback rapidly couples and tunes transcription. *Cell Rep.* **41**, 111492 (2022).
174. Kang, K. *et al.* An improved Tet-on system in microRNA overexpression and CRISPR/Cas9-mediated gene editing. *J. Anim. Sci. Biotechnol.* **10**, 1–12 (2019).
175. Gaidukov, L. *et al.* A multi-landing pad DNA integration platform for mammalian cell engineering. *Nucleic Acids Res.* **46**, 4072–4086 (2018).
176. Chi, X., Zheng, Q., Jiang, R., Chen-Tsai, R. Y. & Kong, L. J. A system for site-specific integration of transgenes in mammalian cells. *PLoS One* **14**, 1–14 (2019).
177. Matreyek, K. A., Stephany, J. J., Chiasson, M. A., Hasle, N. & Fowler, D. M. An improved platform for functional assessment of large protein libraries in mammalian cells. *Nucleic Acids Res.* **48**, 1–12 (2020).
178. Jacobsen, L. B., Calvin, S. A. & Lobenhofer, E. K. Transcriptional effects of transfection: The potential for misinterpretation of gene expression data generated from transiently transfected cells. *Biotechniques* **47**, 617–624 (2009).
179. Cao, B. & Mao, X. The ubiquitin-proteasomal system is critical for multiple myeloma: implications in drug discovery. *Am. J. Blood Res.* **1**, 46–56 (2011).
180. Zhao, L., Zhao, J., Zhong, K., Tong, A. & Jia, D. Targeted protein degradation: mechanisms, strategies and application. *Signal Transduct. Target. Ther.* **7**, (2022).
181. Chassin, H. *et al.* A modular degron library for synthetic circuits in mammalian cells. *Nat. Commun.* **10**, (2019).
182. Yesbolatova, A. *et al.* The auxin-inducible degron 2 technology provides sharp degradation control in yeast, mammalian cells, and mice. *Nat. Commun.* **11**, (2020).
183. Iwamoto, M., Björklund, T., Lundberg, C., Kirik, D. & Wandless, T. J. A general chemical method to

regulate protein stability in the mammalian central nervous system. *Chem. Biol.* **17**, 981–988 (2010).

184. Sabapathy, K. & Lane, D. P. Therapeutic targeting of p53: All mutants are equal, but some mutants are more equal than others. *Nat. Rev. Clin. Oncol.* **15**, 13–30 (2018).

185. Li, B. & Dou, Q. P. Bax degradation by the ubiquitin proteasome-dependent pathway: Involvement in tumor survival and progression. *Proc. Natl. Acad. Sci. U. S. A.* **97**, 3850–3855 (2000).

186. Ward, C. L., Omura, S. & Kopito, R. R. Degradation of CFTR by the ubiquitin-proteasome pathway. *Cell* **83**, 121–127 (1995).

187. Viotti, C. *ER to Golgi-Dependent Protein Secretion: The Conventional Pathway*. (2016). doi:10.1007/978-1-4939-3804-9

188. Benham, A. M. Protein Secretion and the Endoplasmic Reticulum. *Cold Spring Harb. Perspect. Biol.* (2012). doi:10.1101/cshperspect.a012872

189. Chung, C. T., Niemela, S. L. & Miller, R. H. One-step preparation of competent Escherichia coli: transformation and storage of bacterial cells in the same solution. *Proc. Natl. Acad. Sci. U. S. A.* **86**, 2172–2175 (1989).

190. Martella, A., Matjusaitis, M., Auxillos, J., Pollard, S. M. & Cai, Y. EMMA: An Extensible Mammalian Modular Assembly Toolkit for the Rapid Design and Production of Diverse Expression Vectors. *ACS Synth. Biol.* **6**, 1380–1392 (2017).

191. Yang, L. *et al.* RNA-Guided Human Genome Engineering via Cas9. *Science (80-.).* 823–827 (2013).