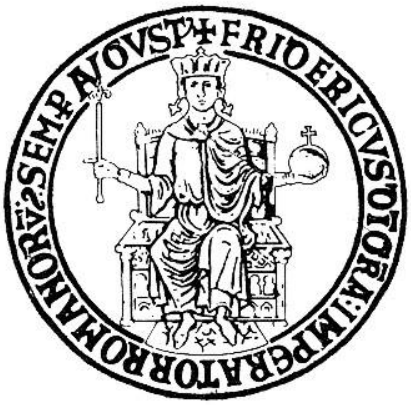
UNIVERSITY OF NAPLES FEDERICO II



PH.D. THESIS IN

INDUSTRIAL PRODUCT AND PROCESS ENGINEERING

DEPARTMENT OF CHEMICAL, MATERIALS AND INDUSTRIAL PRODUCTION ENGINEERING (DICMAPI)

XXXV CYCLE

The CASwitch: a Coherent Feed Forward Loop synthetic gene circuit for tight multi-level regulation of gene expression

Supervisor
Ch.mo Prof. Diego di Bernardo

Candidate
Giuliano De Carluccio
Matr. DR993931

ACADEMIC YEAR 2022/2023

Contents

Abstract

Synthetic biology is now an established biological engineering discipline that combines molecular biology and engineering knowledge with the ultimate goal of programming new cell functions using sophisticated gene circuits. During the last two decades, synthetic biology has grown into a thriving field that deploys an extensive toolkit of knowledge, devices, design strategies, and that holds great promise for the future. Despite the progress made, most current developments are not immediately translatable to “outside-the-lab” contexts.

This thesis focuses on the use of synthetic biology to improve current state-of-the-art inducible systems for controlled modulation of gene expression. Inducible gene expression systems allow to switch on, or off, transcription of a gene of interest by treating cells or even whole organism with a chemical compound. As such, they are indispensable research tools in the life sciences and in biotechnology.

However, state-of-the-art inducible gene expression systems all exhibit unintended basal gene expression, commonly known as *leakiness*, which affects their modularity and increases the likelihood of accidental activation. This makes it difficult to predict their behavior and limits their potential applications in biotechnological and clinical settings.

The goal of this thesis is to develop a novel inducible gene expression system that goes beyond the state-of-the-art *Tet-On3G* system, the most broadly used inducible gene system in mammals, by means of a synthetic biology approach. To date, the Tet-On3G encompasses more than 7000 publications and several implementations in inducible animal models in research contexts. However, its deployment for applications “out-side the lab” is severely hampered by its high level of leakiness.

In this thesis, I built on the Tet-On3G, by wiring it into a multi-level regulated synthetic gene circuit implementing a coherent feed forward regulatory motif (CFFL-4), endowing the Tet-On3G with quasi-zero leakiness while retaining high maximum inducible expression, resulting in more than one order of magnitude increase in fold induction levels

compared to the current state-of-the-art. This resulted in the generation of a new tight inducible gene system in mammalian cells that I called it the *CASwitch*, for its capacity to switch gene expression off or on at will by means of a CRISPR-Cas13d endoribonuclease.

To prove the usefulness of the CASwitch in applications of practical interest, I first modified the CASwitch to build two *whole-cell biosensors*: a copper biosensor able to detect copper in growing medium, and lysosomal stress biosensor able to sense the activation of the stress-responsive transcription factor TFEB. Both biosensors resulted in considerable improvements in dynamic range, detection resolution, and response reliability, as compared to their state-of-the-art counterparts.

Finally, I also applied the CASwitch to enable the development of factory cells able to produce Adeno Associated Virus vectors (AAVs) on demand. Specifically, I modified the CASwitch to achieve tight but inducible control of toxic viral gene expression, essential for AAV production and demonstrated inducible production of AAVs by means of doxycycline. These results hold great promise for further feasible engineering of a stable AAV producer HEK293 cell line.

Chapter 1

Introduction to Synthetic Biology

1.1 Synthetic Biology

Synthetic biology is an exciting interdisciplinary field that combines the knowledge and principles from molecular biology, system biology, and several engineering disciplines to develop application-driven new artificial biological functionalities.

Synthetic biology builds upon the discovery that networks of interacting genes and proteins regulate cellular functions and response to environment. By the end of the 1990s, it was gradually acknowledged that genes and proteins could be regarded as modular parts that could be systematically and rationally manipulated to engineer specific functions in cells. As complement to the top-down approach of systems biology, a bottom-up approach was proposed, which could exploit an increasing array of molecular "parts" to forward-engineer regulatory networks. This approach could be employed both to examine the functional organization of natural systems and to generate artificial regulatory networks with potential applications in biotechnology and health. This approach led to the emergence of the field of Synthetic Biology as a formal biological engineering discipline in the early 2000s, with the development of simple gene regulatory circuits that performed artificial functions in analogy to electrical circuits, such as the toggle-switch and repressilator circuits (Cameron et al., 2014; Gardner et al., 2000; Michael B. Elowitz & Stanislas Leibler, 2000).

The following two decades have seen a tremendous increase in the availability and characterization of genetic parts and modules that “in principle” allow the engineering and

programming of several cell functionality with user-defined complexity and purpose (M. Xie & Fussenegger, 2018). Although viewing parts and modules as functional units with defined input/output relationships that can be connected to form systems of higher complexity is a convenient abstraction for design, the reality is that the behavior of these modules is often not ideal, showing high leakage and properties alteration. This lack of modularity of basic parts makes the realization of synthetic genetic circuits often a lengthy and ad hoc process and finding solutions to this problem is an ongoing major research concern in synthetic biology (Del Vecchio et al., 2016).

Indeed, the long-term goal of synthetic biology could be regarded as the development of reliable and sophisticated gene circuits that serve as a kind of "genetic software" to program cellular functions analogously to the case of electronic computers. The applications of synthetic biology are broad and have the potential to revolutionize various industries, including medicine, agriculture, and biotechnology, with the possibility to fabricate, for example, organisms that could clean hazardous waste in inaccessible places, to use plants to sense chemicals and respond accordingly, to produce clean fuel in an efficient and sustainable fashion, or to recognize and destroy tumors (Purnick & Weiss, 2009).

1.2 Synthetic gene circuits: engineering gene expression

One of the possible applications of a synthetic gene circuits is the control of gene expression in a user-defined predictable manner. The quantitative and spatio-temporal control of the expression of a gene regulating a specific cellular function indeed allows control over that cellular function.

Synthetic circuits consist of simple functional units known as genetic parts. Those parts can control the rate of gene expression at the transcriptional, post-transcriptional, translational, and post-translational level. These genetic parts can thus be assembled to form modules (or devices) that can encompass specific functions.

So far, genetic circuits have been developed using two major design paradigms (Purcell & Lu, 2014). The first is the digital paradigm, which uses the analogy with binary codes for design (Brophy & Voigt, 2014), resulting in circuits performing functions such as logic gates (Bonnet et al., 2013), memory elements (L. Yang et al., 2014), an oscillator (Michael B. Elowitz & Stanislas Leibler, 2000), a toggle switch (Gardner et al., 2000), and a digitizer (Rubens et al., 2016), for instance. The second is the analog paradigm which takes advantage of circuits designs naturally present in cells to construct synthetic circuits able to perform complex temporal dynamics computations over a continuous input range (Teo et al., 2015). Mathematical modelling and control theory principles have been implemented in both paradigms with great impact on creating essential circuit design knowledge (Del Vecchio et al., 2016).

How genetic parts are designed to interact in a synthetic circuit often underpins its function. The pattern of interconnections among genetic parts is known as *topology*, different circuit topologies can thus endow a synthetic circuit with distinct functions. Recurring wiring patterns are known as *motifs*. The same small set of network motifs appear to serve as the building blocks of transcription networks from bacteria to mammals, endowing them with specific properties (Alon, 2007; Mangan & Alon, 2003). Two well-described motifs are feedbacks and feed-forward loops. It has been demonstrated, for instance, that these motifs can provide functions such as robustness to noise, noise-filtering, improved temporal response and robustness to genetic context (Qian et al., 2018). Therefore, synthetic circuits can be designed to have specific topologies to endow a system with defined properties.

1.3 A synthetic biology approach to engineer inducible gene expression systems

Inducible systems for controlled modulation of gene expression are indispensable research tools in the life sciences and in biotechnology. A reliable and well-established configuration

consists of inducible gene expression systems regulating gene transcription. In these systems, the gene expression control is operated by a trans-activator, or repressor, that acts on a downstream gene of interest (GoI). In mammalian cells, inducible synthetic trans-regulators possess a ligand-responsive sequence-specific DNA-binding domain (DBD), usually derived from prokaryotic regulator protein families, that is fused to an eukaryotic transcriptional modulator enabling positive or negative regulation of transcription. The binding of the ligand modulates the affinity of the DBD for a cognate DNA target sequence through tethering or other allosteric mechanisms, and allows the trans-regulator to target a promoter, where it triggers transcriptional activation or repression (Ausländer & Fussenegger, 2016).

The most established and broadly applied mammalian inducible gene expression system is the tetracycline-inducible gene expression system (TET-system) (Siddiqui et al., 2022), schematized in **Figure 1**.

The TET-system comprises two complementary systems, initially described as the tTA dependent (Gossen & Bujard, 1992) and rtTA dependent (Gossen et al., 1995) expression systems. They are now commonly referred to as the Tet-Off System (i.e. tTA dependent) and the Tet-On System (i.e. rtTA dependent). In each system, a recombinant tetracycline-controlled Transcription Activator (tTA) or a reverse-tetracycline Transcriptional Activator (rtTA) is able to bind to its cognate binding site, the tet operators (tetO), in a tetracycline responsive promoter, pTET (Fig. 1).

In the presence of the effector substance tetracycline (Tc) or one of its derivatives, doxycycline (dox), rtTA binds to the promoter, hence inducing transcription. On the contrary, the interaction between tTA and DNA is prevented by doxycycline, hence interrupting the expression of the target gene. Thus, the way the two versions of the Tet-system respond to tetracyclines is opposite. A disadvantage of the Tet-Off system is that the Tc or dox have to be removed to activate gene expression. Although this can be achieved relatively easily in small cell culture experiments by thorough washing of the cells and replacement of the culture medium, it is more problematic when large cell cultures are used

and in animal experiments. For this reason, the Tet-On was developed and is usually preferred to the Tet-Off system for its higher versatility (T. Das et al., 2016).

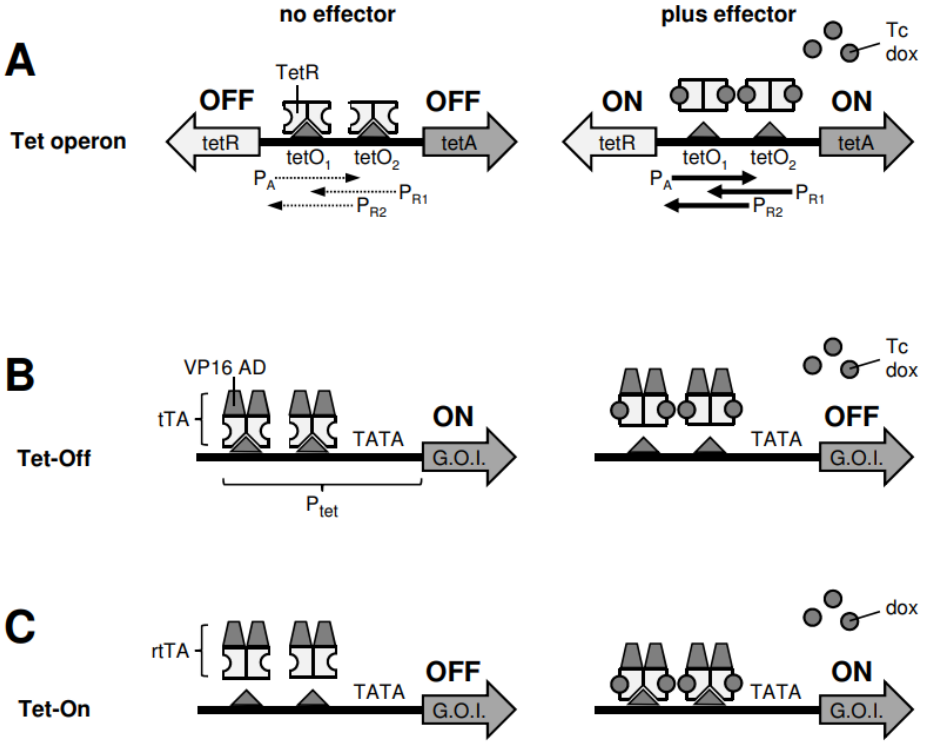


Figure 1 Tc-controlled regulation of gene expression. (A) Tn10 tet operon. In E. coli, TetR binds as a dimer to the tetO1 and tetO2 sites in the Tn10 tet operon. This interaction blocks the activity of the underlying promoters (PA, PR1 and PR2) and inhibits transcription of the tetA and tetR genes. Binding of Tc or dox triggers a conformational switch in TetR that prevents tetO binding and results in the activation of TetA and TetR production. (B) The Tet-Off system. Fusion of TetR to the activation domain of the herpes simplex virus VP16 protein (VP16 AD) resulted in the Tc-controlled transcriptional activator (tTA). Binding of tTA to the Ptet promoter that consists of 7 tetO sequences fused to a minimal TATA-box containing eukaryotic promoter, activates expression of the downstream positioned gene-of-interest (G.O.I.) Binding of Tc or dox induces a conformational change in the TetR domain of tTA, which prevents tetO binding and switches gene expression off. (C) The Tet-On system. The reverse-tTA (rtTA) variant exhibits a reverse phenotype and does not bind tetO in the absence of an effector. Bind- ing of dox triggers a conformational switch in rtTA, which allows tetO binding. Subsequent activation of the Ptet promoter drives expression of the downstream positioned gene. The initial version of rtTA had a low affinity for Tc and was not activated by this compound. From (T. Das et al., 2016).

In more details, the Tet-On system is composed by the following parts:

- **The reverse-tetracycline Trans-Activator (rtTA).** It consists of the reverse Tetracycline repressor (rTetR) to which the herpes simplex virus (HSV)-VP16 transcription activation domain (AD) is fused. The rTetR DNA binding domain results from few point mutations within the original E.Coli-derived Tetracycline Repressor TetR, endowing it with reverse tetracycline responsiveness, hence rtTA requires

tetracyclines for binding to tetO. The rtTA has been through extensive optimization resulting in the latest commercially available version, known as **rtTA3G** which is part of the **Tet-On3G** gene exrpression system (Clonetech). The rtTA3G features a high-affinity rTetR fused to a triplet of the 12-aa minimal AD derived from the HSV-VP16 (Zhou et al., 2006).

- The **pTET**, this is a synthetic promoter comprising a minimal RNA polymerase II promoter preceded by multimerized tetO sequences. This arrangement makes the activity of pTET dependent on the binding of rtTA. Its design has been through extensive adaptations regarding the source of the minimal promoter and the specific configuration of the tetOs. The original version consisted of a CMV minimal promoter fused to an array of seven tetO sequences in the pTET-1. The prevailing nomenclature for this promoter in eukaryotic is pTRE standing for tetracycline responsive element. The latest optimized version of this promoter commercially available with the name of pTRE3G, as part of the Tet-On3G system (Clonetech), contains a tetO heptamer with equally sized random spacer sequences preceding a modified CMV minimal bearing a 5'-UTR derived from a plant RNA virus turnip yellow mosaic virus (TYMV) (Loew et al., 2010).

Ideally, an optimal inducible gene expression system should exhibit essentially no expression of the target gene in the absence of the inducer molecule (i.e. doxycycline), while enabling maximal expression in its presence (Siddiqui et al., 2022; Yeoh et al., 2019). However, a major drawback of the Tet-On3G system is actually its high basal expression in the absence of inducer molecule, a phenomenon known as *promoter leakiness*. Its causes are the intrinsic rtTA-independent activity of the pTet promoter because of the presence of binding sites for general transcription factors in its minimal promoter, and the residual binding of rtTA to pTRE even in the absence of doxycycline (T. Das et al., 2016).

Very tight control of gene expression is particularly valuable in situations where leaky gene expression is undesirable, such as with the expression of a toxic protein, or when short transient induction is preferable to long-term induction. In the context of human therapy,

for example, some therapeutic proteins can carry severe toxicities such as Hepatocyte growth factor (HGF) that stimulates liver regeneration and can be used to treat hepatic injury but is also a tumorigenic factor. Or in metabolic engineering, where using inducible gene systems with a high basal level of enzyme expression can lead to toxicity in producer cells (Cubillos-Ruiz et al., 2021; Purnick & Weiss, 2009)

The ease of use of the Tet-On system led to its broad application with more than 7000 publications and implementation in several inducible animal models (https://www.tetsystems.com). However, its leakiness represents a great hurdle for its effective implementation in biotechnology and in the clinic. Reducing the leakiness of the Tet-On3G system, resulting in a very tight inducible gene expression system, would retain its ease of use thus its broad applicability, and would make it effectively useful for biotechnological and translational applications.

1.4 Synthetic circuit topologies to reduce leakiness

The integration of transcriptional and post-transcriptional control in synthetic gene circuits has been successfully used to decrease leakage and achieve tight gene expression regulation (Deans et al., 2007; Greco et al., 2020; Litovco et al., 2021; Origel Marmolejo et al., 2020; Pieters et al., 2021). Moreover, the interplay between transcriptional and post-transcriptional control has also been utilized to design gene circuits that mimic half-adder circuits capable of performing basic arithmetic (Auslände et al., 2012) and displaying oscillatory gene-expression behavior (Tigges et al., 2009). Indeed, the key for leakage reduction in these synthetic circuits is how the combination of transcriptional and post-transcriptional regulations was designed. Specifically, these synthetic circuits all implement a feed-forward loop (FFL) regulatory motif, as schematized in **Figure 2**.

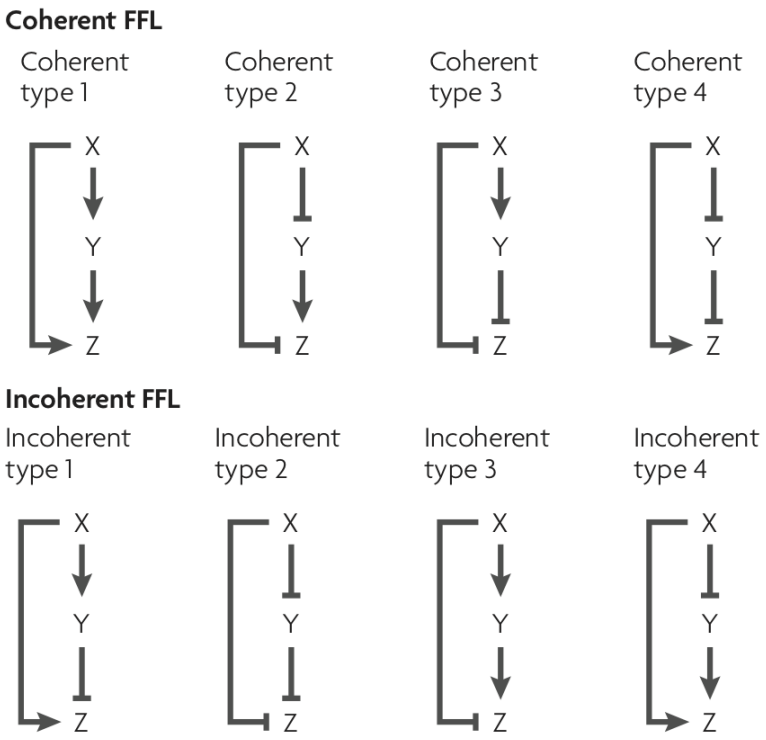


Figure 2 Feedforward loops (FFLs). The eight types of feedforward loops (FFLs) are shown. In coherent FFLs, the sign of the direct path from main input X to main output Z is the same as the overall sign of the indirect path through the intermediate output Y. Incoherent FFLs have opposite signs for the two paths. Adapted from (Alon, 2007)

The feedforward loop (FFL) is a pattern made up of three parts, including a main input "X", an intermediate output "Y", and a main output "Z"; where X regulates Y, and both jointly regulate the main output Z. The FFL can assume eight possible structural types, because each of the three interactions in the FFL can be activating or repressing. Four of these configurations are referred to as "coherent," in which the sign of the direct regulation path (from X to Z) is the same as the overall sign of the indirect regulation path (from X through Y to Z). The remaining four structures are referred to as "incoherent," in which the signs of the direct and indirect regulation paths are opposite (Mangan & Alon, 2003).

Synthetic gene circuits implementing a coherent feed forward loop (CFFL) have been demonstrated to mitigate leakage and to increase the fold-change of the output, enabling a more stringent gene expression control (Nitzan et al., 2017; Pieters et al., 2021). Recently, the implementation of a coherent-feedforward type 4 (CFFL-4) motif, in which the direct regulation pathway is activating (X → Y) and the indirect pathway is formed by cascading

two inhibitors (X --|Y --|Z), has been found to endow established transcriptional inducible gene systems with reduced leakiness and enhanced fold induction in bacteria cells (Litovco et al., 2021). In this CFFL-4 motif circuit, for a low level of input X, the inhibitor Y is highly active and strongly reduces the leakage of the output. Instead, for a high level of input X, the inhibitor level Y is low and therefore, the maximum output level is kept high, thus this circuit design reduces only the basal output without impacting the maximum output, leading to increased fold induction levels.

Aim of the thesis

The engineering of a synthetic gene circuit implementing a CFFL-4 motif in mammalian cells could be an optimal solution to reduce the leakiness of the state-of-the-art Tet-On3G system but without affecting its maximum expression. High induction levels combined with low leakiness would result in an excellent dynamic range allowing regulation of gene expression over several orders of magnitude. This new inducible gene expression system could be effectively used in translational and biotechnological implementations.

This thesis focuses on the design and construction of a synthetic gene circuit implementing a CFFL-4 regulatory motif, resulting in a new multi-level tight controller of gene expression in mammalian cells: the CASwitch. The thesis will also provide a proof-of-principle of applications of the CASwitch technology for the engineering of a stable Adeno-Associated Vector packaging HEK293 cell line, and for the development of reliable Mammalian Whole-Cell Biosensors.

The research presented in this thesis shows the potential of a synthetic biology approach as a disruptive technology generator where traditional incremental technology approaches fall short and provides useful insights about synthetic gene circuits design in mammalian cells.

Chapter 2

Introduction to Adeno-Associated Vector Manufacturing

2.1 Adeno-Associated Virus (AAV) biology

Adeno-associated virus (AAV) is a very small (~22nm), non-enveloped, non-pathogenic virus belonging to the family of Parvoviridae and genus Dependovirus. This genus comprises ssDNA satellite viruses which require the host cell co-infection by other "helper" viruses to complete its lytic cycle. Known AAV helper viruses are Adenovirus and Herpesvirus.

AAV is characterized by a linear ssDNA genome of about 4.7kb in size. It is flanked by two T-shaped sequences named Inverted Terminal Repeats (ITRs) that serve as viral origins of replication and packaging signal. The genome encodes for both non-structural (Rep) and structural (Cap) genes on the same DNA strand. Transcription of those gene is regulated by the three promoters P5, P19 and P40. P5 transcripts are expressed first, yielding Rep78 and 68 by means of alternative splicing; followed by those from P19, producing Rep52 and 48; then those from P40, that are translated into VP1 and VP2 through alternative splicing, and into VP3 by ribosome leaky scanning mechanism, in a 1:1:10 ratio essential for correct capsid assembly (Fig.1). Rep78/68 allows genome replication through rolling-hairpin mechanism, Rep52/48 its packaging into capsid formed by VP1,2,3 (Balakrishnan & Jayandharan, 2014).

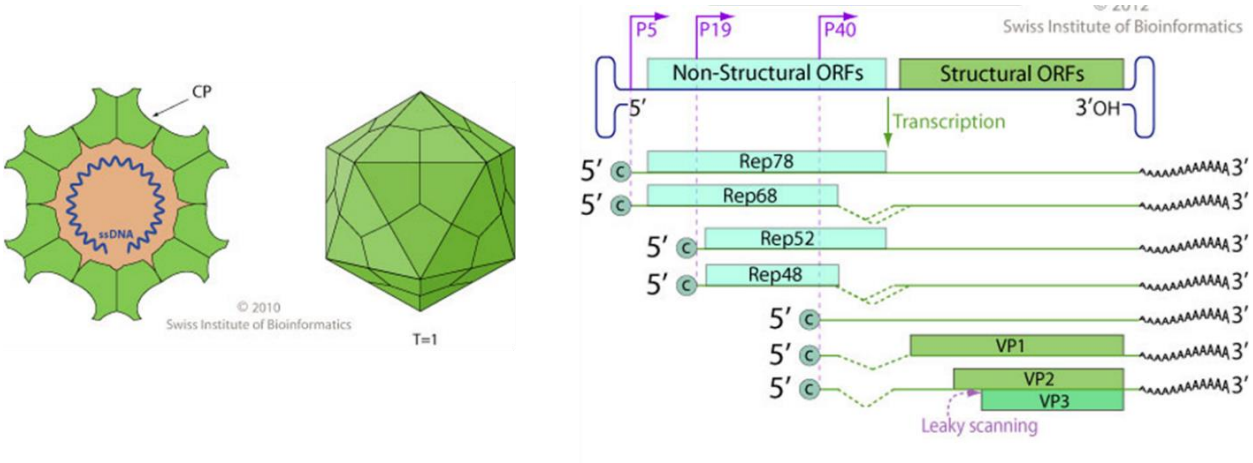


Figure 2.1 Adeno Associated virus virion (Left) and genomic structure (Right). CP, Capsid proteins. ssDNA, single strand DNA. ssDNA genome of AAV is packaged into an icosahedral capsid without envelope. The genome of AAV encodes two genes Rep and Cap, producing non-structural and structural proteins, respectively, required for AAV genome replication, packaging, and capsid assembly. Rep encodes for four proteins: Rep78, Rep68, Rep52 and Rep40. Cap encodes for VP1, VP2 and VP3. (Adapted from https://viralzone.expasy.org/)

Recombinant Adeno-Associated Virus Vectors (rAAV) harness the ITRs to replicate and package an exogenous DNA element, replacing the Rep/Cap genes with an expression cassette of a gene of therapeutic relevance. The first to demonstrate the utility of AAV as mammalian gene transfer vehicle was Hermonat et al. in 1984 (Hermonat & Muzyczka, 1984). To date, Adeno-associated Vectors represent the most promising gene therapy vector due to their minimal immunogenicity, tissue-selective serotypes, amenability to capsid modifications, and clinical successes. With hundreds of ongoing clinical trials, AAV vectors are expected to address a wide range of conditions in the coming decade. However, upcoming increase in production demand cannot be met by current manufacturing technologies. Several Biotech companies are now competing to develop newer manufacturing strategy to enable effective upscaling of rAAV production without losing vector yield and quality (Penaud-Budloo et al., 2018).

2.2 Established AAV production methods and their drawbacks

So far, distinct AAV production methods have been developed however the most established ones are: (i) the transient triple transfection in mammalian cells; (ii) replication defective-helper virus co-infection of mammalian cell lines, and (iii) infection of insect cells with recombinant baculoviruses (BEVs) (Merten, 2016).

Transient triple transfection of HEK293 cell line is the most broadly used production method. It is commonly used in academic pre-clinical context, mostly due to its manufacturing versatility at small scale. In this method three plasmids are transfected into HEK293 cells: (1) a transfer plasmid, (2) a Rep-Cap plasmid, and (3) a helper plasmid. The transfer plasmid contains the transgene expression cassette flanked by ITRs. The Rep-Cap plasmid encodes Rep and Cap proteins required for capsid formation, transgene replication and its packaging. Finally, the helper plasmid works as a proxy for helper virus co-infection. The most common helper plasmid encodes for the Human Adenovirus 5 (HAdV5) genes, including adenoviral E2A, E4 and VARNA genes, essential to help AAV vector particles formation. The remaining essential adenoviral helper genes, namely E1A and E1B, are supplied by HEK293 cells themselves as these are stably integrated in their genome.
Whereas this method is helpful in pre-clinical developmental stages, as it allows systematic modification of the capsid or of the transgene sequence, it lacks scalability. Indeed, the triple transfection method requires adherently growing cells, therefore to scale-up the process parallel culture plates are needed, but this actually involves numerous practical difficulties. The growth of mammalian cells on a surface also presents variability as it can be difficult to maintain consistent parameters such as pH and oxygen concentration across multiple culture vessels. Calcium phosphate and Polyethylenimine are commonly used transfection agents, but their transfection efficacy can be impacted by pH and impurities, resulting in variations from batch to batch. Liposomal transfections have high efficiency and low cytotoxicity but require more expensive reagents. Due to all these factors, scaling up the culture of mammalian cells grown on a surface remains challenging and costly (Srivastava et al., 2021).

This method can be improved by means of HEK293 growing in suspension, but it still relies on transfection (Grieger et al., 2016).

Other systems have been developed that could address the scalability limitation avoiding the difficulties of large-scale transfections infection of mammalian cell lines with replication-defective recombinant helper virus, and infection of insect cells with recombinant baculoviruses (BEVs).

In the Herpes simplex virus (HSV) complementation platforms HEK293 cell line is infected with ICP27-deficient recombinant Herpes Simplex Virus (HSV) vectors which are replication-incompetent. Often two replication-deficient HSVs are employed: one carrying the desired transgene expression cassette and another carrying the AAV Rep-Cap genes. A major drawback of HSV methods is that infection of HEK293 cells with high titer of both rHSVs is required to produce AAV vector, hence manufacturing of two distinct HSV vectors to high titer prior to using them for making the AAVs, adding to the overall complexity of the process (Robert et al., 2017).

Production via recombinant baculovirus (BEVs) employs infection of cultured Spodoptera frugiperda Sf9 cells with a BEV transducing Rep78/52 and VPs coding sequences and the other one carrying the transgene expression cassette. This is possible because these insect cells do not require additional helper viral genes to induce the expression of Rep and Cap genes. Moreover, they are amenable to the constitutive expression of Rep proteins, whereas for mammalian cells, such as HEK293, these are toxic. Hence, this property allowed to develop an Sf9 cell line stably integrated with AAV Rep and Cap genes, where AAV production is then induced by the infection with only one recombinant baculovirus providing the AAV transgene expression cassette. A key drawback of insect cell techniques is that they typically lead to altered capsid compositions (abnormal VP1, VP2, VP3 ratios) and therefore the resulting AAVs exhibit lower therapeutic potencies, thus further improvements are required to ameliorate vector titer, quality, and potency (Kondratov et al., 2017).

Currently, state-of-the-art methods for d AAV vector production platforms, all have limitations when it comes to scaling up for clinical production. Some methods, such as transient triple transfection, are practically difficult to scale-up, while others, like BEVs complementation platforms, can be more easily scaled-up but may compromise vector quality. Therefore, there is a clear unmet need for improvements in AAV production platforms for large scale clinical AAV manufacturing.

A revolutionary strategy in this context would be the development of a mammalian AAV peoducer cell line with inducible expression of AAV Rep and Cap genes, thus enabling AAV production on demand. This feat has not been achieved so far because of the toxicity of the AAV Rep proteins in mammalian cells that requires tight gene regulatory strategies to prevent their expression in uninduced conditions.

2.3 Design hurdles in Adeno-Associate Vector packaging HEK293 cell line

An AAV packaging cell line should stably harbor in its genome inducible Rep/Cap gene and helper genes. The stable integration of a specific transgene expression cassette flanked by ITRs in addition, could led to the generation of an AAV producer cell line.

An AAV packaging cell line based on the HEK293 cell line is preferred because it is an already approved manufacturing cell substrate in the clinical setting. However, the E1A gene present in HEK293 cell line causes the constant transactivation of Rep proteins from the endogenous AAV P5 and P19 promoters. Since Rep proteins are known to inhibit cell proliferation by inducing cell-cycle arrest or induce apoptosis (Berthet et al., 2005; Schmidt et al., 2000), the isolation of a stable HEK293 packaging or producer cell line is not trivial. Therefore, expression of Rep proteins has to be tightly controlled to avoid their toxicity, hence cell clone death, but at the same time it has to be rescuable when production is needed.

So far, attempts at developing such inducible Rep constructs have been reported (Grimm et al., 1998; Li et al., 1997; Ogasawara et al., 1998; Q. Yang et al., 1994), however AAV P5 promoter replacing with heterologous ligand-responsive promoters, besides resulting in Large Rep leaky expression, could not control the expression of the small Rep52/40 proteins stemming from AAV P19 promoter, which is internal to the Rep78/68 coding sequence. Replacing that promoter means both large and small Rep proteins coding sequences disruption. For those reason, other attempts have been made using different cell lines, such as HeLa cell line in which E1A and E1B expression was controlled (Qiao, Li, et al., 2002). But also in this case, E1A and B expression control resulted to be leaky leading to instability of obtained cell clones and infeasibility to isolate a high-titer producer cell line.

All these failed attempts at the turn of the millennium brought to the establishment of other packaging/producer cell lines development strategies, all having in common the use of recombinant Adenovirus vector infection. In these Adenoviral complementation AAV production platforms packaging cell lines are often based on HeLa cells, although A549 cells have also been used (Thorne et al., 2009). In this case, packaging cell lines contains only Rep and Cap genes and AAV vector production is induced by infection with two recombinant Adenoviral vectors: one having functional E1A/E1B genes but temperature sensitive E2B DNA polymerase, the other containing the AAV transgene expression cassette in place its E1 region. If a producer cell line, already bearing the AAV transgene expression cassette, is used the production requires the infection with only the temperature sensitive recombinant Adenoviral vector. The major drawback for this production strategy is that it requires the removal and inactivation of the Adenovirus, moreover the temperature-sensitive adenovirus mutants are prone to reversion and are difficult to produce and to characterize. For these reasons this production platform has never been implemented for large-scale AAV production (Merten, 2016).

The establishment of producer cell lines was also reported in HEK293 cells in the context of the Adenoviral complementation platform. In this case, expression of the Rep proteins is tightly controlled by the CRE-LoxP recombinase system. A polyadenylation signal flanked by

two LoxP sequences has been inserted immediately downstream the AAV P19 promoter in other to disrupt both Large and Small Rep coding sequences in the Rep/Cap genome of AAV2, then this construct has been stably integrated in HEK293 cell line. Although, achieving tight control of Rep expression, production of rAAV from these cells has to be triggered by infection with a recombinant, E1-deleted, Adenoviral vector expressing the CRE recombinase (Qiao, Wang, et al., 2002). The advantage of using these "producer" HEK293 cell lines is that already provides E1A and E1B helper functions allowing the production with a replication-incompetent Adenovirus. However, because recombination events that reconstitute replication-competent Adenovirus particles might occur an Adenoviral purification step is still required for a GMP purposes.

Overall, reported attempts show that achieving tight control of the expression of toxic Rep proteins is a prerequisite to successfully engineer an AAV packaging HEK293 cell line, however, even though not appointed in those studies also adenoviral helper genes are toxic and should be tightly controlled in order to avoid packaging HEK293 cell clone death. Regarding Rep proteins, the main design hurdle to tight control their expression is related to Rep gene structure. Replacing the P5 promoter with heterologous promoters even being a valid solution to control Rep78/68 proteins expression, it cannot be employed to control Rep 52/40 proteins expression, because replacement of the internal AAV P19 promoter would cause the distruption of both large and small rep coding sequences.

The development of a synthetic gene circuit implementing a CFFL-4 regulatory motif endowing the Tet-On3G system with tight inducible gene expression control both at transcriptional and post-transcriptional levels could be a valuable solution to achive regulation of both large and small Rep proteins, and helper genes, enabling the generation of an AAV packaging HEK293 cell line. Specifically, the post-transcriptonal control would make possibile to regulate small Rep proteins and it would also contribute to the trasncriptional control exherted on large Rep and Helper genes, enahcing the tighteness of their expression regulation.

Chapter 3

Construction of the CASwitch: a multi-level regulated coherent feedforward loop gene circuit in mammalian cells.

This chapter outlines the process of designing and constructing a multi-level regulated Coherent Feedforward Loop type-4 (CFFL-4) in mammalian cells, starting from the rational design of the synthetic circuit to its final implementation in HEK293T cells, going through the construction and characterization of the single biological parts composing the circuit. This CFFL-4 circuit, which I named CASwitch, works as a tight inducible gene expression system in mammalian cells, with improved gene expression switching capabilities compared to the state-of-art Tet-On3G system.

Specifically, here I first describe how I built the circuit by engineering two key components, a tetracycline-repressible promoter (pCMV/TO) and a CRISPR-Cas13d endoribonuclease (RfxCas13d or CasRx), to create a CFFL-4 topology. Finally, I will conclude the chapter by benchmarking the engineered CASwitch against the Tet-On3G.

The applications of the CASwitch range from the generation of stable cell lines with inducible expression of toxic genes, i.e. for the generation of an inducible adeno-associated vectors, to the construction of amplified whole-cell biosensors.

3.1 Introduction

As shown in Figure 3.1, a coherent feed-forward loop type-4 (CFFL-4) is characterized by a input species X that activates the main output species Z, while repressing the intermediate output species Y, which in turn represses Z (Mangan & Alon, 2003). In the context of transcriptional networks in bacterial cells, the type-4 CFFL is capable of suppressing leakiness in the expression of species Z, thanks to the presence of species Y, and of achieving high-fold-change activation (Litovco et al., 2021). Hence, I chose this topology as it represents a good starting to improve on the state-of-the-art inducible gene expression system in mammalian cells, i.e. the Tet-On3G system, which is characterized by high levels of leaky expression in the absence of its inducer molecule (doxycycline). Indeed, in the past few years, many efforts have been carried out to solve this problem. However, the incremental technological improvement approach did not completely solve the problem, with the latest version of the Tet-On system (Tet-On3G) still exhibiting significant leakiness levels.

Therefore, as shown in Figure 3.1B, in the CASwitch circuit the main activation pathway X → Z consists of the Tet-On3G system, where the tetracycline-inducible synthetic transcription factor rtTA3G is species X and it directly binds the pTREG promoter driving expression of the reporter gene (Gaussia Luciferase) representing species Z. To implement the sequential repression cascade (X --|Y--|Z), the first crucial necessity was to find a solution to enable the species Y-driven repression of the main output Z (i.e. Y--|Z), compatible with the implementation of the pTRE3G promoter. The Y-mediated repression of Z is a prerequisite for CFFL-4 implementation, and in the last few years, several repression solutions have been adopted. In prokaryotic systems, inverted promoters that harness transcriptional interference or RNA toehold switches have been used (Greco et al., 2020; Litovco et al., 2021). Unfortunately, these solutions cannot be used in mammalian cells; therefore, other strategies have been developed. In particular, miRNAs or composite promoters, usually consisting of two adjacent DNA sequences bound by two functionally opposite transcription factors, have been used (Origel Marmolejo et al., 2020; Z. Xie et al., 2011). Interestingly,

Marmolejo et al. implemented a CFFL-4 circuit by employing a composite promoter, which can be activated by the tTA transcriptional factor and repressed by an ERT transcriptional repressor to integrate X and Y signals. However, this solution was not optimal for reducing the leakiness in Z expression. In contrast, miRNAs can be used to repress the main output Z by acting post-transcriptionally on the 3'-UTR of the Z species; therefore, they have wider applications. However, to our knowledge, inducible miRNA expression by means of rtTA3G transcription factor is not efficient and needs to be further optimized. Because to these limitations, as shown in Figure 3.1B, I decided to overcome them by means of the recently discovered Ruminococcus flavefaciens CRISPR-Cas13d (RfxCas13d or CasRx) endoribonucleases (Konermann et al., 2018). Similar to miRNAs, CasRx can efficiently post-transcriptionally repress a target gene with a modified 3'UTR, however, in contrast to miRNAs, CasRx is a standard protein coding gene and hence reliable expression control solutions compatible with the rtTA3G transcription factor are available. Finally, to implement the X-mediated repression of Y (i.e. X --|Y), I decided to rely on the pCMV/TO tetracycline repressible promoter (ThermoFisher) to enable the rtTA3G (i.e. species X) to inhibit CasRx expression (i.e. species Y).

Thus, as shown in Figure 3.1B, in my CFFL-4 genetic circuit implementation: species X is represented by the tetracycline-inducible rtTA3G transcription factor; the intermediate output Y is the CasRx endonuclease, driven by tetracycline repressible promoter pCMV/TO; finally, the main output species Z is the Gaussia Luciferase driven by the tetracycline responsive pTRE3G promoter, and degraded by the CasRx endonuclease thanks to a modified 3'UTR.

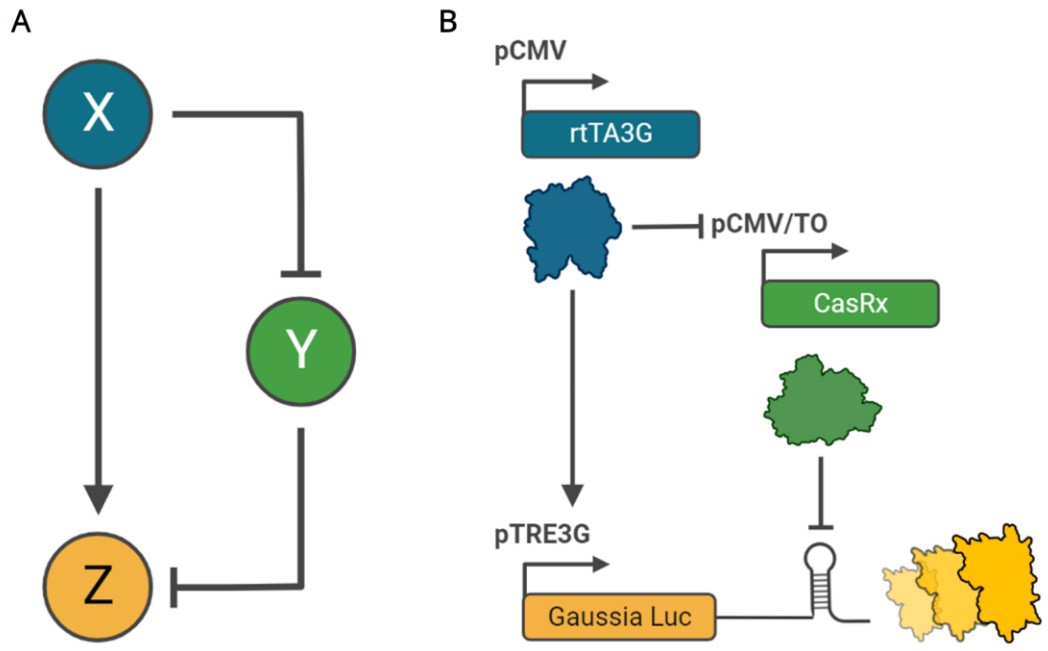


Figure 3.1. Coherent Feed Forward type-4 (CFFL-4) circuit topology and CASwitch implementation in mammalian cells. (A) Schematic of the type-4 CFFL (CFFL-4) topology. X, main input; Y, intermediate output; Z, main output; when X is active, it simultaneously induces Z while repressing Y. When X is inactive, there is no inhibition of Y, which in turn represses the main output Z. **(B)** CASwitch implementation in mammalian cells. X, reverse tetracycline transactivator (rtTA3G, Clonetech); Y, Ruminococcus Flavefaciens CRISPR-Cas13d (CasRx); Z, Gaussia Luciferase. In the presence of doxycycline, rtTA3G is active and transcriptionally induces Z through the pTRE3G promoter, and transcriptionally inhibits CasRx via the pCMV/TO promoter. In the absence of doxycycline, rtTA3G is not active; hence, there is no CasRx inhibition, which can post-transcriptionally repress Z through the endonucleolytic cleavage of its transcript mediated by the recognition of a cleavage site (stem-loop structure).

3.2 Experimental implementation and characterization of transcriptional repression of the pCMV/TO promoter by the rtTA3G transcription factor.

The pCMV/TO promoter is derived from the CMV promoter modified to harbor two Tetracycline Operon (TO) sequences downstream of the TATA-binding box (Thermo Fisher). As shown in **Figure 3.2A**, these two TO sequences are bound by the reverse tetracycline transcriptional activator (rtTA3G) in the presence of tetracycline or its analog doxycycline.

When rtTA3G binds to these sequences, it creates a steric hindrance for the mammalian Pol II at the promoter transcriptional start site (TSS), thus resulting in transcriptional inhibition.

In the CFFL-4 circuit implementation, I used the pCMV/TO promoter to negatively control the expression of the CasRx in the presence of doxycycline, as shown in **Figure 3.1B**. CasRx inhibition is required to release suppression of the main output Z, which is simultaneously activated by rtTA3G in the presence of doxycycline through the pTRE3G promoter.

The efficiency of rtTA3G-mediated repression of expression from the pCMV/TO promoter was assessed using a Gaussia/RedFirefly Dual Luciferase assay, as shown in Figure 3.2A, where the Gaussia Luciferase reporter gene is cloned downstream of the pCMV/TO promoter (pCMV/TO-gLuc). To characterize rtTA3G repression of the pCMT/TO promoter, HEK293T cells were transfected with the a constant molar amount of the plasmid constitutively expressing rtTA3G (pCMV-rtTA3G), while the pCMV/TO-gLuc plasmid was transfected at increasing molar amounts, thus encompassing several different molar ratios of the two plasmids (pCMV-rtTA3G : pCMV/TO-gLuc) ranging from 1:0.1 to 1:10 as shown in Figure 3.2B,C. As a negative control “Mock”, cells were transfected with only the reporter plasmid (pCMV/TO-gLuc) to assess the promoter expression in the absence of the rtTA3G. Cells in all conditions were also transfected with a pCMV-Red Firefly Luciferase reporter plasmid, employed as a transfection efficiency normalizer, and with a pcDNA3.1 empty plasmid (Thermo Fisher) to transfect the same total amount of DNA between all conditions. Finally, in all experimental conditions, cells were treated with increasing amounts of doxycycline (0, 10, 100, 1000 ng/mL.)

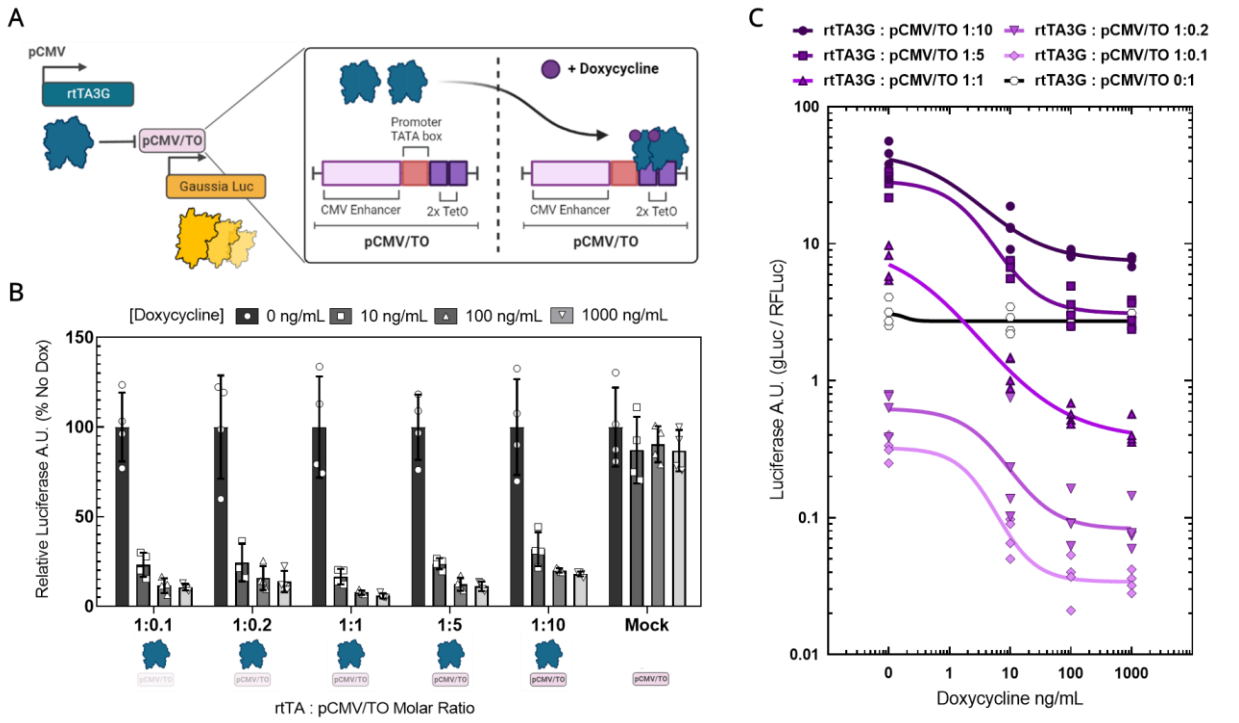


Figure 3.2. Experimental characterization of the pCMV/TO promoter repression by the rtTA3G transcription factor . A) pCMV/TO mechanism of transcriptional inhibition by Pol II steric hindrance. Starting from the 5' end, the pCMV/TO promoter is composed of a CMV enhancer and a TATA binding box, followed by two Tetracycline Operon (TetO) binding sequences. In the presence of doxycycline, rtTA3G can bind these sequences sterically, blocking the binding of the mammalian general transcription machinery to the TATA binding box, thus inhibiting the PolII-mediated transcription start. (B) pCMV/TO promoter repression in the presence of rtTA3G. In the presence of a 1000ng/mL concentration of doxycycline, the rtTA3G represses transcription from the CMV/TO promoter by more than 90%. A.U., arbitrary units of luminescence. (C) The data in panel B are here shown in logarithmic value and as dose-response curves. A negative correlation is present between pCMV/TO-driven luciferase expression and doxycycline. Tuning the relative molar amount of the pCMV/TO-gLuc plasmid to that of rtTA3G caused a vertical shift in the dose/response curve.

As a measure of promoter activity, arbitrary luciferase units (A.U.) were calculated as the ratio between the Gaussia and the Red Firefly luciferase values, and the Relative Luciferase A.U. was obtained as a percentage of the Luciferase A.U. at 0 ng/mL of doxycycline. The bar graph in Figure 3.2 B clearly shows that the pCMV/TO promoter is inhibited in the presence of doxycycline only if rtTA3G is expressed, suggesting that this behavior can be explained by the steric hindrance caused by rtTA3G in the presence of the inducer molecule. At 1000 ng/ml doxycycline, Gaussia luciferase expression decreased to approximately 10% of its

initial value for all the tested molar ratios (Figure 3.2B). Thus, no variation in repression efficacy between different molar ratios was observed, indicating that a small amount of rtTA3G is sufficient to saturate the pCMV/TO promoter. However, the molar ratio can be used to tune the initial and final absolute values of a gene expressed from the pCMV/TO promoter, in this case the Gaussia Luciferase as reported in **Figure 3.2C**. Overall, the pCMV/TO promoter proved to be a reliable way to repress downstream gene expression in the presence of doxycycline by more than 90% of its initial value.

3.3 Experimental implementation and characterization of post-transcriptional repression of gene expression by the CasRx endonuclease.

The Ruminococcus Flavefaciens CRISPR-Cas13d (RfxCas13d or CasRx) belongs to the Class 2 Type VI system Cas13 family of single-effector enzymes, comprising guide RNA-directed ribonucleases within fmy subtypes (Cas13a-d). Each of them exhibits significant sequence divergence apart from the two consensus HEPN (higher eukaryotes and prokaryotes nucleotide-binding domain) RNase motifs, R-X4–6-H. Cas13 enzymes process pre-crRNA into mature crRNA guides in a HEPN-independent manner to defend against viral infection, followed by HEPN-dependent cleavage of a complementary "activator" target RNA in trans. Upon binding of a constant direct repeat (DR) region (36 nt) derived from the characteristic repeats of CRISPR arrays, Cas13 enzymes cleave the pre-crRNA into a mature crRNA, which is composed of the constant DR and a spacer region complementary to the target protospacers (Konermann et al., 2018; C. Zhang et al., 2018), as schematically depicted in **Figure 3.3 A**. Recently, it has been reported that the pre-crRNA processing feature of CasRx can be employed to direct the cleavage of user-defined transcripts by taking advantage of the CasRx direct repeat (Rfx-DR) as an endoribonuclease recognition motif (DiAndreth et al., 2022). Unlike other Cas13 endoribonucleases, CasRx remains bound to its 3‘-cleaved product, therefore I could harness its cis-acting crRNA processing activity by placing the

cognate DR in the 3'-UTR (UnTranslated Region) of a target gene transcript, as shown in **Figure 3.3B**. In this way, CasRx binds the DR and thus cleaves the polyadenylation signal from the target gene transcript, causing its degradation by cellular exonucleases.

To confirm this post-transcriptional repression strategy and to quantify its efficacy, a Gaussia/RedFirefly Dual Luciferase assay was performed, as shown in **Figure 3.3B-C**. The Rfx-DR was cloned in the 3'-UTR of a constitutively expressed Gaussia Luciferase transcript, yielding the pCMV-gLuc-DR(Rfx) plasmid (Figure 3.3 B). Then, HEK293T cells were transfected with the same fixed molar amount of this plasmid (pCMV-gLuc-DR(Rfx)) and with increasing molar amounts of pCMV/TO-CasRx(NLS)-T2A-mCherry plasmid, encompassing several different molar ratios ranging from 1:0.1 to 1:10. As negative control, I included a condition without CasRx transfection (Mock). The pCMV/TO-CasRx(NLS)-T2A-mCherry plasmid includes the CasRX augmented with two SV40 nuclear localization signals at its N- and C- terminals, as it has been demonstrated that this modification results in a great improvement in the direct repeat cleaving efficacy (Konermann et al., 2018). Furthermore, I added a T2A skipping ribosome sequence to co-express the red fluorescent mCherry reporter, as a plasmid transfection control. As an additional negative control, cells were transfected with a reporter plasmid without Rfx-DR in its 3' UTR (pCMV-gLuc) at the same fixed molar amount and with increasing molar amounts of the pCMV/TO-CasRx(NLS)-T2A-mCherry plasmid. All experimental conditions were transfected with a Red Firefly reporter plasmid, employed as a transfection efficiency normalizer, and with a pcDNA3.1 empty plasmid to transfect the same total amount of DNA between all conditions (Figure 3.3 B)

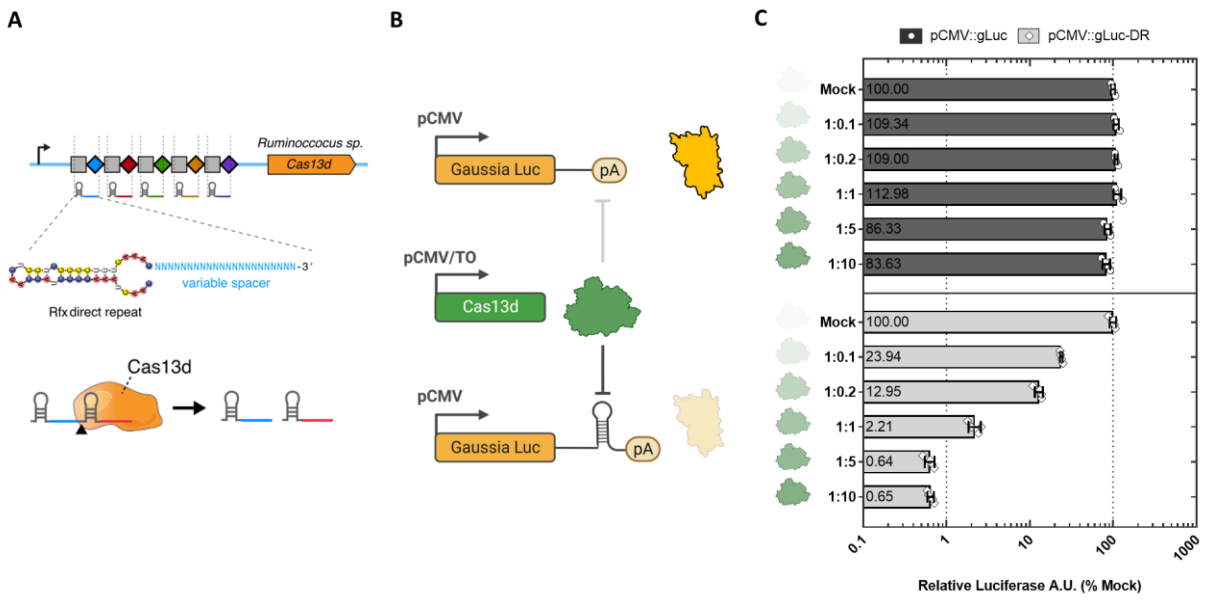


Figure 3.3 Experimental characterization CasRx-mediated post-transcriptional repression mechanism. A) Rfx-Direct Repeat (Rfx-DR) derived from the CRISPR array of Ruminococcus Flavefaciens Cas13d (adapted from (Konermann et al., 2018)). B) CasRx-mediated repression assay A pCMV/TO-driven CasRx can only cut a Gaussia Luciferase transcript bearing the Rfx-DR in its 3'UTR, causing its poly-A deprotection and degradation by cellular exonucleases. C) CasRx-mediated repression. The molar amount of the transfected CasRx-expressing plasmid with respect to the Gaussia Luciferase target is negatively correlated with luciferase expression. Ten-fold molar increase corresponds to one order of magnitude in luciferase repression.

As a measure of reporter activity, Luciferase Arbitrary Units (A.U.) were calculated as the ratio between the Gaussia and Red Firefly Luciferase values, and the relative luciferase A.U. were obtained as a percentage of the luciferase A.U. value in the mock condition. The bar graph in Figure 3.3 C shows that CasRx post-transcriptionally represses the reporter gene only if its transcript harbors a direct repeat. Moreover, these data prove that CasRx is an efficient post-transcriptional repressor that lowers target gene expression by 20% at a low molar ratio (1:0.1) and decreases it by two orders of magnitude at a higher molar ratio (1:5) (Figure 3.3C). Together, these results corroborate the post-transcriptional repression strategy based on CasRx as a reliable solution to implement Y-mediated repression of the main output Z in the CFFL4 topology.

3.4 Experimental implementation and characterization of the double-repression branch of the CFFL4 topology (X--|Y--|Z).

Litovco et al. showed that the strength of the intermediate output Y-mediated repression of the main output Z is a key parameter that determines the fold-change activation level of the CFFL-4 system. In particular, as shown in **Figures 3.4A**, in silico simulations of the CFFL4 circuit by Litovco et al., 2021, show that an increase in the strength of the Y-mediated repression of Z is positively correlated with the fold-change activation level of the output Z. However, in their study, activation of the main input X is assumed to completely inhibit Y. This assumption, however, does not apply to my system, in which I demonstrated that pCMV/TO can only partially be inhibited by the rtTA3G in the presence of doxycycline. Therefore, I performed an experiment as depicted in **Figure 3.4B**. Specifically, the assay consists of a cascade repression in which rtTA3G (pCMV-rtTA3G, representing species X) transcriptionally inhibits the CasRx endonuclease through the pCMV/TO promoter (pCMV/TO-CasRx-T2A-mCherry, representing species Y), which in turn post-transcriptionally represses a constitutively expressed Gaussia luciferase reporter (pCMV-gLuc-DR, representing species Z). To tune the repression strength exerted by the CasRx, I decided to control the molar ratio of pCMV/TO-CasRx(NLS)-T2A-mCherry and pCMV-gLuc-DR plasmids (Figure 3.4 B, dial), as indicated by results obtained in **Figure 3.3C**.

Therefore, in the experiment reported in **Figure 3.4C**, HEK293T cells were transfected with plasmids encoding the parts described above, i.e. pCMV-rtTA3G, pCMV/TO-CasRx(NLS)-T2A-mCherry and pCMV-gLuc-DR plasmids, at two different relative concentration (1:1:5 and 1:5:5); to which I will refer as Low and High, representing CasRx-mediated low and high repression, respectively. I also included two negative "Mock" controls where the pCMV-rtTA3G plasmid was not transfected, corresponding to 0:1:5 and 0:5:5 relative concentrations, in which CasRx is fully expressed thus constitutively inhibiting Gaussia luciferase expression, thereby setting the system baseline responses. Finally, as a positive control, I transfected the pCMV-gLuc plasmid without the DR sequence at a 1:5:5 molar ratio

(MAX in Figure 3.4C), thus setting the system maximal response because it is insensitive to CasRx repression, and to which all other response conditions were normalized. This enabled me to obtain a clear readout of the dynamic range of the systems at different repression strengths. All conditions were tested at six doxycycline concentrations (0, 2.5, 10, 25, 100, and 1000 ng/ml).

The results of these experiments are reported in **Figure 3.4C**, where it is clear that increasing concentrations of doxycycline result in increasing levels of Gaussia Luciferase for two different concentrations of the Gaussia luciferase carrying plasmid relative to CasRx encoding plasmid (shown as green triangles and circles); this behavior can be explained by the doxycycline-mediated increase in rtTA3G activity resulting in higher transcriptional repression of the pCMV/TO promoter driving CasRx, and thus a reduced post-transcriptional cleavage of the Gaussia Luciferase 3'UTR. Thus, a positive correlation between Gaussia luciferase expression and doxycycline concentration is expected. Moreover, as expected, lower molar ratio of the Gaussia Luciferase and CasRx plasmids resulted in greater absolute luminescence (green triangles versus green circles in Figure 3.4C). Finally, the results for the negative "mock" controls in Figure 3.4C (red squares versus red rhombi) show that an increase in the amount of pCMV/TO-CasRx(NLS)-T2A-mCherry plasmid relative to that of the Gaussia Luciferase plasmid corresponds to a higher repression strength. Therefore, it is possible to tune the CasRx repression strength by controlling the molar ratio between pCMV/TO-CasRx(NLS)-T2A-mCherry and pCMV-gLuc-DR plasmids (Figure 3.4 B, dial)

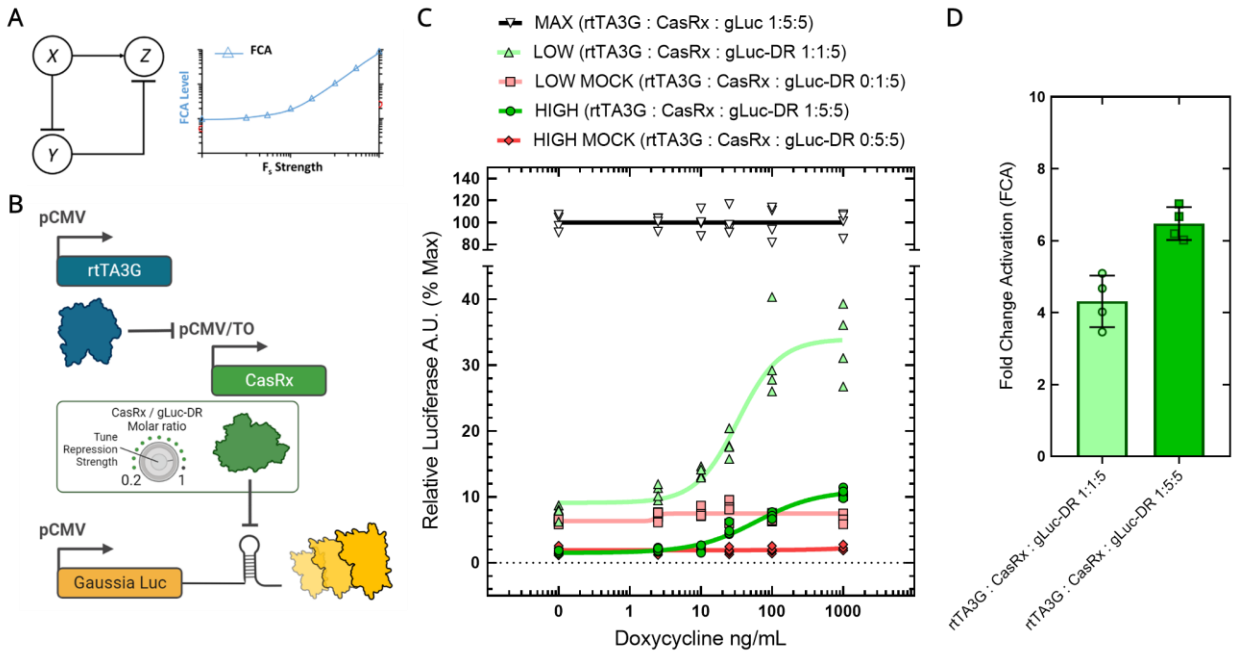


Figure 3.4 pCMV/TO-mediated inhibition of CasRx repression and characterization. A) Fold-change activation levels positively correlate with Y repression strength in an ideal CFFL-4 topology (in silico simulation), where X-mediated inhibition of Y is complete (adapted from (Litovco et al., 2021)) B) Scheme of the experiment to assess CMV/TO-mediated inhibition of CasRx repression. In the absence of doxycycline, CasRx is constitutively expressed by the pCMV/TO promoter, causing full repression of the gLuc target gene bearing the DR in its 3'UTR. In the presence of doxycycline, pCMV-driven rtTA3G binds to pCMV/TO, inhibiting CasRx expression and relieving the repression of the gLuc reporter gene. This results in an increase in luciferase activity. The CasRx repression strength can be tuned by regulating the molar ratio of the pCMV/TO-CasRx-T2A-mCherry and pCMV-gLuc-DR plasmids (dial). (C) pCMV/TO-mediated inhibition of CasRx repression. Doxycycline-induced repression relies on pCMV/TO-mediated inhibition of CasRx expression. Dark green curves: stronger CasRx repression; light green curves: weaker CasRx repression. Black curve: gLuc target plasmid without Direct Repeat; (D) Maximum fold-change activation of the Gaussian luciferase expression. Dark green: stronger CasRx repression; light green: weaker CasRx repression.

3.5 Construction and characterization of the CASwitch: a CasRx-based Coherent Feed Forward Loop Type-4 gene circuit in mammalian cells

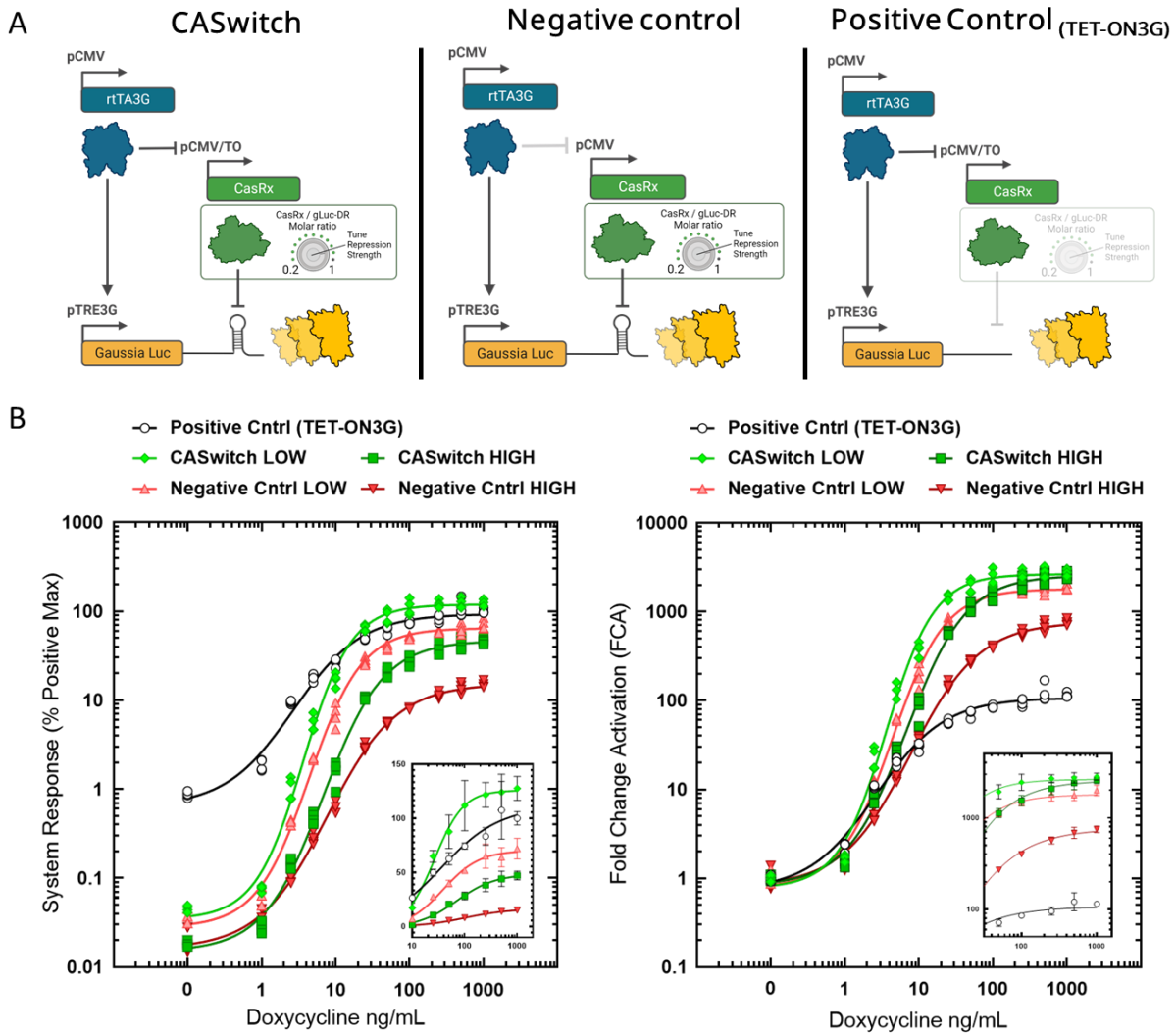


Figure 3.5 Composition and characterization of the CFFL-4 circuit implementation: The CASwitch. A) Schematic of the CASwitch characterization assay CASwitch is a CFFL-4 circuit topology implementation where the main input X is rtTA3G, the intermediate output Y is CasRx, and the main output Z is Gaussia Luciferase bearing the Rfx-DR in its 3'UTR. In the absence of doxycycline, constitutive pCMV/TO-expressing CasRx cleaves all gLuc leaky transcripts stemming from the pTRE3G promoter, quenching system leakiness. Tuning of CasRx repression strength was achieved by controlling the CasRx/gLuc-DR plasmid molar ratio (Dial). In the presence of doxycycline, rtTA3G transcriptionally activates gLuc expression through the pTRE3G promoter and inhibits CasRx expression through the pCMV/TO promoter, inducing maximum activation of the system. In the negative control, pCMV/TO was substituted with a pCMV promoter, causing doxycycline-insensitive constitutive expression of CasRx, and thus a constant repression of gLuc target expression. Tuning of CasRx repression strength was achieved by

controlling the CasRx/gLuc-DR plasmid molar ratio (Dial). CasRx has also been used in the Tet-On3G system to account for burden bias; thus, a gLuc target plasmid without DR was employed to avoid dampening the maximum response of the system. B) Normalised luminescence as a function of Doxycyclyne for the indicated circuits. Inset: plots in linear scale to better show the changes for increasing levels of doxycycline.

The CASwith, my CFFL-4 gene circuit implementation, consists of three plasmids, as schematized in **Figure 3.5A**:

1. pCMV-rtTA3G plasmid, constitutively expressing the rtTA3G doxycycline responsive transcriptional activator and representing the CFFL-4 circuit's main input X
2. pCMV/TO-CasRx(NLS)-T2A-mCherry, encoding a doxycycline transcriptionally repressible CasRx endonuclease and acting as the CFFL-4 circuit's intermediate output Y (Chapter 3.2 and 3.4)
3. pTRE3G-gLuc-DR, representing the CFFL-4 circuit's main output Z, where the Gaussia Luciferase reporter gene expression is transcriptionally activated by the rtTA3G in the presence of doxycycline, and post-transcriptionally repressed by the CasRx through the Direct Repeat-mediated mRNA degradation (Chapter 3.3)

I transfected HEK293T cells with these three plasmids (pCMV-rtTA3G : pCMV/TO-CasRx(NLS)-T2A-mCherry : pTRE3G-gLuc-DR) at a 1:5:5 and 1:1:5 molar ratio respectively; for simplicity of notation, I will refer to the 1:5:5 molar ratio as "CASwitch High," and to the 1:1:5 molar ratio as "CASwitch Low". As negative controls, cells were transfected with a modified version of the CasRx plasmid which this time is driven by a constitute pCMV promoter that thus cannot be repressed by the rtTA3G as shown in Figure 3.5A, at the same two molar ratios used for the CASwitch. Finally, as a positive control, cells were transfected with a modified version of the Gaussia Luciferase plasmid where the Gaussia Luciferase lacks the DR in its 3' UTR (pTRE3G-gLuc) at a single molar ratio of 1:5:5. This positive control is essentially the classic state-of-the-art Tet-On3G gene expression system.

All conditions were tested at increasing doxycycline concentrations (0, 1, 2.5, 5, 10, 25, 50, 100, 250, 500, and 1000 ng/ml). In addition, all experimental conditions were transfected with a Red Firefly reporter plasmid, employed as a transfection efficiency normalizer, and

with a pcDNA3.1 empty plasmid to transfect the same total amount of DNA between all conditions. Arbitraty Luciferase units (A.U.) were calculated as the ratio between the Gaussia and Red Firefly Luciferase values; these relative luminescence values are shown in **Figure 3.5B** as a percentage of the luciferase activity at 1000 ng/mL of doxycycline in the Positive control. As shown in **Figure 3.5B**, both CASwitch Low and CASwitch High in the absence of doxycycline exhibited a greatly reduced luminescence level, hence a much lower leakiness, when compared to the positive control (Tet-On3G system) (black circles versus green rhombi and squares), with the "CASwitch High" exhibiting a lower luminescence level in the absence of Doxycycline than the one exhibited by the "CASwitch Low". Of note, the "CASwitch Low" reaches a maximum luminescence level similar to that of the positive control for 1000 ng/mL of Doxycycline (Figure 3.5B). Both negative control curves (Negative Low and Negative High) were characterized by an overall lower system response when compared to their cognate CASwitch Low and CASwitch High (red curves versus green curves), thus proving the pCMV/TO inhibition contribution to CFFL-4 system performance. Indeed, CASwitch systems show the same leakiness values as the negative controls but achieve a much higher doxycycline-induced expression. Furthermore, in **Figure 3.5B** (right panel), I calculated the fold change activation (FCA) values, which confirmed my hypothesis that in my implementation, a stronger repression does not correspond to higher fold change levels, as demonstrated by the "CASwitch High" response curve having lower leakiness value but also lower Fold Change Activation when compared to the "CASwitch Low". The FCA graph again shows the contribution of pCMV/TO implementation to CFFL-4 systems performance, endowing them with overall higher FCA values than their cognate negative controls.

Moreover, I observed a different shape of the dose-response curve among the CASwitch systems, Tet-On3G system, and negative controls, where CASwitch Low and High seemed to have steeper responses to doxycycline. For this reason, I plotted the Luciferase A.U. values as percentage of their respective maximum values and calculated the effective concentration (EC50) and Effective Hill Coefficient (H_{eff}) parameters, as reported **in Figure**

3.6. The results of this calculation are shown in **Figure 3.6A** where CASwitch Low activates at a lower concentration compared to TET-ON3, suggesting an increased sensitivity to doxycycline with a doubled effective Hill Coefficient. In their seminal paper, Mangan & Alon showed that comparing CFFL-1 and CFFL-4 topologies, only the first endowed a system with a higher effective hill coefficient (Mangan & Alon, 2003) (figure 3.6B); however, here I report a different result, showing that my CFFL-4 topology implementation endows the Tet-On3G system with increased sensitivity doubling its effective Hill coefficient (figure 3.5.2 A).

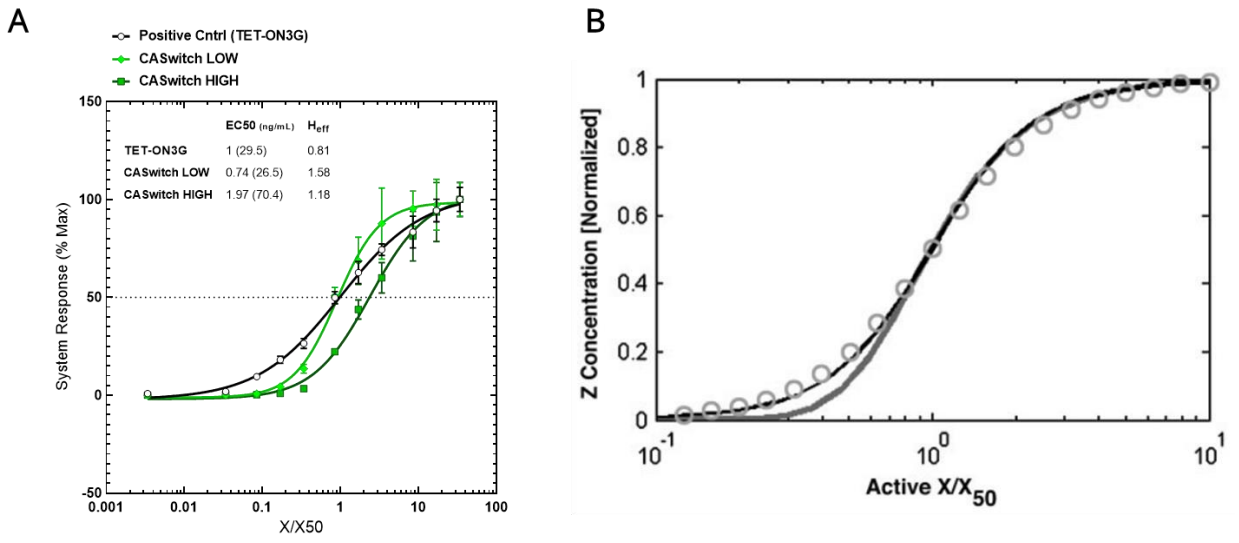


Figure 3.6 CASwitch shows an increased effective Hill coefficient. A) Normalized CASwitch system responses to its maximum value as a function of doxycycline activity (X/X50). The graph shows the response curve for CASwith a weaker CasRx repression strength (light green line), stronger CasRx repression (dark green line), and the Tet-On3G system. Both CASwitch systems shoId an increased effective Hill Coefficient compared to that of the Tet-On3G system, suggesting a steeper response to doxycycline. B) Apparent cooperativity of the steady-state Z response as a function of the X activity. The graph shows the z(x) response curve for type 1 (thick line), type 4 (thin line) FFLs, and a simple regulation system (○), adapted from (Mangan & Alon, 2003).

Finally, I carried out a "stress test" on the CASwitch system to check whether it retains the same features when using alternative promoters to express the rtTA3G. Thus, I tested the CASwitch using three different promoters with increasing expression strength, with pPGK being the weakest, pEF1a the one with intermediate expression power, and pCMV being the strongest. The results are shown in **Figure 3.7**, where it can be appreciated that the

CASwitch retains its features for all the tested promoters. However, when using the weakest pPGK promoter caused the CASwitch's maximum response to stand at approximately half of that of the Tet-On3G system. The explanation for this behavior, compared to high- and medium-expression promoters (pCMV and pEF1a), might be that weak promoters do not express enough rtTA3G to efficiently block pCMV/TO, resulting in residual CasRx expression, which dampens the maximum response of the CASwitch system. It is likely that in a relatively low rtTA3G scenario, the pTRE3G and pCMV/TO promoters compete for rtTA3G binding, where pTRE3G, having more TetO sites, is more responsive to low amounts of rtTA3G. However, CASwitch retained high fold-change activation values compared to the Tet-On3G system in all cases.

Finally, normalizing the system responses to their maximum (Figure 3.7 right panels), I still observed an increase in the effective Hill coefficient for all the tested promoters. Together, these data demonstrate the broad applicability of the system, especially for ligand responsive promoter-mediated biosensing applications, where a specific molecule-responsive promoter can be used to drive rtTA3G. The application of the CASwitch system will result in increased fold-change activation and enhanced sensitivity.

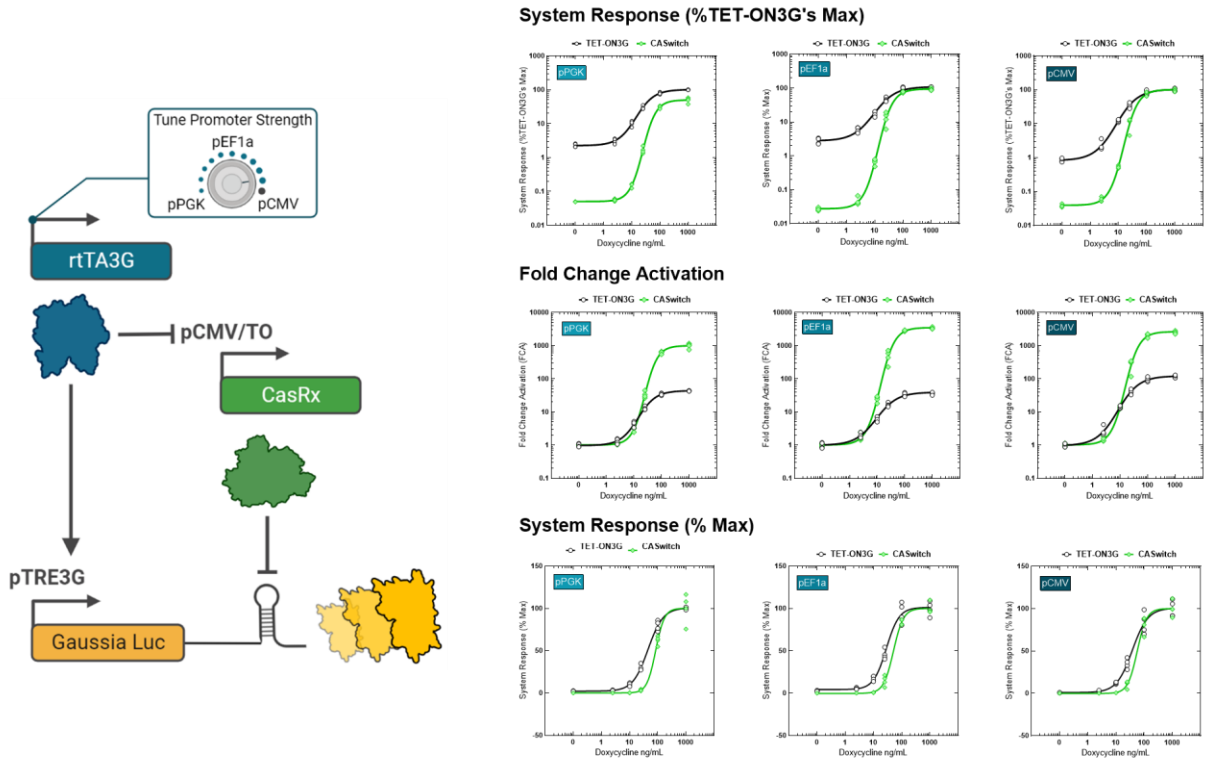


Figure 3.7 CASwith promoter transcription strength stress test. A) Stress test composition. Three different promoters with as many transcriptional strengths as possible were used to drive the expression of the main input rtTA3G. pPGK, weak; pEF1a, medium; pCMV, strong transcription (power dial). B) Left, CASwitch, and Tet-On3G systems' responses to doxycycline normalized to Tet-On3G's maximum response; Center, CASwitch, and Tet-On3G fold-change activation to doxycycline; Right, CASwitch and Tet-On3G systems' responses to doxycycline normalized to their respective maximum responses. Dark line: Tet-On3G, Light green Line: CASwitch.

3.6 Conclusion

Overall, the strategies I set forth to implement the CFFL-4 circuit topology proved to be valid and reliable. I called this CFFL-4 gene circuit the CASwitch because it is an improved gene expression control system able to "switch" expression of a gene of interest in an inducible manner.

The CASwitch system is characterized by at least one magnitude lower leakiness compared to the state-of-the-art Tet-On3G inducible gene expression system, while maintaining a high maximum response. Together, these characteristics result in more than one magnitude

higher fold-change activation at saturating concentrations of doxycycline. Furthermore, the CFFL-4 implementation endows the CASwitch with a steeper response to doxycycline as shown by higher effective Hill coefficient. I successfully employed the CASwitch system in two proof-of-principle applications.

In the first one, I harnessed the CASwitch as dynamic range amplifier of pre-established biosensors which dynamic ranges were limited by the low transcriptional strength of their responsive promoters. In this implementation of the CASwitch, the rtTA3G driven by the responsive promoter amplifies output reporter expression, enabling detection of small changes in promoter activity. On the contrary, CasRx buffers against noise amplification and reduce basal activity when the promoter is inactive, overall improving the signal-to-noise ratio and the dynamic range of the transcriptional-based biosensor.

In the second implementation, I used the CASwitch to tightly control the expression of viral toxic genes required to produce Adeno Associated Vectors (AAV) in HEK293 cell line, AAV vector inducible production. In this case, the CasRx post-transcriptionally represses toxic viral genes expression without compromising essential viral genome structure. Instead, in the presence of doxycycline the rtTA3G-mediated inhibition of the CasRx, and transcriptional activation allow high expression of viral genes required to produce AAV vectors. This achievement will grant the the generation of an AAV packaging HEK293 cell line.

Chapter 4

Amplified Biosensing harnessing the CASwitch System

4.1 Whole-cell biosensors

Whole-cell biosensors (WCBs) are devices that can sense and report on bioactive molecules, chemicals, or even pathogens relevant to human health that are present within the cell or in the surrounding microenvironment, introducing a cheap, portable, and simple methods for detecting molecules of interest (Lim et al., 2015). Applications range from toxicology and drug discovery, to biomanufacturing and biosecurity. For human health WCBs are appealing especially as alternatives to animal models. In this context, WCBs can be used in cell-based assays for Phenotypic drug discovery (PDD) aiming to identify drug candidates that modulate a physiologically-relevant biological system or cellular signaling pathway and screen for drug efficacy (Beitz et al., 2022). WBCs-based assays are target-agnostic and provide the opportunity to identify therapeutics that resolve disease-associated phenotypic deficits even when the disease etiology remains obscure, and targets are undefine. In order to identify drug candidates that translate to effective therapeutics in the clinic, WCBs-based assays that faithfully report on disease-relevant processes are required. Often, WCBs are transcription-based biosensor obtained by engineering a promoter regulated by a transcription factor whose activity is modulated by a cellular signaling pathway. However, transcription-based WCBs are not efficient as they have limited linear and dynamic range, and exhibit “leakiness” because of a non-zero basal expression. Moreover, developing a

biosensor requires a cumbersome trial-and-error approach as initial prototypes usually perform poorly. Hence, deploying a WCB that can identify toxic compounds for toxicology, or novel molecules that translate to effective therapeutics in the clinic remains a substantial challenge. The CASwitch technology would yield a fast-to-implement strategy for biosensor design by simply using the analyte-sensitive promoter in a novel configuration endowing the resulting biosensor with greatly improved performances.

4.2 Dynamic range amplification of an intracellular copper biosensor

Copper (Cu) is a trace element essential for all organisms that utilize oxygen as an electron acceptor during respiration. As a transition metal, Cu serves as a cofactor for several redox enzymes. The same redox-active properties also render the metal potentially toxic as it can induce free radical formation and cause direct damage to proteins, lipids, and DNA. To maintain copper homoeostasis, remarkably efficient mechanisms have evolved to regulate copper import, export, and cellular distribution. Severe disorders may arise as a result of disruption of copper homoeostasis. Recessive mutations in the gene encoding ATP7A (copper-transporting P1B-type ATPase) result in Menkes disease, a fatal X-linked neurodevelopmental disorder. Similarly, recessive mutations in ATP7B result in Wilson disease, which is characterized by a combination of liver disease and neurological and psychiatric problems that affect children and young adults (Van Den Berghe et al., 2007). Develop a WCB featuring an amplified report on intracellular concentration of copper can enable the development of reliable cell-based assays that can identify molecules able to reestablish copper homeostasis, which can be further developed into drugs capable of counteracting the symptoms of copper accumulation-related diseases. Moreover, the investigation of hit compounds mechanism of action could pave the road to new insights about those diseases and, in general, about copper homeostasis regulation. Therefore, I sought to amplify the dynamic range of a pre-developed transcriptional-based copper biosensor by harnessing the CASwitch technology, as an easy and fast-to-implement strategy to develop novel WCBs.

Engineering the CASwitch-amplified copper biosensor

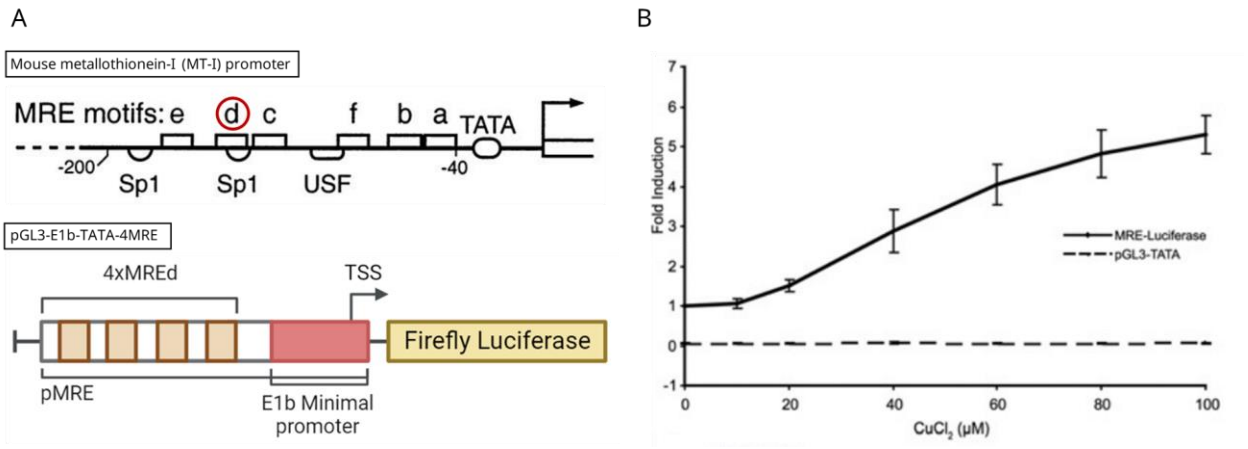


Figure 4.1 Original pMRE biosensor performance. A) Composition of the Wild-type Mouse MT-I and of its synthetic derivative: the pMRE promoter. The MT-I promoter contains six metal response elements (MRE a to f), binding sites for Sp1 and USF (adapted from (B. Zhang et al., 2003)). Synthetic pMRE promoter construct with four equidistant MREd sites upstream of an Adenoviral E1b minimal promoter driving the Firefly Luciferase reporter gene (pGL3-E1b-TATA-4MRE). B) Fold-change activation of pMRE copper biosensor at increasing concentration of copper. pGL3-E1b-TATA-4MRE or the control vector (pGL3-TATA) ware transiently transfected in HEK293T cells. After transfection, cells were incubated for 24 h with different concentrations of CuCl2. (adapted from (Van Den Berghe et al., 2007)

To construct the CASwitch-amplified biosensor, I started improving on a previously established copper biosensor developed by Van Den Berghe et al. (Van Den Berghe et al., 2007). This transcription-based copper biosensor features a synthetic metal-responsive promoter called pMRE, which composition is shown in Figure 4.1A. It contains four copies of the strongest MREd element (4xMREd) derived from the mouse metallothionein-I (MT-I) wild-type promoter (B. Zhang et al., 2003), followed by the Adenoviral E1b minimal promoter, endowing the pMRE with the TATA binding box and the transcription start site. Each MREd element contains a consensus binding site "TGCRCNC" bound by the metal response element binding transcription factor 1 (MTF-1) in the presence of zinc (Zn), copper (Cu), or cadmium (Cd). In the copper biosensor by Van Den Berge et al., pMRE drives the

expression of a Firefly Luciferase reporter gene in a copper dose-responsive manner, reaching a maximum of five-fold induction at 100 μM of $CuCl_2$ (Figure 4.1B). Therefore, to amplify the dynamic range of this copper biosensor, I plugged the pMRE promoter "copper sensor module" into the CASwitch dynamic range amplifier circuit. As shown in **Figure 4.2A**, in this configuration the pMRE drives the expression of the rtTA3G in a copper-dependent manner. In the presence of doxycycline (1000 ng/mL), rtTA3G is constitutively active and fully induces the expression of the firefly luciferase through the pTRE3G promoter, thus, a direct correlation between the intracellular concentration of copper (i.e. pMRE activation) and the luciferase expression is established, resulting in the amplification of the output in the CASwitch-amplified copper biosensor. Instead, in the absence of copper there is no rtTA3G expression meaning no inhibition of the CasRx, hence allowing reduction of the pTRE3G promoter basal expression (i.e. leakiness) through the DR-mediated post-transcriptional degradation mechanism. Because the amplification of luciferase expression can be achieved by the Tet-On3G as well, by means of the rtTA3G positive regulation of the firefly luciferase expression through the pTRE3G promoter, I chose to compare Tet-On3G amplification performance to the one of the CASwitch system (Figure 4.2 A versus B). Tet-On3G system has no means for leakiness reduction, thus it should merely increase the absolute luciferase expression without achieving any dynamic range amplification. CASwitch-amplified and Tet-On3G-amplified biosensors response and fold-induction to increasing copper concentrations have been compared to those ones of the original biosensor as shown in Figure 4.2 D and E.

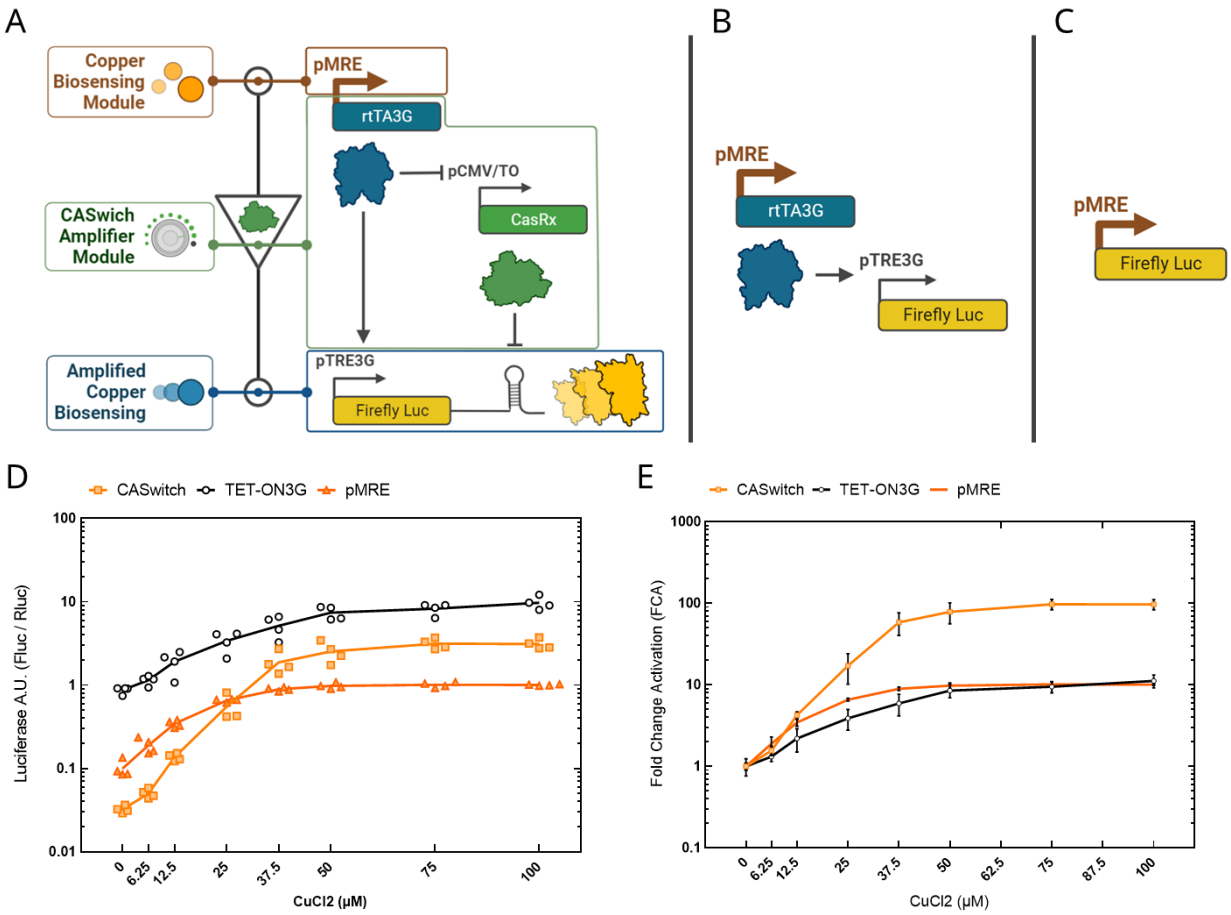


Figure 4.2 CASwitch-amplified, Tet-On3G-amplified, and original copper biosensors performances compared. A) CASwitch-amplified copper biosensor composition. The pMRE promoter represents the "Copper sensing module" plugged into the CASwitch "Amplifier module" where the pMRE drives the expression of the rtTA3G. The rtTA3G expression directly correlate with pMRE promoter activation which depends on the intracellular copper concentration. In turn, in the presence of doxycycline, the rtTA3G expression directly correlates with firefly luciferase expression in the CASwitch system, resulting in the amplification of the output in the CASwitch-amplified copper biosensor. B) Tet-On3G-amplified copper biosensor. The pMRE drives the expression of the rtTA3G, which in turn regulate the expression of the firefly luciferase through the pTRE3G promoter in the presence of doxycycline. C) Original pMRE copper biosensor regulating the firefly luciferase reporter gene. D) CASwith-amplified (light orange curve), Tet-On3G-amplified (black curve), and original pMRE copper (light orange curve) biosensors dose response curves at increasing concentration of $CuCl_2$ (µM). Data points of n=4 technical replicates. E) CASwith-amplified (light orange curve), Tet-On3G-amplified (black curve), and original pMRE copper (light orange curve) fold induction to $CuCl_2$ (µM). Mean and s.d. of n=4 technical replicates.

Indeed, the Tet-On3G-amplified dose-response curve (black circles) in Figure 4.4D represents the upward vertical scaling of the one of the original pMRE biosensor (dark orange triangles), do not achieving any fold-induction amplification as depicted in Figure 4.2F. Instead, the CASwitch-amplified biosensor dose-response curve (light orange cubes), although not reaching the same maximum luciferase expression of the Tet-On3G system, is characterized by having the lowest leakiness and an higher maximum luciferase expression than the pMRE original biosensor, resulting in enhanced fold-induction values immediately after a $CuCl_2$ concentration of 12.5 μM, as shown in Figure 4.2 D and E. The maximum response dampening observed in the CASwitch-amplified biosensors compared to the Tet-On3G one was expected (black circles versus orange cubes at $CuCl_2$ 100 μM) since I observed the same outcome when using the low-strength promoter pPGK, meaning that the maximum amount of rtTA3G expressed by a saturated pMRE promoter is not enough to completely inhibit the CasRx through the pCMV/TO promoter, resulting in a dampened maximum response in the CASwitch-amplified system.

Even that, the CASwitch amplification strategy outperformed the Tet-On-3G one, endowing the original copper biosensor with a final fold-induction value higher by one order of magnitude at 100 μM of copper. Moreover, the dynamic range amplification obtained harnessing the CASwitch amplifier circuit endows the amplified copper biosensor with increased detection at lower copper concentrations compared to the Tet-On-amplified and original pMRE biosensors. Overall, these results demonstrate a reliable implementation of the CASwitch system as a tool to enhance the performance of an established copper biosensor, supporting the vision of the CASwitch-mediated dynamic range amplification as a fast-to-implement starting strategy to develop novel WCBs-based assays.

To further prove the broad applicability of the CASwitch system, I decided to employ it to amplify the fold-induction of another promoter workings as cellular lysosomal stress biosensor previously developed in the diBernado's laboratory by Dr. Nicoletta Moretti: the pNiClear. (Unpublished data)

4.3 Dynamic range amplification of a lysosomal stress biosensor

Lysosomes are organelles that are central to the degradation and recycling processes in animal cells. Lysosomal biogenesis and function are regulated by the master regulator Transcription Factor EB (TFEB), which is involved in the coordinated transcriptional control of a network of genes, called the coordinated lysosomal expression and regulation (CLEAR) network. Genes in this network are characterized by having promoters containing a palindromic 10–base pair “GTCACGTGAC” motif, called CLEAR site, bound by TFEB following amino-acid starvation, aberrant lysosomal storage or other lysosomal stress conditions (Sardiello et al., 2009). TFEB controls multiple aspects of lysosomal function beyond biogenesis, including autophagy and lysosomal exocytosis. The role of TFEB in lysosomal function control can be exploited to promote cellular clearance in several disease conditions. This approach has been tested in several cellular and mouse models of human diseases resulting from the accumulation of undegraded substances, such as lysosomal storage diseases (LSD), Parkinson's disease, Alzheimer's disease, and diet-induced obesity (Puertollano et al., 2018).

Engineer a WCB featuring an amplified reporting on lysosomal stress can enable the development of reliable cell-based assays that can identify molecules able to control and induce lysosomal clearance of undegraded molecules, which can be further developed into drugs capable of counteracting the symptoms lysosomal storage diseases and neurodegenerative diseases.

Engineering the CASwitch-amplified lysosomal stress biosensor.

The pNiClear promoter is a TFEB-responsive synthetic promoter composed of seven repeats of the CLEAR site “GTCACGTGAC” upstream of a CMV minimal promoter, as shown in Figure 4.3. This promoter can be used to drive the expression of a firefly luciferase reporter gene, establishing a lysosomal stress biosensor responsive to TFEB activation. Unfortunately, this biosensor is unable to sense the activation of endogenous TFEB, but responds only to exogenous TFEB overexpression, regardless of the cell lysosome stress status (unpublished

data). This due to the very low expression level of endogenous TFEB and probably due its low transcriptional activation capacity, which makes of the pNiClear a very weak promoter, unable to detect physiological changes in TFEB activity. Therefore, I decided to amplify the dynamic range of the pNiClear promoter harnessing the CASwitch system to endow the original pNiClear lysosomal stress biosensor with the ability to detect endogenous TFEB activity physiological changes.

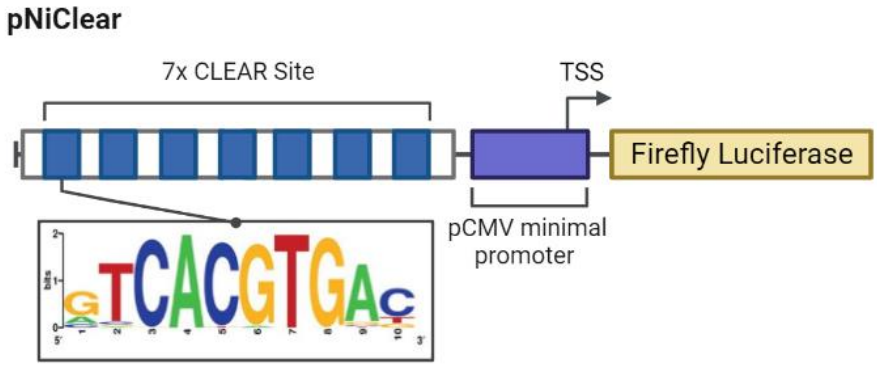


Figure 4.3 Schematic of pNiClear promoter. The pNiClear promoter developed by Dr Nicoletta Moretti in the di Bernardo's lab consists of a construct with 7 CLEAR sites upstream of a CMV minimal promoter driving the expression of the firefly luciferase reporter gene.

I plugged the pNiClear promoter "sensor module" into the CASwitch circuit "amplifier module", allowing the pNiClear to drive the expression of the rtTA3G, as shown in Figure 4.4A. In this configuration, in the presence of doxycycline there will be a direct correlation between the activity of the pNiClear promoter and the expression of the firefly luciferase through the constitutively active rtTA3G. This correlation that leads to luciferase expression amplification also applies to the Tet-On3G system (Fig. 4.4B). For this reason, I decided to compare the performance of the CASwitch-amplified lysosomal stress biosensor to that one of the Tet-On3G-amplified. In order to induce the pNiClear promoter, the TFEB-inducing small molecule Torin-1 was used to activate endogenous TFEB. Torin-1 is a potent and selective ATP-competitive inhibitor of mTOR (mammalian target of Rapamycin) kinase. Inhibition of mTOR cause inhibition of TFEB phosphorylation; dephosphorylated TFEB translocates into the nucleus where it transcriptionally activates its target genes. CASwitch-

amplified and Tet-On3G-amplified biosensors' response and fold-induction to increasing Torin-1 concentrations have been compared to those ones of the original biosensor as shown in Figure 4.4 D and E.

Similar results to the copper implementation were observed. The Tet-On3G-amplified dose response curve (black circles) represents the upward vertical scaling of the pNiClear original biosensor one (dark blue triangles in Fig. 4.4 D), resulting in no fold-induction amplification as depicted in Figure 4.4 E. Instead, the CASwitch-amplified biosensor dose response curve (light blue cubes in Fig. 4.4 D) is characterized by having the lowest leakiness value and the maximum luciferase expression similar to that one of the original pNiClear biosensor at saturated concentration of Torin-1 (light blue cubes versus dark blue triangles at 1000 μM of Torin-1 in Fig 4.4 D). Even though the CASwitch-amplified biosensor has an always lower luciferase expression compared to the original pNiClear biosensor, it manages to obtain higher fold-induction respect to the other biosensors, immediately after 100μM of Tornin-1 (light blue cubes Fig. 4.4 E). In this implementation, the CASwitch circuit yielded a weaker dynamic range amplification, achieving a two-times increase of the original pNiClear fold induction value at maximum concentration of Torin-1 (light blue cubes versus dark blue triangles Fig. 4.4 E), compared to the one order of magnitude fold induction increase obtained in the copper implementation. This amplification discrepancy might be explained by the absolute transcriptional strength difference of the two implemented promoters, where the pMRE is a stronger promoter than the pNiClear, which confirms to be a very weak promoter. Indeed, comparing the relative luciferase expression dose-response curves of the CASwitch-amplified biosensors between the two implementations (light orange cubes in Fig 4.2D versus light blue cubes in Fig. 4.4E) suggests a weaker pCMV/TO-mediated inhibition of the CasRx in the pNiClear application with an overall low luciferase expression for the CASwitch-amplified lysosomal stress biosensor probably caused by very low amount of expressed rtTA. In this very low rtTA3G expression context, the seven TetO-binding sites on pTRE3G could have prevailed, eventually subtracting the already scarce pool of rtTA3Gs

from the two TetO-binding sites of pCMV/TO, resulting in an insufficient inhibition the CasRx.

These results highlight the need for optimization of the pCMV/TO-mediated transcriptional inhibition of the CasRx in CASwitch circuit when using very weak promoters. Even that, also in this implementation of the pNiClear promoter the application of the CASwitch dynamic range amplifier circuit resulted in fold induction amplification, supporting the vision of the CASwitch of a broadly applicable tool to enhance the performance of an established biosensor.

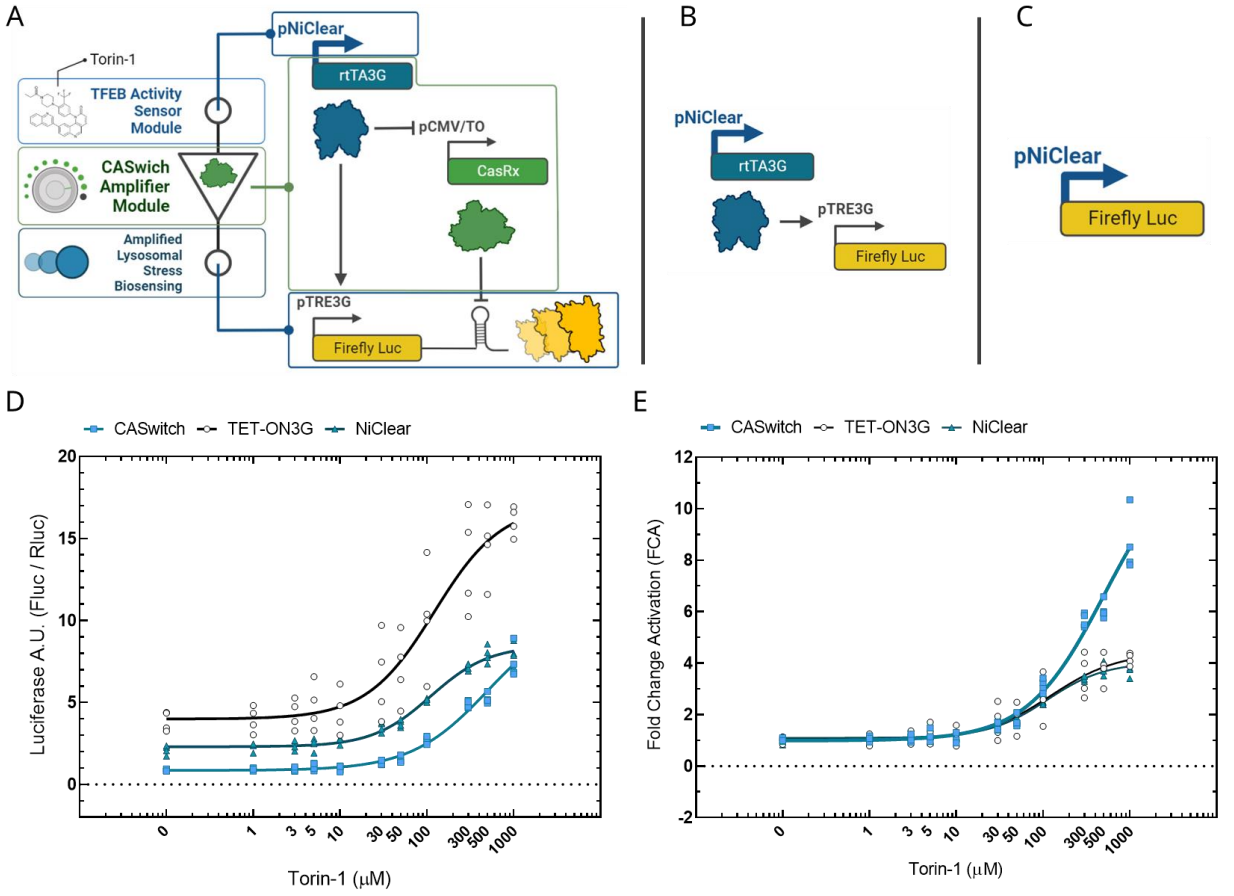


Figure 4.4 Schematic of the CASwitch-amplified lysosomal stress biosensor and its response to copper. A) CASwitch-amplified lysosomal stress biosensor composition. The pNiClear promoter represents the lysosomal stress sensing module plugged into the CASwitch amplifier module by letting the pNiClear drive the expression of the rtTA3G in the CASwitch system. This, in saturated doxycycline concentration condition, allows to directly correlate the lysosomal stress-induced TFEB activity sensed by the pNiClear promoter to the firefly luciferase

expression, representing the amplified output of the CASwitch-amplified lysosomal stress biosensor. B) Tet-On3G-amplified lysosomal stress biosensor. There the pNiClear drives the expression of the rtTA3G, which in turn regulate the expression of the firefly luciferase through the pTRE3G promoter. C) Original pNiClear lysosomal stress biosensor directly expressing the firefly luciferase reporter gene. D) CASwith-amplified, Tet-On3G-amplified and original pNiClear lysosomal stress biosensors' responses to Torin-1. E) CASwith-amplified, Tet-On3G-amplified and original pNiClear lysosomal stress biosensors' fold-change activation to Torin-1.

Chapter 5

Application of the CASwitch for recombinant Adeno-Associated Vector (rAAV) production.

The CASwitch system can be used as an improved inducible gene expression system in mammalian cells that goes beyond the state of the art thanks to the its “tight” gene expression control system guaranteeing a very low leakiness, large dynamic range and high maximum expression values.

As previously described in Chapter 2, the shortcomings of current recombinant adeno-associated vector (rAAVs) manufacturing technologies is a pressing problem preventing wide adoption of AAV-based drugs in the clinical setting. Indeed, current attempts to develop a new manufacturing method based on the standard HEK293 cell line are falling short, mostly due to the extreme toxicity of AAV and Helper viral genes after stable genome integration (Qiao, Li, et al., 2002). Therefore, to allow for stable genome integration of these viral genes and to avoid their toxicity, I employed the CASwitch system to tightly control their expression with the ultimate goal of enabling the generation of an rAAV HEK293 inducible producer cell line, with a huge impact on the development of novel rAAV-based gene therapy.

5.1 Design of a CASwitch-compatible gene set for inducible rAAV vector production

An inducible rAAV producer HEK293 cell line should harbor integrated and rescuable transgene DNA, AAV Rep and Cap genes, as well as essential helper genes from other helper viruses. To enable the expression control of this set of genes by the CASwitch, the first step was to re-engineer these genes to be compatible with the post-transcriptional "leakiness-quenching" mechanism exerted by the CasRx. Therefore, I first identified the Helper genes that are strictly required for rAAV production and included them into a bicistronic transcriptional unit making it compatible with the CASwith system by adding the DR sequence in its 3'-UTR. Similarly, I also modified the AAV-2 Rep and Cap genes by adding in the DR in their shared 3'UTR. Additionally, these engineered transcriptional units were placed downstream a pTRE3G promoter, allowing for doxycycline-induced gene expression in the CASwitch system, as schematically shown in Figure 5.3A. In what follows, I will describe in more detail all the steps I performed to design the CASwitch-based rAAV production system.

Human Adenovirus 5 minimal helper gene set characterization: design of a μHelper plasmid

The adeno-associated virus (AAV) belongs to the Dependoparvovirus family, which means that it depends on helper factors for replication. These helper factors can be provided by co-infection of helper viruses such as adenoviruses, herpesviruses, and papillomaviruses. Human Adenovirus type 5 (HAdV5) is the most studied AAV helper virus, and its genes are used as helper factors in most established rAAV manufacturing protocolos. HAdV5 genes that are necessary and sufficient to sustain rAAV vector production are five: E1A, E1B, E2A, E4, and VARNA (Viral Associated RNA) (Balakrishnan & Jayandharan, 2014; Geoffroy & Salvetti, 2005; Meier et al., 2020). Generally, they are required for efficient AAV gene

expression, DNA replication, and packaging. The E1A and E1B genes are stably expressed in HEK293 cells. Instead, E2A, E4, and VA RNA should be exogenously supplied to induce the production of AAV vectors. In the transient triple transfection manufacturing method, these genes are supplied by a Helper plasmid, thus substituting the Human Adenovirus 5 for its helper functions. The E2A, E4, and VARNA genes are transactivated by E1A in the HEK293 cell line meaning that the first challenge to stably integrate these genes is that it will result in their constitutive toxic expression, thereby causing cell death (Figure 5.1A).

One solution could be to decouple these genes from the E1A-mediated transcriptional regulation by replacing their wild-type promoters with synthetic promoters. However, E2A and E4 are both polycistronic genes. The E2A gene is transcribed from two different promoters (E2-early and E2-late) and subjected to distinct splicing patterns, giving rise to two E2A DNA Binding Protein (DBP) isoforms with alternative N-terminal protein sequences (Cooper et al., 1995; Donovan-Banfield et al., 2020; Zajchowski et al., 1985) (Figure 5.1A). Instead, E4 encodes for seven different proteins (ORF1, ORF2, ORF3, ORF3/4, ORF4, ORF6, and ORF6/7) through a complex pattern of differential splicing that produces at least 18 distinct mRNAs with different 3'-terminal sequences (Täuber & Dobner, 2001) (Figure 5.1A). Hence, modifying the promoters of E2A and E4 genes would not be the right solution to engineer an rAAV producer cell line in my case, because it would not be suitable for the CASwitch implementation since E4 gene transcripts do not share the same 3'-UTRs. Moreover, the wild-type E2A and E4 genes further encode for several other viral products on the opposite strand with unknown functions for rAAV production (e.g. L4, L3, E3 genes' products in Figure 5.1A), which could raise safety concerns in the long run. Therefore, I decided that the most suitable and optimal solution was to select only the gene products of E2A and E4 required to sustain rAAV vector production in order to rationally implement them in the CASwith system and to tightly control their toxic expression.

Therefore, I investigated in the literature the adenoviral gene selection process that resulted in the creation of the adenoviral Helper plasmid and in the establishment of the Helper virus-free transient transfection production method. Currently, two adenovirus

helper plasmids are used to manufacture rAAV. Plasmid pXX6 was developed by Xiao et al. and contains the E2A, VA-RNA, and E4 genes (Xiao et al., 1998), as well as a similar construct developed by Matsushita et al., harboring the same genes (Matsushita et al., 1998). In particular, Matsushita et al. have shown that the gene products E2A(DBP) and E4(Orf6), together with the VARNA gene, are sufficient to produce the same amount of vector produced by an HAdV5 co-infection (Matsushita et al., 1998). Similar findings were obtained for the VA-RNA gene. The HAdV5 VARNA gene encodes two VA-RNA: VA-RNAI and VA-RNAII, where VA-RNAI has been found to play a predominant pro-viral role (Figure 5.1A). The most established and well-understood role of VA RNAI is as an efficient inhibitor of the innate immune protein, double-stranded (ds)RNA-activated kinase (PKR). Through this activity, VA-RNAI blocks the cellular response to dsRNA, a potent pathogen-associated molecular pattern (PAMP), avoiding general protein translation shut-off, thereby allowing HAdV5 and AAV transcripts to be translated (Vachon & Conn, 2016). Therefore, as with the E2A and E4 genes, only a small part of the VARNA gene, the VA-RNAI (~200nt) is required for rAAV vector production. In summary, I found that only two coding sequences, E2A(DBP) and E4(Orf6), and one non-coding sequence, VA-RNAI, are necessary and sufficient to sustain rAAV vector particle formation. However, even though these genes have been identified in the literature, a smaller version of the Helper plasmid carrying only those gene products has, to my knowledge, never been developed before. For this reason, I decided to clone these gene products on one construct that I called Adenoviral μHelper plasmid and tested if it was capable of inducing rAAV vector production.

As shown in Figure 5.1B, I thus cloned the E2A(DBP) and E4(Orf6) on a single transcription unit using two distinct established strategies to design mammalian bicistronic transcripts: the ECMV-IRES and the P2A-skipping ribosome sequences, additionally placing the VA-RNAI immediately after this transcriptional unit. The E2A(DBP), E4(Orf6) and VA-RNAI sequences were cloned from the “p884 Helper AD5” plasmid (a gift of TIGEM Vector Core Facility), a distant derivative of the pXX6 Xiao et al. helper construct. For each bicistronic strategy, I swapped the E2A(DBP) and E4(Orf6) positions, accounting for the reported loss in

translation efficiency of the coding sequence in the second position, characterizing both ECMV-IRES and P2A sequences. This approach yielded four Adenovirus μHelper plasmids (AuH4-8), to which I added the VA-RNAI sequence immediately after the transcriptional unit encoding E2A(DBP) and E4(Orf6), thus generating four additional Adenovirus μHelper plasmids (AuH1-4), as shown in Figure 5.1B.

In all the eight resulting plasmids, the bicistronic transcription unit, encoding adenoviral helper genes is driven by a strong pCMV promoter, and is followed by an SV40 polyadenylation signal. I then evaluated their capacity to elicit rAAV production by comparing the manufacturing yield obtained from each plasmid to that obtained by the parental p884 Helper AD5 plasmid. Results are reported in Figure 5.1C. I observed that all AuH plasmids were able to induce rAAV production, albeit with different yields. AuH plasmids with E2A(DBP) in the first position elicited higher yields than those with E4(Orf6) in the first position (odds versus even numbered AuHs). Moreover, among odds or even AuHs, i.e among plasmids sharing the same relative position between E2A(DBP) and E4(Orf6) genes, it seems that IRES sequences is a better bicistronic strategy compared to the P2A one in the context of rAAV vector production. AuH-3 and AuH-7 are the top performers, producing similar amounts of vector, although their production capacities are lower than that of p884 Helper Ad5 (Figure 5.1C). The AuH-3 and -7 plasmids share the same transcriptional unit, E2A(DBP)-IRES-E4(Orf6), with the only exception being the presence of VA-RNAI in the AuH3 plasmid. These data would suggest that VA-RNAI is not required for rAAV production, as I observed no difference in production yield between plasmids containing VA-RNAI (AuH 1-4) and those that did not (AuH 4-8). However, this observation does not match with findings of X. Xiao et al. and Matsushita et al., neither explains the production yield loss compared to the parental helper plasmid used as manufacturing benchmark (p884 Helper Ad5). Thus, I concluded that AuH plasmids bearing the VA-RNAI sequence might not express it, probably because of the proximity to the SV40 poly-adenylation signal that could exert an inhibitory influence on PolIII-mediated transcription of the VA-RNAI. For this reason, in the future I plan to clone the VA-RNAI sequence before

the pCMV promoter on the reverse strand in a new version of the AuH-3 plasmid, in which Pol-II and Pol-III-mediated transcription should not interfere with each other as they would be going in opposite directions. In conclusion, even though the production yield obtained by the AuH-3 and -7 plasmids should be optimized, I designed a helper bicistronic transcriptional unit, E2A(DBP)-IRES-E4(Orf6), able to supply helper function thus inducing rAAV production and compatible with a CASwitch system implementation.

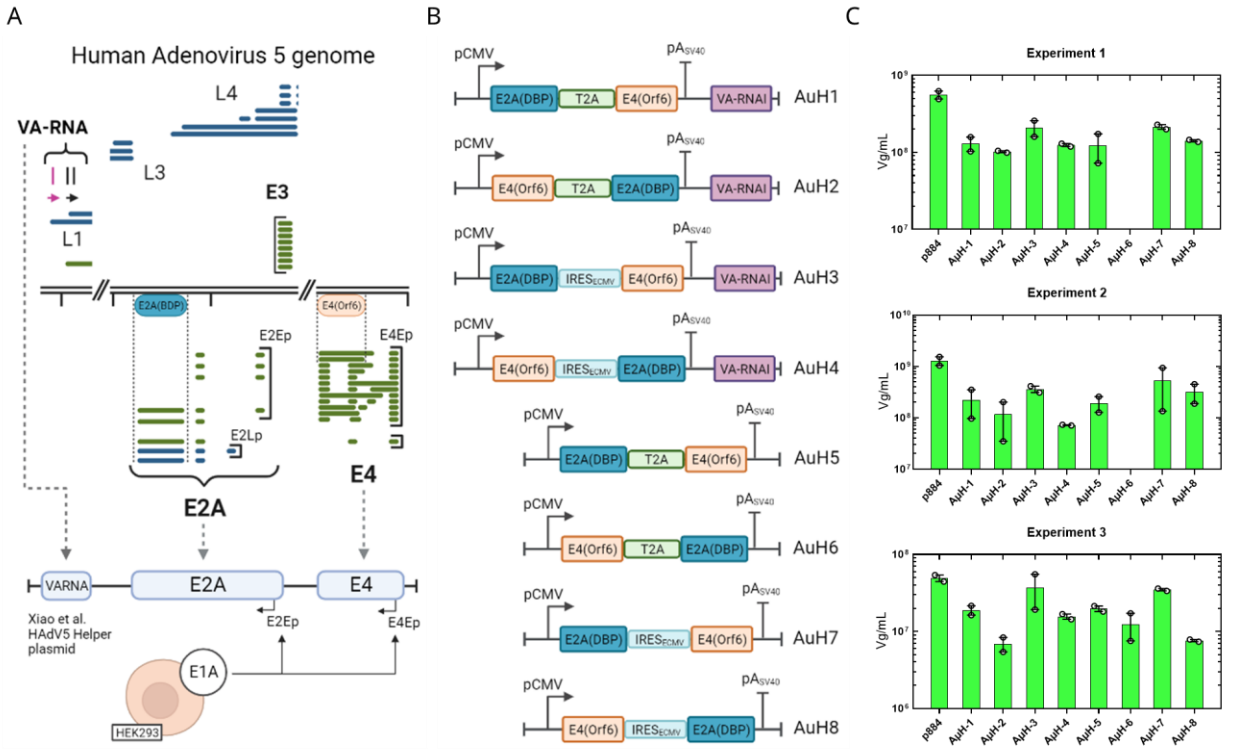


Figure 5.1 Adenoviral µHelper plasmid production screening rationale, schematics, and results. A) Top. Human Adenovirus 5 snippets. E, Early genes (dark green bars); L, Late genes (dark blue bars). Pink arrow indicates VA-RNA-I gene. Bottom. An example representation of the Xiao et al. Helper plasmid. In the HEK293 cell line, E1A transactivates the E2A and E4 genes in the Xiao et al. helper plasmid through the E2-Early promoter (E2Ep) and E4-Early promoter (E4Ep). This makes it impossible for the E2A and E4 genes to be integrated into the genome of HEK293 cells without causing their constitutive toxic expression. B) Schematic representation of adenoviral µH (AµH) plasmids. (C) Quantification the of rAAV yields obtained for each AµH. Three separate experiments were performed to test the rAAV manufacturing yields obtained by transfecting AµH compared to the original parental plasmid p884 Helper AD5. In Experiments 1 and 2, AµH6 was missing because I was not able to clone this construct. After AµH6 successful cloning, I repeated the screening experiment, including the production obtained using AµH6 in the third experiment. For each experiment n=2 technical replicates, bars indicate SE.

5.2 CASwitch-compatible design of AAV-2 Rep/Cap genes and helper bicistronic transcriptional unit.

Unlike Helper genes, the genomic structures of Rep and Cap genes easily lend themselves to the expression control exerted by the CASwitch system. Indeed, even though Rep and Cap genes synthesize several products through three distinct transcription starts (p5, p19, and p40), alternative splicing, and unconventional translation start codon mechanisms, all of the resulting transcripts do share the same 3'-UTR, as shown in Figure 5.2. For this reason, the insertion of a single Direct Repeat in the Rep/Cap genes' shared 3'-UTR enables the repression of their expression by the CasRx in the CASwitch system. Furthermore, this post-translational repression strategy also enables the inhibition of the expression of toxic small Rep genes driven by the AAV p19 internal promoter, which has been a major hurdle in rAAV producer cell line development.

In conclusion, I used the same strategy to design bicistronic transcriptional units encoding for Helper genes and for AAV Rep/Cap by placing the DR in their 3'-UTR, and thus enabling their post-transcriptional repression by means of the CasRx in the absence of doxycycline, as summarized in Figure 5.2.

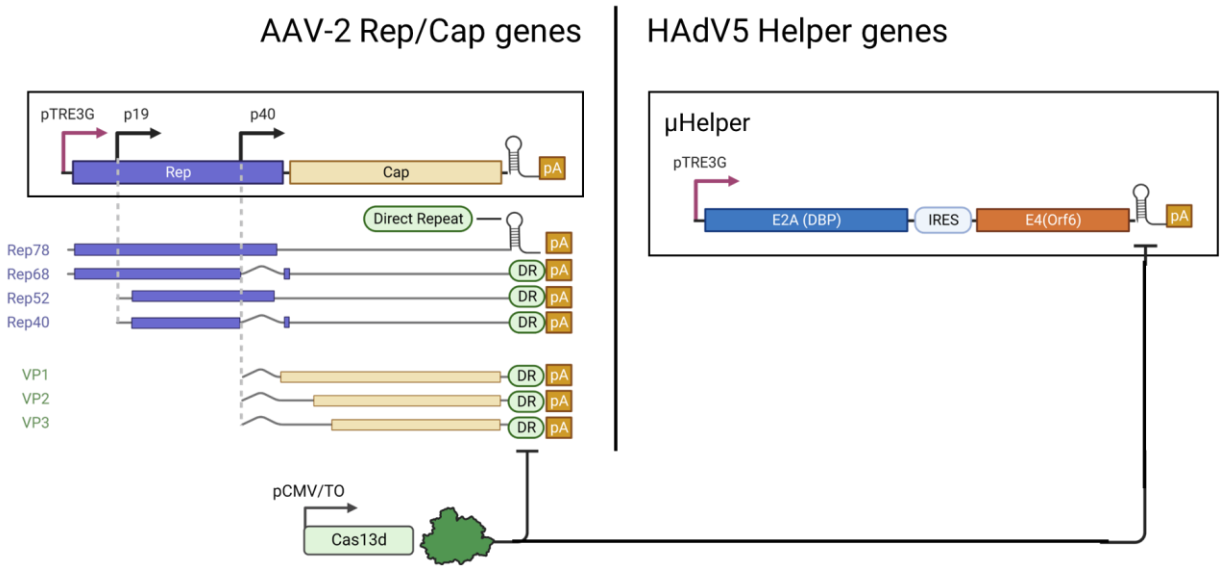


Figure 5.2 Schematics of CASwitch-compatible Rep/Cap and Helper genes. AAV-2 Rep/Cap genes encode Rep78/68 large Rep proteins from the pTRE3G promoter, Rep52/40 small Rep proteins from the p19 internal promoter, and VP1,2,3 capsid proteins from the p40 promoter. All transcripts share the same 3'-UTR, in which a direct repeat is placed, thereby mediating their degradation by recruiting CasRx endoribonuclease. The same consideration applies to the μHelper construct harboring a direct repeat in its 3'-UTR.

Moreover, I employed the pTRE3G promoter in both constructs to drive expression in the presence of doxycycline of both Rep/Cap and Helper genes. Rep and Cap genes sequences were cloned from the p877 Pack 2/2, a gift from the TIGEM's Vector Core Facility, encoding the wild-type AAV-2 Rep and Cap genes.

5.3 On Demand recombinant Adeno-Associated Virus vector (rAAV) production using CASwitch.

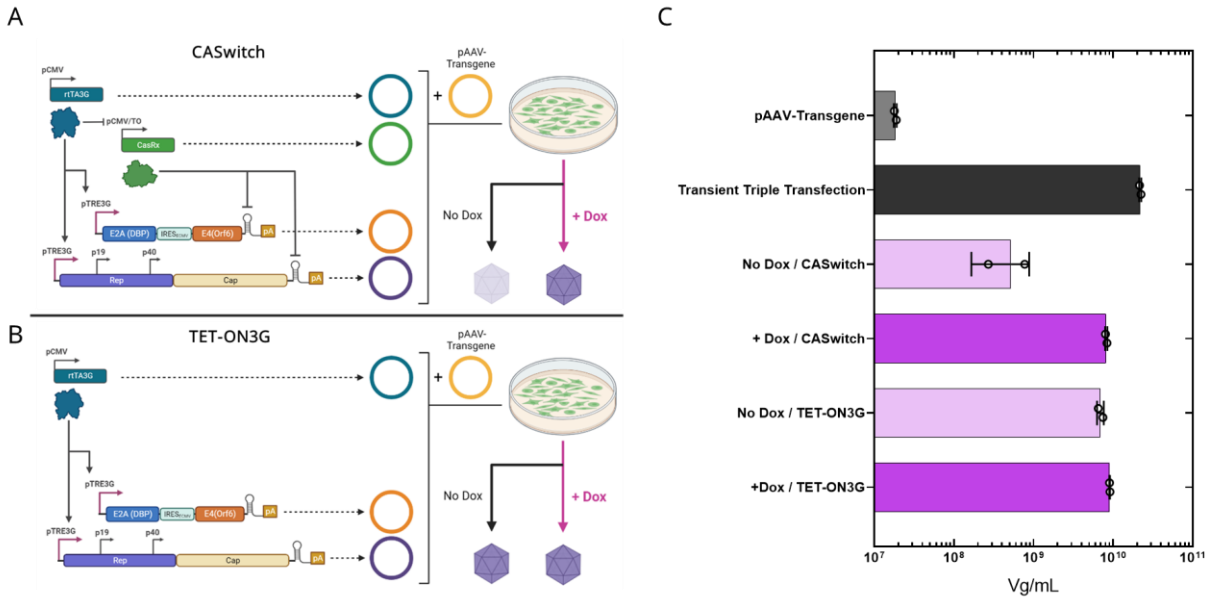


Figure 5.3 CASwith enables inducible rAAV vector production. A) Schematic of the CASwitch implementation for inducible rAAV production. The CASwitch implementation is composed of: i) a pCMV-rtTA3G plasmid (dark blue circle), constitutively expressing rtTA3G and representing the CASwitch main input; ii) a pCMV/TO-CasRx(NLS)-T2A-mCherry plasmid (green circle), encoding a doxycycline transcriptionally repressible CasRx endonuclease and acting as the intermediate output of the CASwitch system; iii) a pTRE3G-E2A(BDP)-IRES-E4(Orf6)-DR plasmid (orange circle), representing the first main output of the CASwitch implementation, who express the Human Adenovirus 5 helper genes: E2A DNA Binding protein (DBP) and E4(Orf6) on the same transcript through the ECMV IRES from a pTRE3G promoter; iv) a pTRE3G-Rep2/Cap2-DR plasmid (purple circle), as second main output, encoding for the AAV-2 Rep and Cap genes. Both constructs iii) and iv) harbor a direct repeat in their respective 3'-UTR, allowing the CASwith to quench their leakiness in the absence of doxycycline. Plasmids i) to iv) together with a pAAV-transgene plasmid (yellow circle) were transfected into HEK293 cells and then treated or not with 1000 ng/mL of doxycycline. Doxycycline should induce viral gene expression, resulting in viral vector production; however, its absence should not. B) The same experimental procedure was performed to test the Tet-On3G system, with the exception that the pCMV/TO-CasRx(NLS)-T2A-mCherry plasmid was not transfected. Given the high leaky expression characteristic of the Tet-On3G system, some vector production should also result in the absence of doxycycline. C) rAAV production titration using absolute real-time PCR. pAAV-Transgene, vector obtained transfecting HEK293 cells with only transgene plasmid; Transient Triple Transfection, vector obtained transfecting HEK293 cells with pAAV-Transgene, p877-Rep2/Cap2, and p884-Helper-AD5; No Dox / -Dox CASwitch, vector produced by doxycycline-uninduced and -induced CASwitch system; No Dox / + Dox Tet-On3G, vector produced by doxycycline-uninduced and -induced Tet-On3G system. n=2 biological replicates, bars indicate SD.

In the CASwitch implementation for inducible rAAV production, the designed expression units encoding the Rep/Cap and Helper genes, respectively represent the main output (Figure 5.3A), where the pTRE3G promoter allows for Rep/Cap and Helper genes expression only after doxycycline administration; while the DR sequence enables transcript degradation by the CasRx endoribonuclease, thus strongly reducing their "leaky" expression in the absence of doxycycline.

Once all the necessary parts had been implemented, I was able to evaluate the capacity of the CASwitch system to induce rAAV vector production only following doxycycline administration. To this end, I transfected HEK293T cells with the constructs described above, i.e. pTRE3G-E2A(DBP)-IRES-E4(Orf6)-DR (orange circle) and pTRE3G-Rep/Cap-DR (purple circle), and the parts composing the CASwitch system, i.e. pCMV-rtTA3G (dark blue circle) and pCMV/TO-CasRx(NLS)-T2A-mCherry (green circle), as shown in Figure 5.3A. To compare the CASwitch performance with that of the state-of the-art Tet-On3G mammalian expression system, I used a control where the pCMV/TO-CasRx-T2A-mCherry plasmid was not transfected, as shown in Figure 5.3B. Both CASwitch and Tet-On3G systems were tested at doxycycline concentrations of 0 ng/mL and 1000 ng/mL, representing the "No Dox" and "+ Dox" conditions, respectively. In addition, both systems were transfected with the pAAV-Transgene plasmid (yellow circle) carrying the GFP coding sequence. As negative control, HEK293 cells were transfected only with the pAAV-Transgene plasmid (pAAV-Transgene), whereas, as production positive control, the traditional transient triple transfection method was used, i.e. by transfecting HEK293 cells with p887 pack 2/2, p884 Helper AD5, pAAV-Trangene plasmids.

The rAAV production yields were quantified by means of absolute real-time PCR and results are reported in Figure 5.3C. I found that the CASwitch system was indeed able to induce vector production to a similar extent as the Tet-On3G system in the presence of doxycycline, suggesting the CASwitch retention of high maximum expression of essential viral genes. For both systems, the manufacturing yields were lower than that of the transient triple transfection production positive control, which may be due to the helper plasmid used in

both systems, which lacks the essential VA-RNAI gene, as observed in the μHelper screening results in Figure 5.1. Furthermore, the CASwitch lowers AAV vector production of an approximately one order of magnitude in the absence of doxycycline, when compared to the Tet-On3G system, thus confirming that the CASwitch implementation efficiently quenches toxic viral leaky expression in the absence of doxycycline. However, when compared to the negative control, it appears that that CASwith in the absence of doxycycline is able to only partially inhibit the production of vectors. This finding, together with the observation that there is no significant production difference between the induced and uninduced TET-ON3 system, suggest that a very small amount of Rep/Cap and Helper gene expression is enough to induce vector production in a transient transfection context. This could lead to the speculation that if viral genes are able to induce high manufacturing yields even at low expression levels, then vector production should not be affected by the maximum expression of viral genes. For this reason, as a future work, I plan to create two CASwitch-embedded stable cell lines with distinct CasRx repression strengths by tuning the amount of integrated endoribonuclease. Next, I will transiently transfect pTRE3G-Rep/Cap-DR and pTRE3G-Helper-DR plasmids into these cell lines to test CASwitch induction and inhibition of rAAV vector production in a stable cell line context.

In conclusion, data in Figure 5.3 provide a proof of principle application for the CASwitch system as a for on-demand AAV vector manufacturing and inspire us to further develop an AAV vector producer cell line.

Chapter 6

Discussion

“The time has come for synthetic biologists to develop more real-world applications [...] the field has had its hype phase, now it needs to deliver.” , this citation, from an article written in 2010 (Kwok, 2010), is perfect to summarize the landmarks achieved by synthetic biology in its first decade and the challenges that it needed to face at the beginning of its second decade, especially the requirement for some real achievements. Now, at the beginning of its third decade, synthetic biology seems to deliver on the hype with some synthetic biology products already on the market such as engineered bacteria and CAR-T cells; this achievements have been possible thanks to concerted efforts to improve on the design of synthetic gene circuits alongside innovation and the discovery of new technologies. However, some problems remain, such as lack of modularity, susceptibility to burden and cell population variability, just to list some of them, thus still hampering real-world applicability of synthetic biology at its full potential.

This thesis focused on finding a solution to one of these problems, specifically the leakiness in inducible gene expression systems that impedes their deployment in useful contexts as human gene therapy and metabolic engineering, in those cased where leaky exogenous gene expression is toxic, so that a tight inducible gene expression system is required. Hence, I decided to improve on the Tet-On3G, representing the state-of-the-art mammalian gene inducible system, with thousands of established implementations in the research context but severely limited ones in real-world implementations, because of its high leakiness. Building upon the Tet-On3G system, I designed a synthetic circuit that

implements a coherent feed-forward loop type-4 motif (CFFL-4), harnessing the multi-level regulation of gene expression which underpins the central dogma of biology, in order to decrease its leakiness without compromising its high maximum achievable expression.

After introducing the research context in Chapter 1 and 2, I characterized the single genetic parts in Chapter 3 and demonstrated that it is possible use the pCMV/TO repressible promoter to implement the CFFL-4 topology. Specifically, in my implementation, in the presence of doxycycline, the rtTA3G drives the expression of a gene of interest (GoI) from the pTRE3G doxycycline-inducible promoter, while the inhibiting the expression of the CasRx CRISPR-endonuclease from the pCMV/TO promoter. On the contrary, in the absence of doxycycline, the CasRx is expressed and it is able to bind a direct repeat (DR) sequence in the 3'UTR of the GoI thus cleaving the polyA and leading to its degradation, thus endowing this system with a quasi-zero leakiness and high maximum induced expression. I called the resulting synthetic circuit, the CASwith, due to its ability to switch the expression of a gene on or off on demand.

In Chapters 4 and 5 I discussed the potential applications of this new technique. In Chapter 4 , I found that CASwitch could effectively act as a dynamic range amplifier circuit , able to increase the fold induction values of two previously established biosensors, one consisting of a synthetic promoter responsinve to copper and the other to synthetic promoter responsive to lysosomal stress. The CASwitch thus enables the development of a new generation of whole cell biosensors with higher specificity and sensitivity capable of detection at concentrations of target molecules that would be difficult to detect using traditional methods.

In Chapter 5, the CASwitch was used to successfully induce the production of AAV vectors, whereas the Tet-On3G showed no difference between uninduced and induced conditions because of the leakiness in viral gene expression even in uninduced conditions. This indicates that the CASwitch technology can be used to tightly regulate the expression of viral genes required for AAV vector production, thus suggesting that an inducible AAV packaging HEK293T cell line can be indeed developed. If this can indeed be done, it would

provide a completely new way to produce AAVs for gene therapy thus decreasing costs while increasing the population of patients that can be treated.

In conclusion, in this work I addressed the issue of leakiness in current state-of-the-art inducible gene expression systems, a critical problem in the context of toxic protein expression or requirement for a low signal-to-noise ratio, by means of a synthetic biology approach. To date, efforts to solve this problem relied on traditional incremental technology approaches, which required several cycles of trial and errors and were case-specific. Here, I instead showed implemented a general-purpose approach that overcomes current limitations of the traditional approaches.

Chapter 7

Material and methods

Plasmid Construction

Plasmid were constructed by Golden Gate cloning using the EMMA cloning platform (Jones et al., 2019; Martella et al., 2017) or NEBridge® Golden Gate Assembly Kit (BsaI-HFv2, BsmBI-v2) employing custom fusion sites. Genetic parts were amplified by PCR with Platinum™ SuperFi II Green PCR Master Mix (ThermoFisher) following manufacturer protocol with a final concentration of 0.1µM each primer and 1ng of total target plasmid DNA. Custom oligonucleotides for PCRs were synthetized by Eurofins. All cloned constructs and primers used to amplify genetic parts from indicated target plasmids are reported in **Table 1, 2,** and **3**. Gel electrophoresis was performed to check amplicons size. Zymoclean Gel DNA Recovery Kit (Zymo Research) was used for DNA agarose gel extraction according to manufacturer instruction.

Cloning reactions for domesticated and assembled plasmids respectively listed in **Table 1** and **2** were performed according to EMMA cloning domestication and assembly protocols (Jones et al., 2019; Martella et al., 2017).

In order to remove internal BsaI restriction site from E2A(DBP)-IRES-E4(Orf6) and allow its domestication a custom golden gate protocol was set-up. Two primers were designed to flank the sequence to amplify (A-Fw, B-Rv), in this case E2A(DBP)-IRES-E4(Orf6), other two primers were instead designed to flank the internal BsaI site (A-Rv; B-Fw), for a total of four primers, all bearing a BsaI site in their respective overhangs. A-Fw and A-Rv primers amplify the sequence upstream, generating “Amplicon-Part A”. B-Fw and B-Rv primers amplify the

sequence downstream, generating “Amplicon-Part B”. The fusions sites preceding the BsaI site in the overhangs of A-Rv and B-Fw primers were designed to make a silent mutation in the E2A(DBP)-IRES-E4(Orf6) sequence removing the internal BsaI site, and to be complementary each other in this way allowing fragment A and B fusion in the part entry vector during domestication cloning reaction and restoring the entire E2A(DBP)-IRES-E4(Orf6) with no internal BsaI site. Custom Golden gate cloning was used to clone pTRE3G-Rep2-Cap2-DR as well, because Cap gene contains an internal BsmbI site. AAV2 Rep Cap genes were amplified from the large Rep ATG until the Cap stop codon from the p887 pack 2/2 plasmid flanking it with a 5’ overhang harboring a BsaI site and an EMMA G fusion site, and a 3’ overhang harboring an EMMA H fusion site and a BsaI site, in a 5’ - 3’ order. At the same time a pTREG-p7-DR destination plasmid was amplified by inverse-PCR from the pTRE3G-gLuc-DR plasmid endowing it with overhangs containing complementary EMMA G and H fusion sites, in order to accommodate the Rep2-Cap2 amplicon, yielding the pTREG-Rep2-Cap2-DR. Custom Golden Gate cloned plasmids are listed in **Table 3**. Custom Golden Gate cloning reactions were assembled with a molar ration of 1:3 between vector and inserts, 1 µL of BsaI-HFv2 or BsmBI-v2 master mix (NEB), 1 µL of T4 ligase Buffer (NEB) and Milli-Q water to a final volume of 10 µL.

Both EMMA and custom Golden Gate clonings were transformed in Stellar™ Competent Cells (Takara) following manufacturer instructions and plated on Kanamycin (domesticated plasmids) or Ampicillin (Assembled and custom Golden Gate plasmids) agar plates. Colonies were screen by an in-house PCR-colony procedure. Briefly, each picked colony was resuspended in 10 µL of MilliQ-water, of which 5 µL were transferred into 45 ul of Luria Broth, the remaining 5ul were lysed at 95°C for 5’, then added to a PCR colony mix containing 10 ul of MyTaq™ HS Red Mix (Meridian Bioscience) 2.5 uM of each primer and MilliQ-water up to 15 uL, for a total reaction volume of 20 uL.

Cell culture and transfection

The HEK293T cell line (ATCC) was cultured in DMEM Glutamax (Gibco) supplemented with 10% Tet-Free Fetal Bovine Serum (Euroclone) and 1% Penicillin-Streptomycin (Euroclone).

For luciferase assay a reverse PEI transfection protocol was used. The same day of transfection 5x10^4 HEK293T cells/well were seeded in µCLEAR® Black 96 well plate (Greiner) and have been transfected with a transfection mix containing 1 uL of a home-made PEI solution (MW 25000, Polysciences, stock concentration 0,324 mg/ml, pH 7.5) and 200ng of DNA per well. Regarding experiments employing the Gaussia Luciferase, also a plasmid constitutively expressing the RedFirefly luciferase was transfected to normalize Gaussia Luciferase values to transfection efficiency. A similar normalization strategy was adopted with experiments employing the Firefly luciferase, where a plasmid constitutively expressing the Renilla luciferase was transfected to normalize Firefly luciferase values to transfection efficiency. Where reported, HEK293T cells were treated with doxycycline before PEI transfection to a final concentration of 1, 2.5, 5, 10, 25, 50, 100, 250, 500 or 1000 ng/mL). For AAV production a reverse PEI transfection protocol was used as well. The same day of transfection 8x10^5 HEK293T cells/well were seeded in Costar® 6-well Clear TC-treated Multiple Well Plate (Corning) and have been transfected with a transfection mix containing 2ug of DNA and 10 uL of a home-made PEI solution (MW 25000, Polysciences, stock concentration 0,324 mg/ml, pH 7.5) each well.

Luciferase Assay

Luciferase assays for HEK293T cell expressing Gaussia and RedFireFly luciferases were performed by means of Pierce™ Gaussia-Firefly Luciferase Dual Assay Kit (ThermoFisher) employing a custom protocol. 48 hours after transfection, media were aspirated and HEK293T cells in the µCLEAR® Black 96 well plate (Greiner) were immediately lysed with 100ul of 1X lysis buffer provided with the Pierce™ Gaussia-Firefly Luciferase Dual Assay Kit. The Working Solution was assembled with 20ul of Gaussia-Firefly Luciferase Dual Assay Buffer, 0.2µL of 100X Coelenterazine and 0.2µL of 100X reconstituted D-Luciferin per well. Luciferase activity was measured immediately after addition of 20ul of Working Solution each well on a Glomax Discovery plate reader (Promega) using an automatic injector protocol. Luciferase arbitrary units (A.U.) were calculated as the ratio between the Gaussia and the Red Firefly luciferase values. Relative Luciferase A.U. and System response values

were obtained as a percentage of the Luciferase A.U. at 0 ng/mL of doxycycline and at 1000 ng/mL, respectively. Regarding experiments employing the Firefly Luciferase, media were aspirated 48 hours after transfection and HEK293T cells in the µCLEAR® Black 96 well plate (Greiner) were immediately lysed with 100ul of 5X Passive Lysis Buffer (Biotin) diluted in MilliQ-water. Luciferase arbitrary units (A.U. Fluc/Rluc) were calculated as the ratio between the Firefly and the Renilla detected luciferase values for each sample.

AAV vector extraction and qRT-PCR titration in crude lysate

72hr after transfection HEK293T cells were collected, pelleted at 1000xg for 5' and resuspended into 500 uL of PBS-MK (1 × PBS, 1 mM MgCl2, and 2.5 mM KCl), then lysed by 4 cycles of freeze and thaw in a dry ice/ethanol bath; finally, the crude lysate was clarified through centrifugation at 18.000xg for 10'. 5uL of clarified crude lysates containing rAAV particles were incubated with 50 units (U) of DNase I (Roche) for 15 hr at 37°C, and then inactivated at 75°C for 30 min; PCR-grade water was added up to 40uL. Samples were then treated with 10 µL of proteinase K (>600 mAU/mL) at 56°C for 2 hr, followed by incubation at 95°C for 30 min to inactivate the enzyme, for a total reaction volume of 50uL, resulting in the 10-fold dilution of the rAAV sample. Linearized pAAV-Transgene plasmid (XhoI-NheI) was serially diluted in PCR-grade water from 2.5× 10^8 to 25 copies/µL in 10-fold serial dilution to generate standard curves for absolute qPCR quantification. All qPCR reactions were composed of 5ul of diluted processed rAAV sample or serially diluted standard, 0.5 µM forward primer, 0.5 µM reverse primer, and 12,5 ul of LightCycler® 480 SYBR Green I Master (Roche) for a total reaction volume of 25 µL each well. Sequences of PCR primers targeting the BGHpA in the pAAV2.1-EGFP were ordered from Eurofins, Forward Primer: 5′-GCCAGCCATCTGTTGT-3′; Reverse Primer: 5′-GGAGTGGCACCTTCCA-3. qPCR reaction were performed using the LightCycler® 96 Instrument (Roche) with cycling conditions as follows: 95°C for 10 min, followed by 40 cycles of 95°C for 15 s and 60°C for 1 min. Absolute quantification analysis were obtained using the LightCycler® 96 Software (Roche), and the viral genome concentration (Vg/mL) was calculated using the formula: $Vg/mL = [(A/B) \times 10 \times 1000]$, where A is the interpolated total viral DNA amount in each well,

B is the starting volume of the pre-processed rAAV samples (5uL), and 10 and 1000 are the dilution factors.

Acknowledgements

I would like to thank my supervisors, Professor Diego di Bernardo for his support, advice, and mostly for trusting me (sometimes more than I trusted myself). I feel that I have learned a lot scientifically and personally during these years, I am deeply grateful to him. I would also like to thank Professor Alberto Auricchio for his advice on the Adeno-Associated Virus project direction and for providing me with the starting materials I needed to produce them. I want to thank Monica Doria for introducing me to AAV manufacturing and techniques and for supporting me in debugging my production protocols. I appreciate Barbara for our "constructive" debates and especially for her insights into AAV infection, which gave me new energy to continue.

I would like to thank all members of the di Bernardo lab, past and present, for being an open, entertaining, and smart group that provided me with great times and discussions. I would like to thank my Ph.D. companion, Iacopo, for his bureaucratic assistance, for listening to my complaints, and especially for all his memes. I deeply thank Alessio for our great scientific discussions, for his support during difficult times, and for his patience as my flatmate. I also want to thank Virginia for having me as her "capo" and for conducting experiments on the CASwitch-amplified biosensors. I admire her commitment to learning and her relentless effort. I also want to thank Francesco for our thought-provoking discussions and for being a great companion to talk about both serious and silly topics.

I would like to thank my parents for always supporting and loving me in every way and for trusting me. Mom and Dad, I love you.

I also want to thank my in-laws for being like second parents to me during the last few years and for listening to all my complaints and offering kind advice.

Finally, I would like to express my deepest gratitude to my girlfriend, Mara. There are no words to describe how grateful I am to have you by my side. You have been my light, guiding me through the toughest times and never fading. You have taught me how to stay strong

and lighthearted. Your constant support is the only reason this thesis is complete. Thank you seems inadequate to express all my gratitude. Your love is the source of all my happiness. I love you!

Appendix A

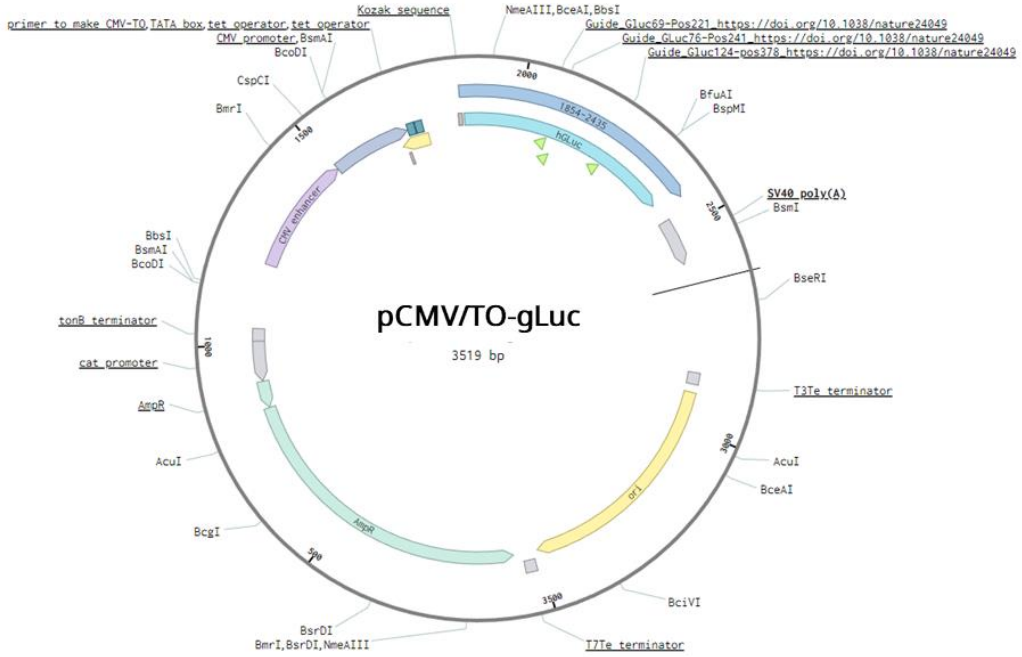


Figure A. 1 pCMV/TO-gLuc vector. pCMV/TO drives the expression of Gaussia Luciferase

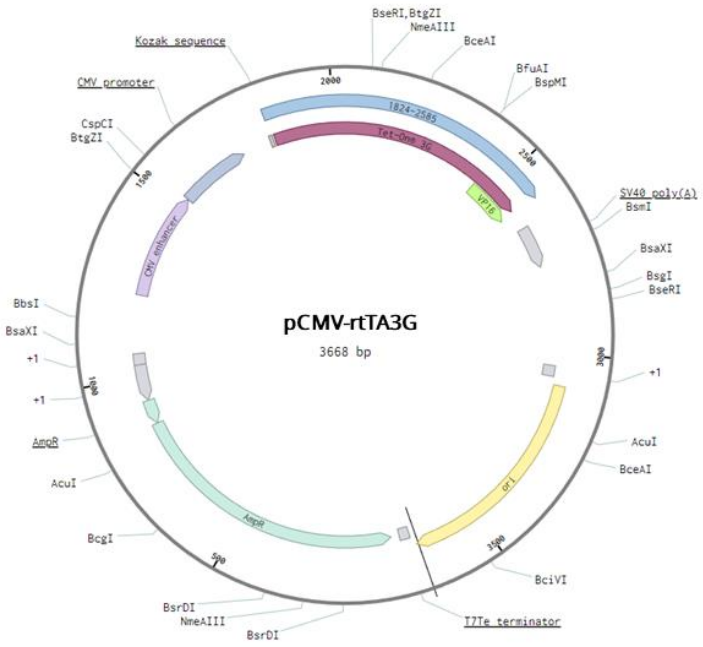


Figure A. 2 pCMV-rtTA3G vector. pCMV promoter drives the expression of the reverse-tetracycline Trans Activator (rtTA3G)

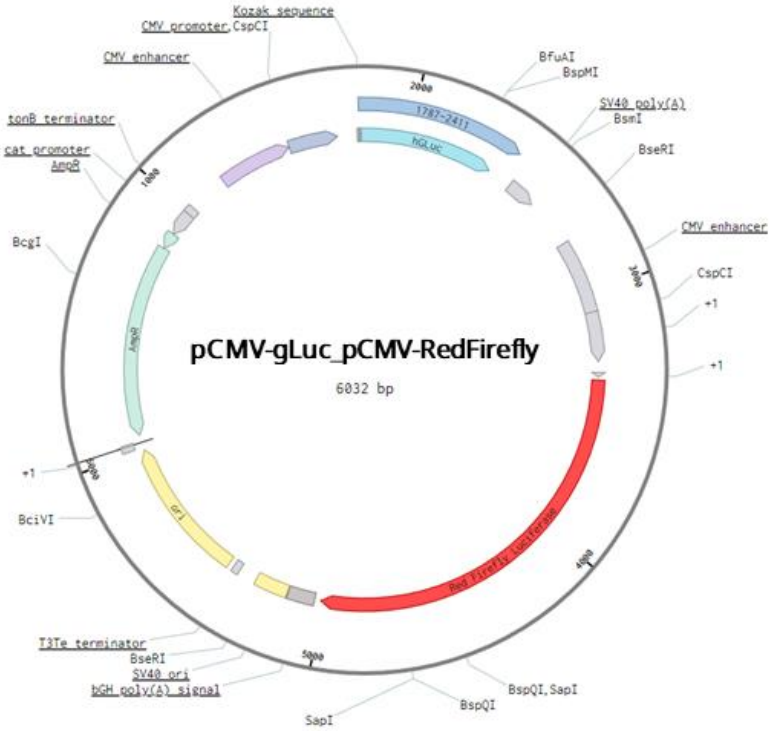


Figure A.3 pCMV-gLuc_pCMV-RedFirefly vector. The first transcriptional unit encoding a Gaussia Luciferase is followed by a second transcriptional unit encoding a RedFirefly Luciferase. Both are driven by the CMV promoter.

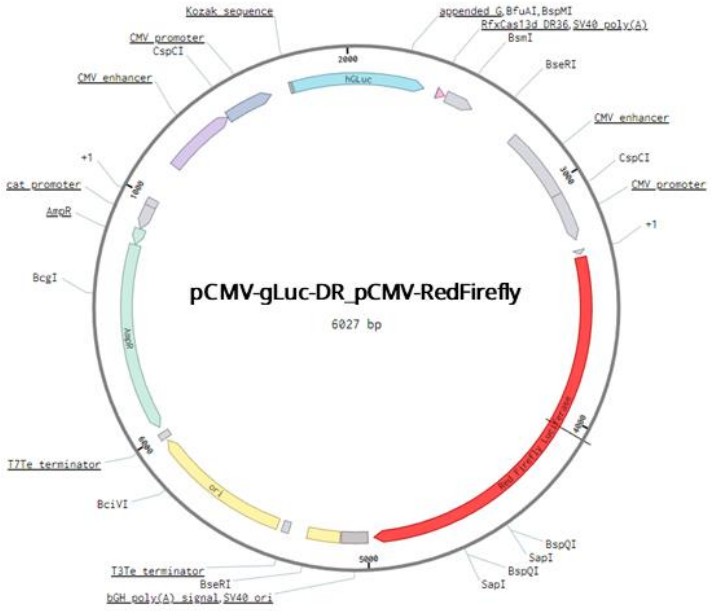


Figure A. 4 pCMV-gLuc_pCMV-RedFirefly vector. The first transcriptional unit encoding a Gaussia Luciferase and containing a Direct Repeat sequence in its 3'UTR is followed by a second transcriptional unit encoding a RedFirefly Luciferase. Both are driven by the CMV promoter.

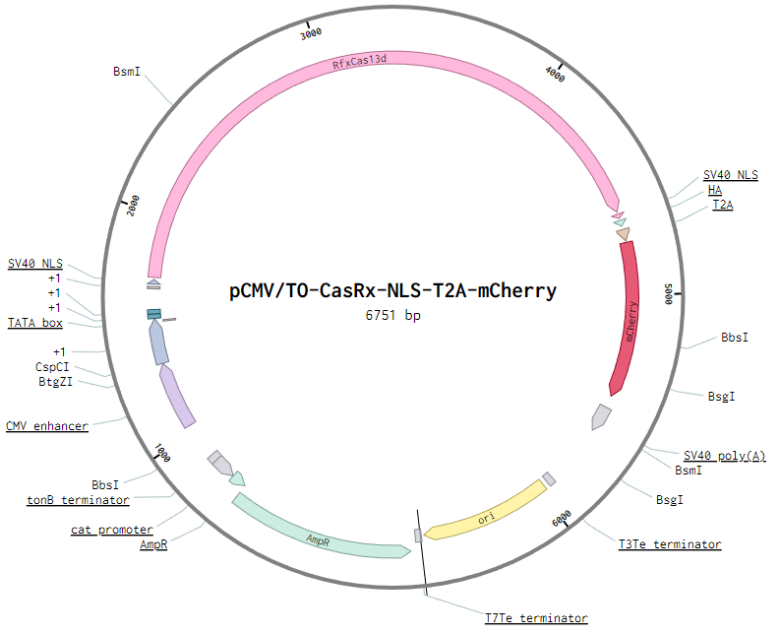


Figure A. 5 pCMV/TO-CasRx(NLS)-T2A-mCherry vector. pCMV/TO promoter drives the expression of the CasRx CRISPR endonuclease and mCherry simultaneously, through the skipping ribosome sequence T2A

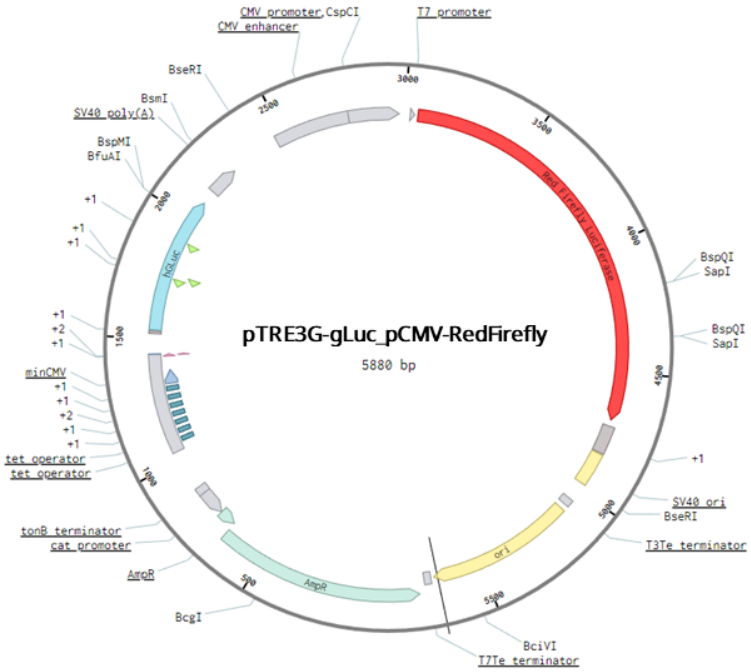


Figure A. 6 pTRE3G-gLuc_pCMV-RedFirefly vector. The first transcriptional unit encoding a Gaussia Luciferase is transcribed from a pTRE3G promoter and is followed by a second transcriptional unit encoding a RedFirefly Luciferase, transcribed from a CMV promoter.

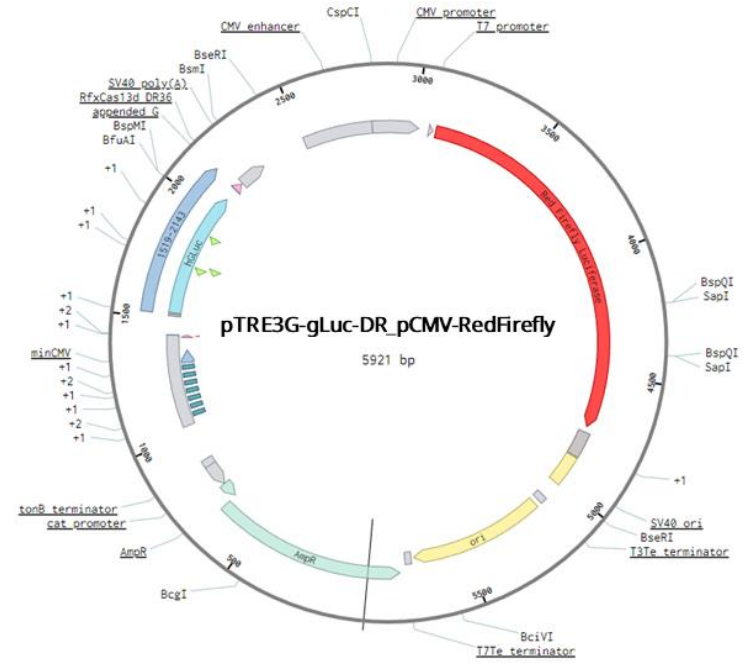


Figure A. 7 pTRE3G-gLuc-DR_pCMV-RedFirefly vector. Figure A. 8 pTRE3G-gLuc_pCMV-RedFirefly vector. The first transcriptional unit encoding a Gaussia Luciferase is transcribed from a pTRE3G promoter and contain a Direct Repeat sequence in its 3'UTR. It is followed by a second transcriptional unit encoding a RedFirefly Luciferase, transcribed from a CMV promoter.

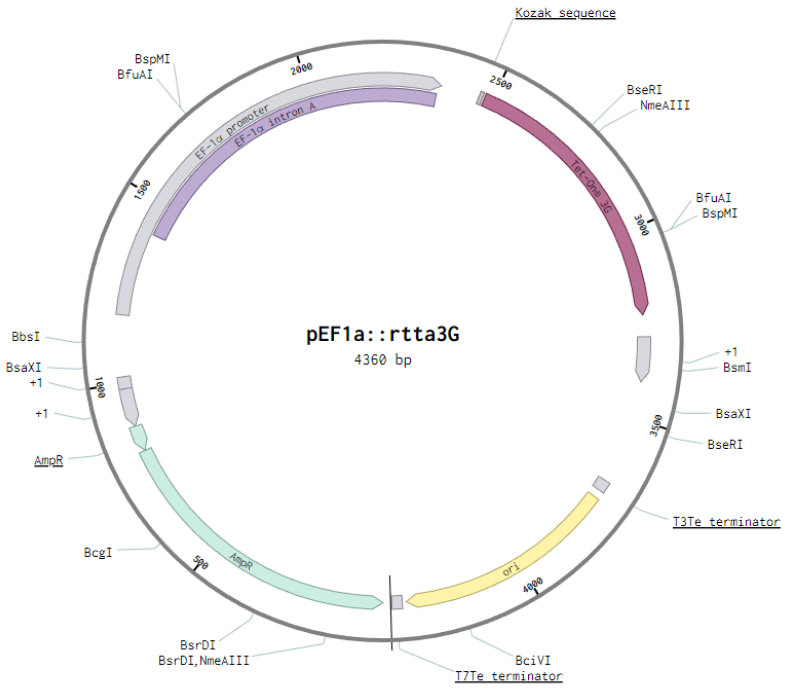


Figure A. 9 pEf1a-rtTA3G vector. The rtTA3G is driven by the human EF1a promoter

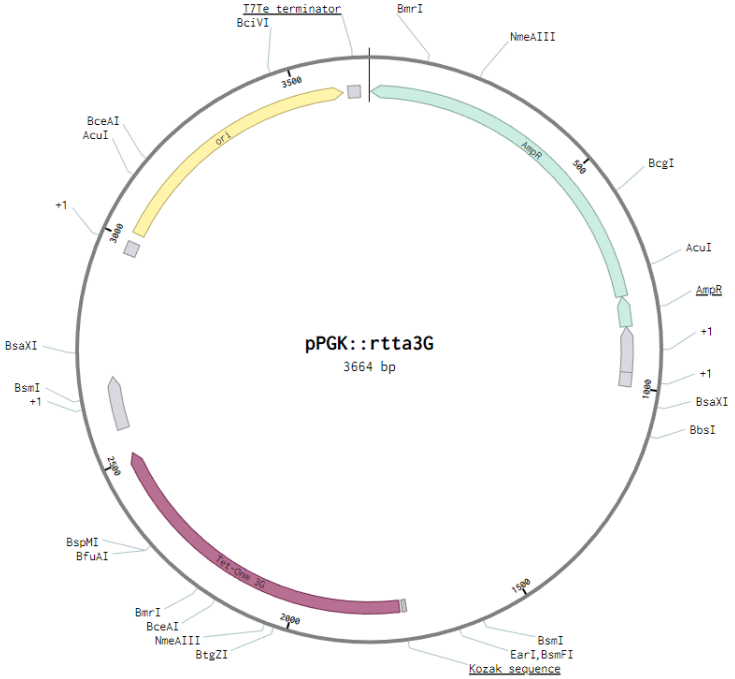


Figure A. 10 pPGK-rtTA3G vector. The rtTA3G is driven by the human pPGK promoter

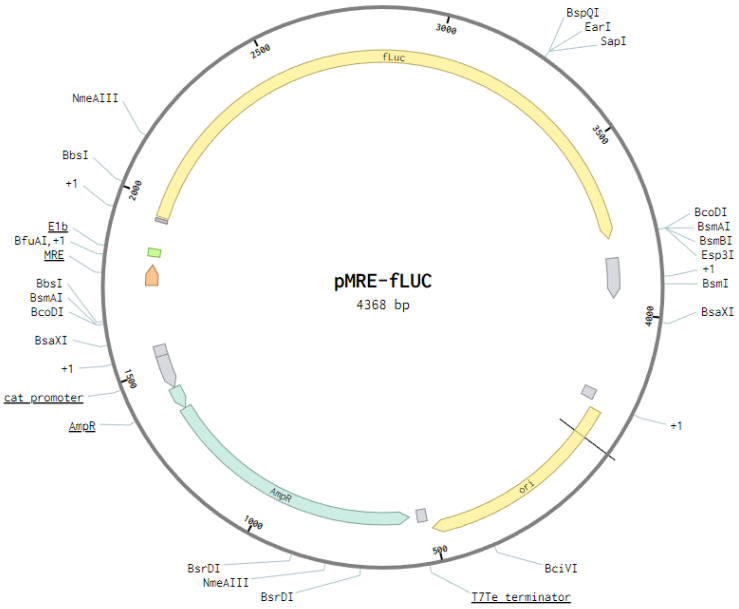


Figure A. 11 pMRE-Fluc vector. Firefly luciferase is expressed from the pMRE synthetic promoter

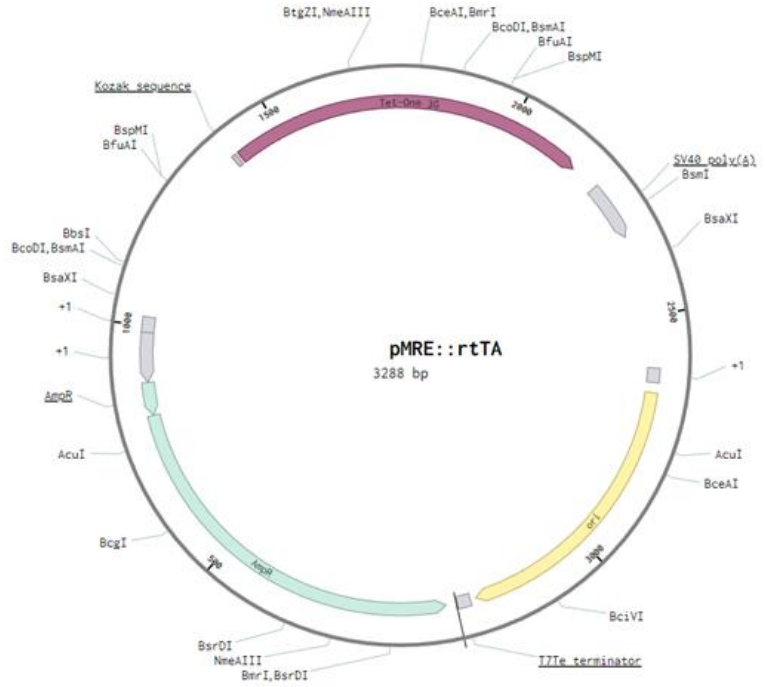


Figure A. 12 pMRE-rtTA3G vector. The rtTA3G is driven by the synthetic pMRE promtoer

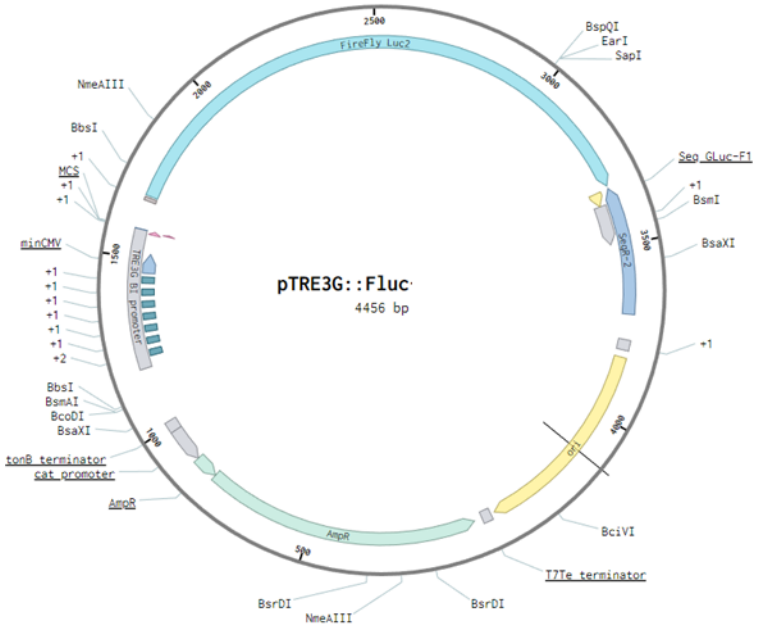


Figure A. 13pTRE3G-Fluc vector. The Firefly luciferase is driven by the pTRE3G promoter

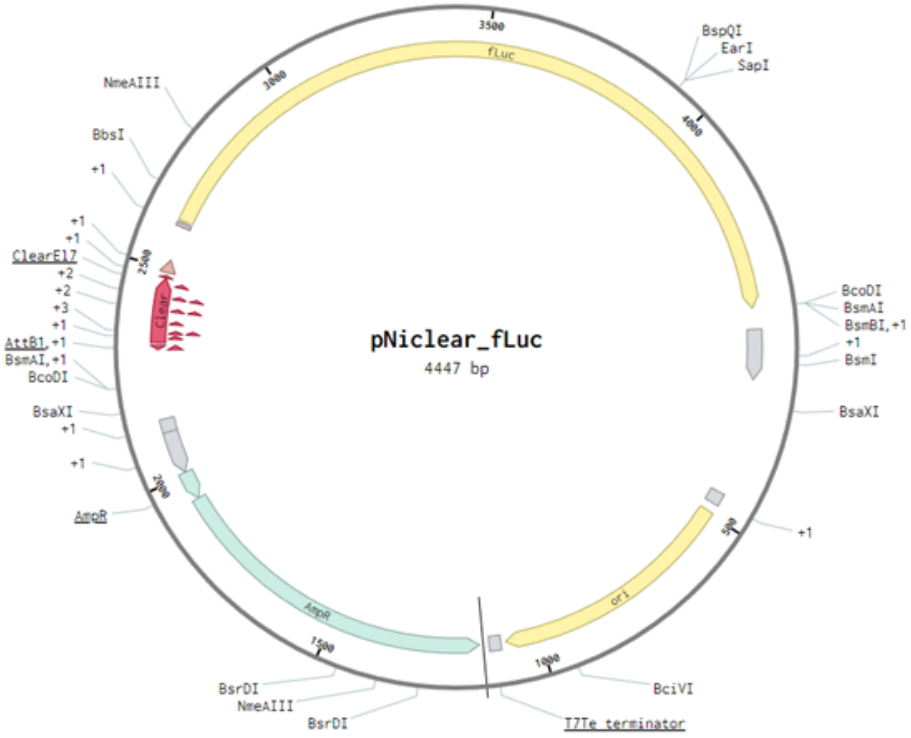


Figure A. 14 pNiClear-Fluc vector. The Firefly luciferase is driven by the synthetic pNiclear promoter.

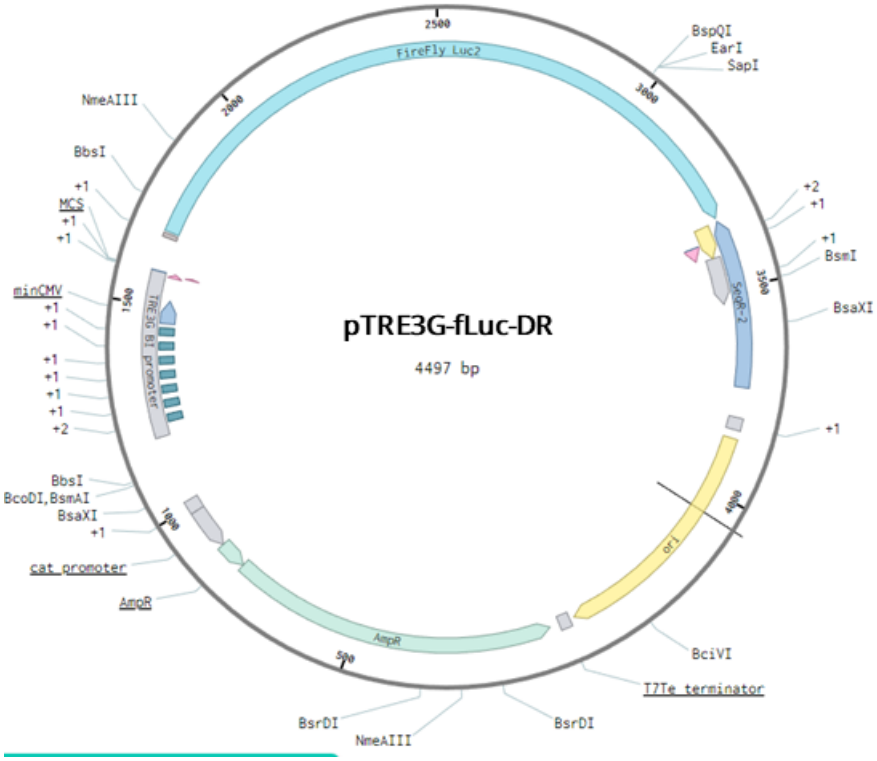


Figure A. 15 pTRE3G-Fluc-DR vector. The Firefly Luciferase is expressed from the pTRE3G promoter and contains a Direct Repeat sequence in its 3'-UTR

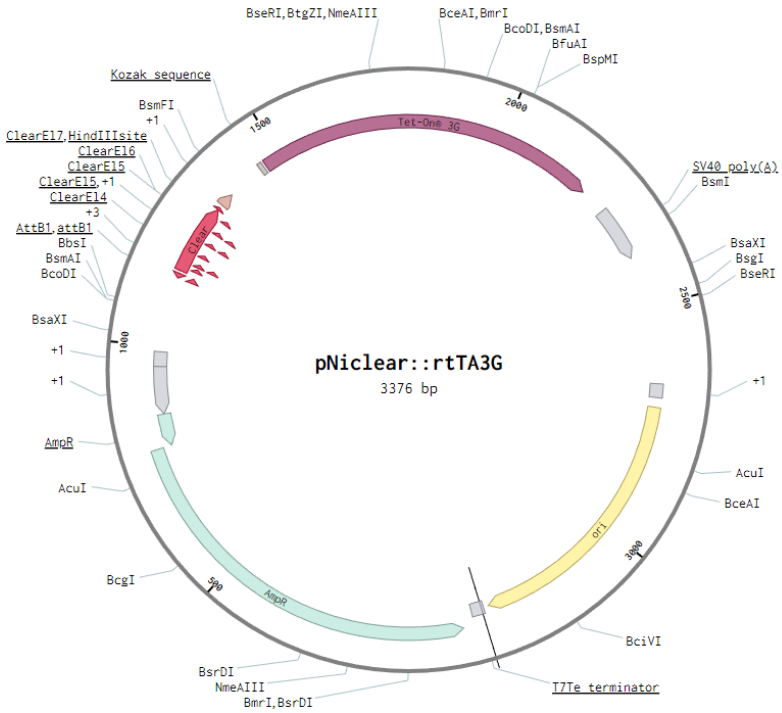


Figure A. 16 pNiClear-rtTA3G vector. The pNiClear drives the expression of the rtTA3G plasmid.

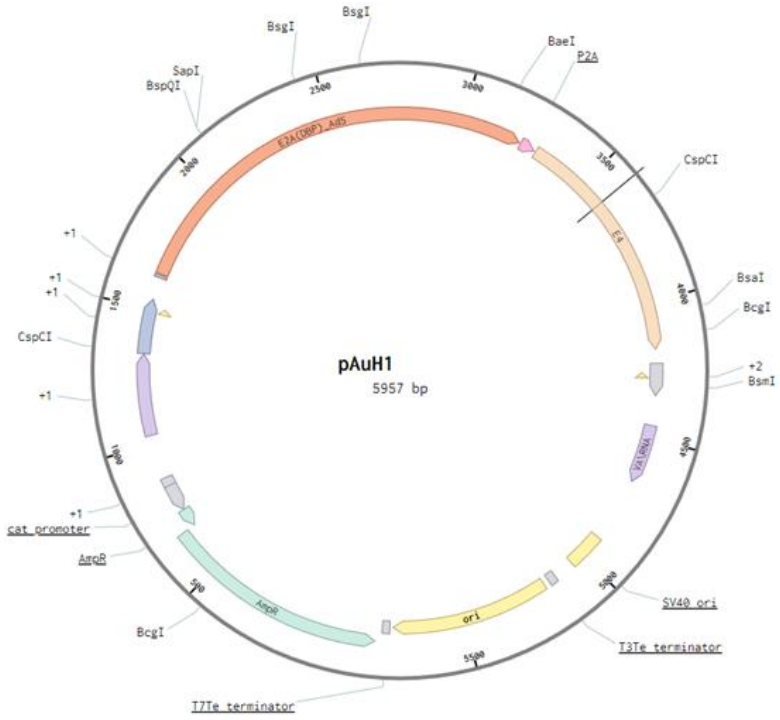


Figure A. 17 Adenoviral μ Helper-1 vector. pCMV-E2A(DBP-P2A-E4(Orf6)-SV40pA-VARNAI

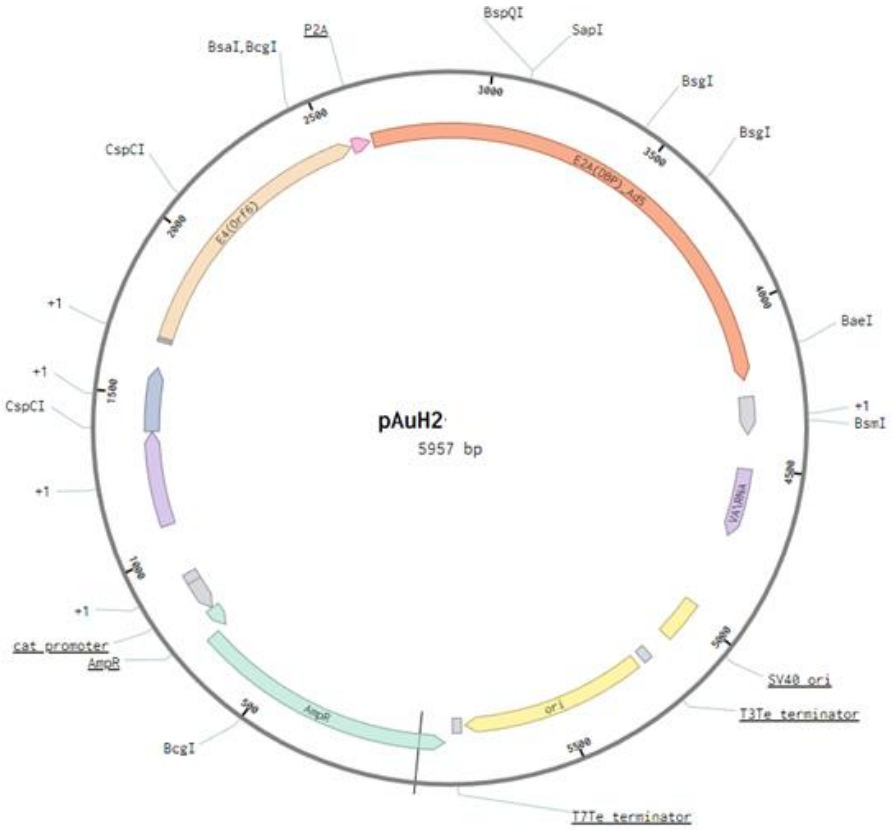


Figure A. 18Adenoviral μ Helper-2 vector. pCMV-E4(Orf6)-P2A-E4(DBP)-SV40pA-VARNAI

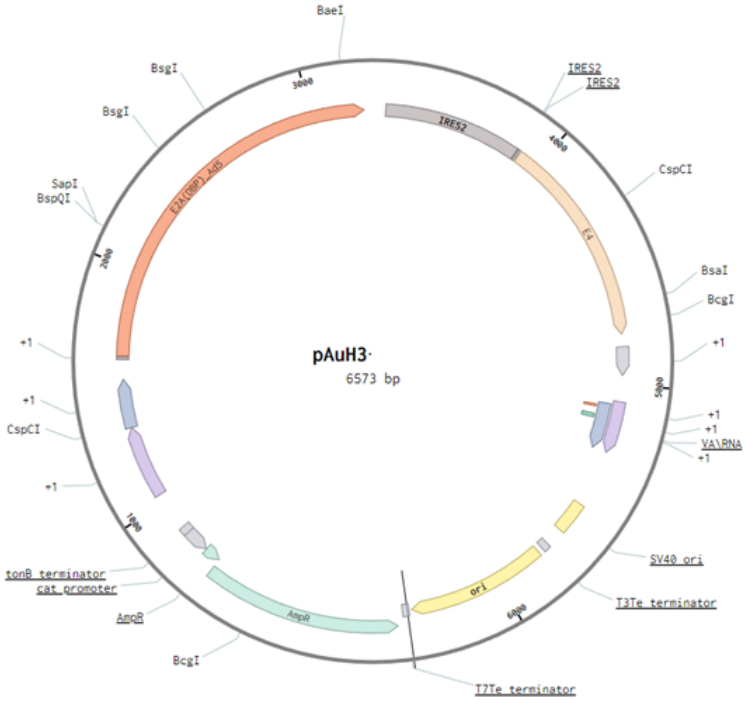


Figure A. 19 Adenoviral μ Helper 3 vector. pCMV-E2A(DBP)-IRES-E4(Orf6)-SV40pA-VARNAI

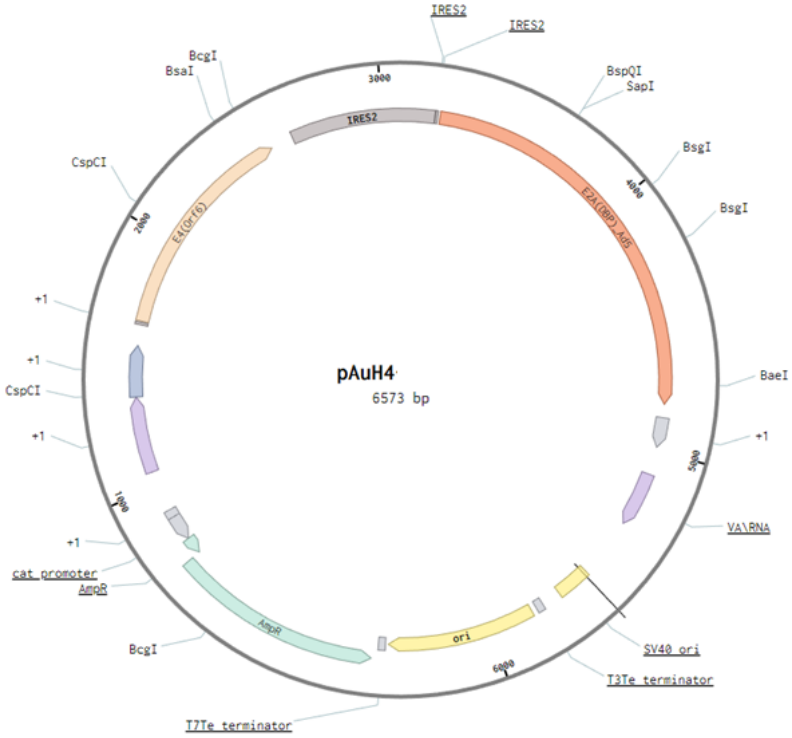


Figure A. 20 Adenoviral μ Helper 4 vector. pCMV-E4(Orf6)-IRES-E4(DBP)-SV40pA-VARNAI

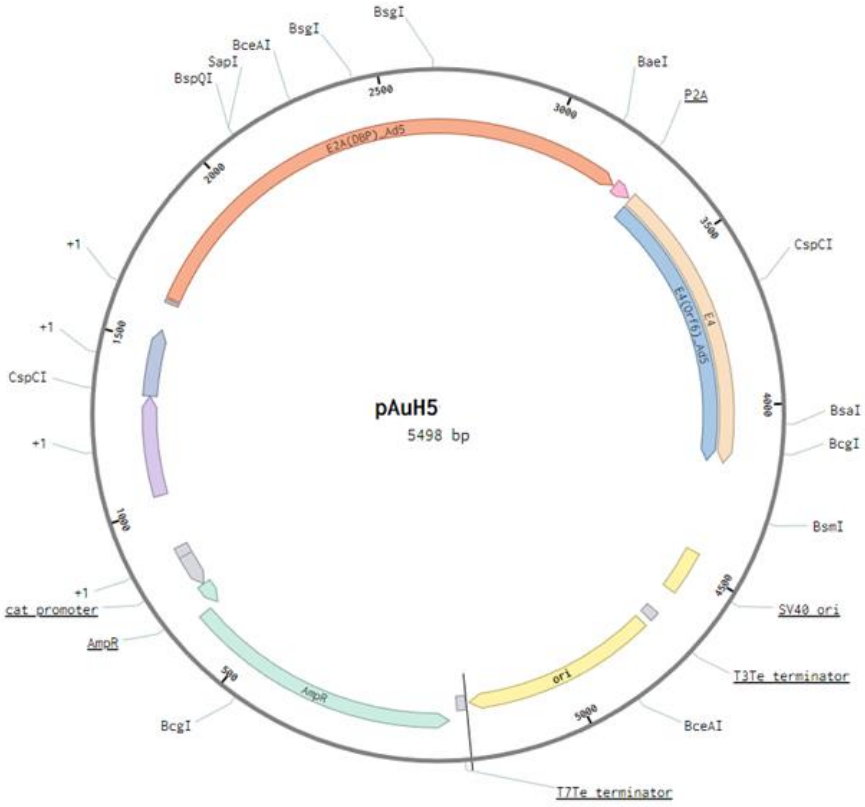


Figure A. 21Adenoviral μ Helper 5 vector. pCMV-E2A(DBP)-P2A-E4(Orf6)-SV40pA

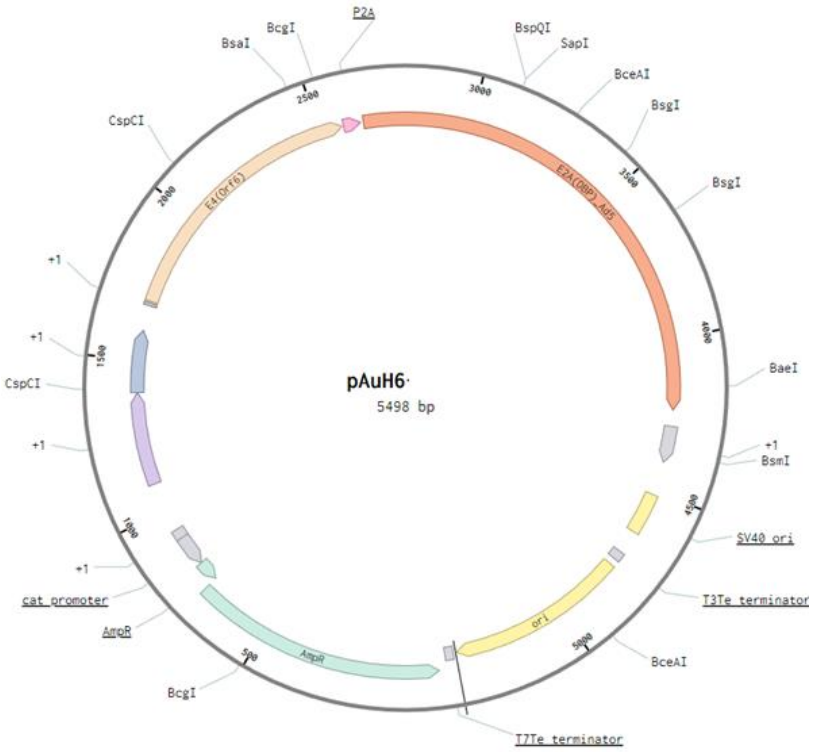


Figure A. 22Adenoviral μ Helper 6 vector. pCMV-E4(Orf6)-P2A-E2(DBP)-SV40pA

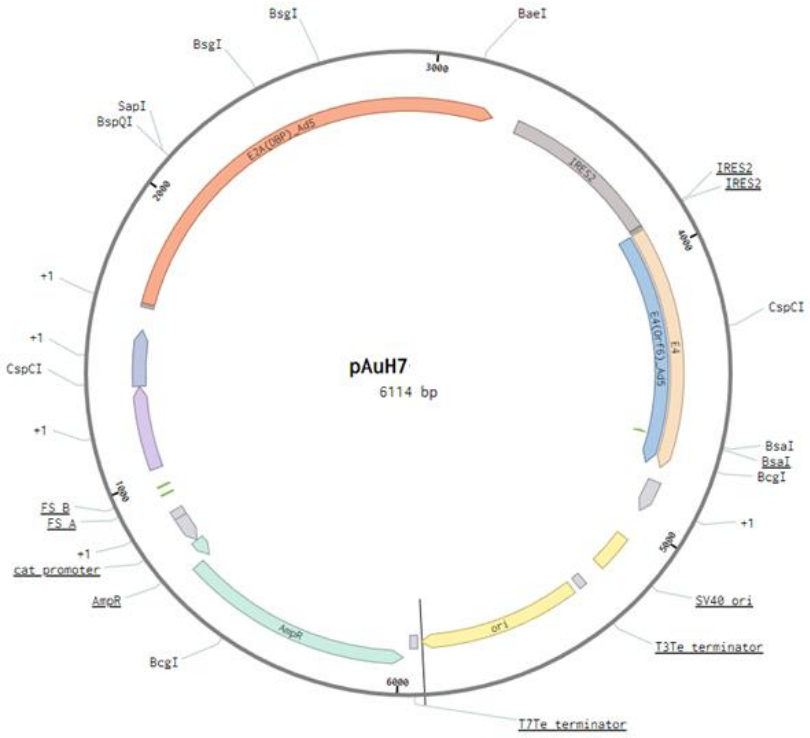


Figure A. 23Adenoviral μ Helper 7 vector. E2A(DBP)-IRES-E4(Orf6)-SV40pA

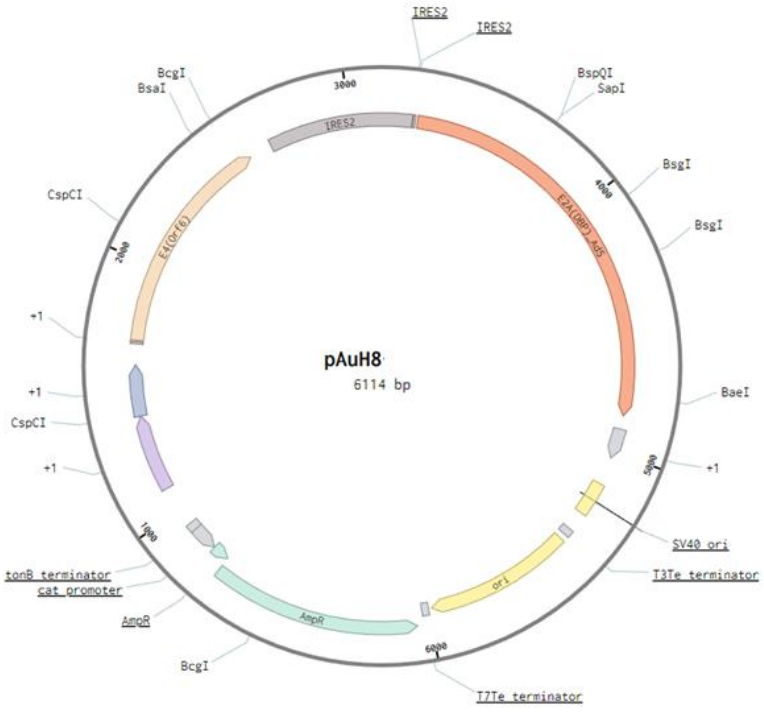


Figure A. 24Adenoviral μ Helper 8 vector. E4(Orf6)-IRES-E2(DBP)-SV40pA

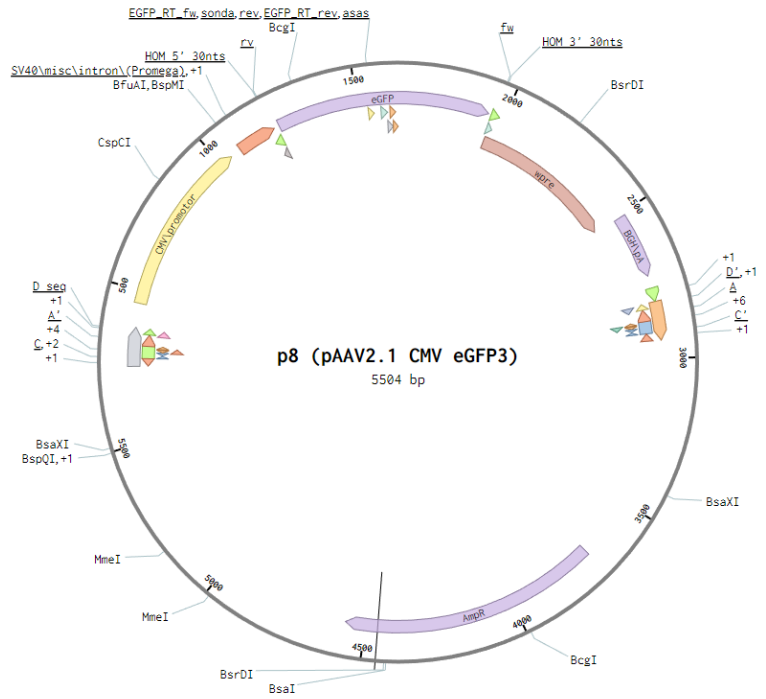


Figure A. 25 p8 pAAV2.1 EGFP vector. A pCMV-EGFP-BGHpA expression unit is flanked by the AAV-2 inverted terminal repeats (ITRs)

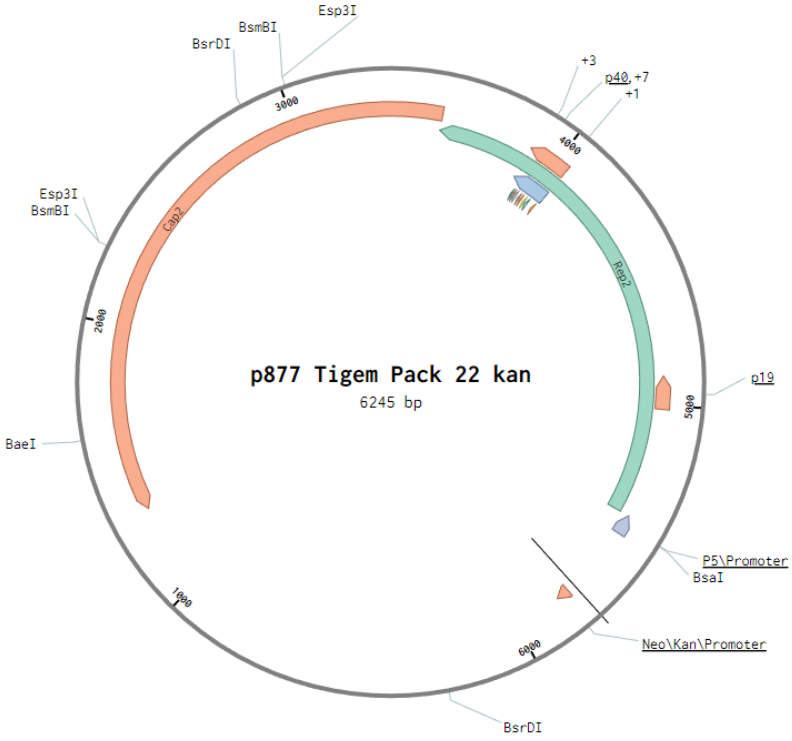


Figure A. 26 p877 pack 2/2 vector. Rep2-Cap2 were cloned into a bacterial plasmid carrying the Kanamycin resistance gene

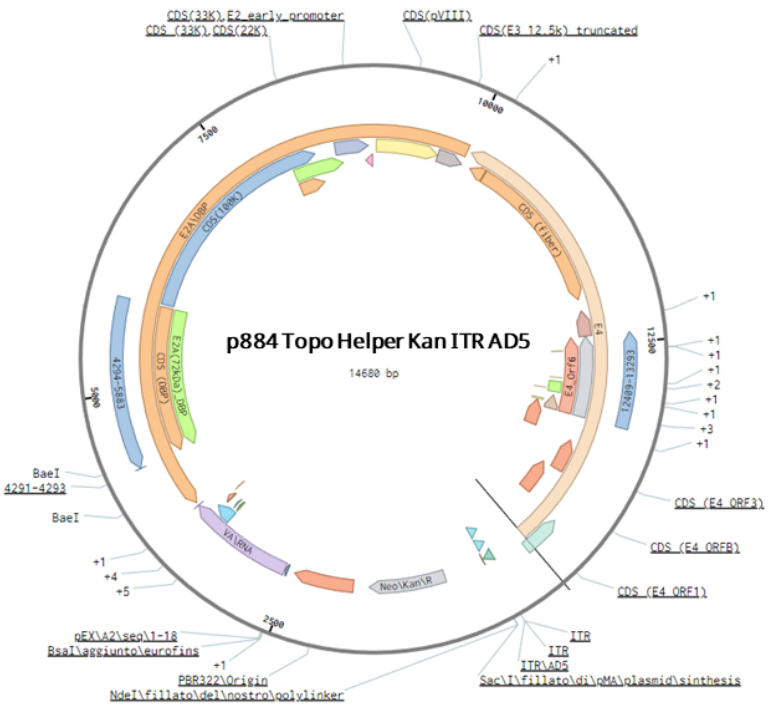


Figure A. 27 p884 Topo Helper AD5 vector. This plasmid encodes for Human Adenovirus 5 E2A, E4 and VARNA genes

Appendix B

Table 1. Domesticated parts with EMMA cloning

Plasmid name	Forward Primer	Reverse Primer	Target
pE3-pCMV/TO	ATAGATGGTCTCAGACTCGTTACATAACTTACGGTAAATGG	ATAGATGGTCTCAGTCCCTCGTCGACGATCTCTATCACTGATAGGGAGATCTCTATCACTGATAGGGAGAGCTCTGCTTATATAGACCTCCC	pE3-pCMV
pE3-pCMV	CACAGGGTCTCAGACTCGTTACATAACTTACGGTAAATG	CACGCGGTCTCAGTCCAGCTCTGCTTATATAGACC	p_i/A8 AAV2.1 CMV EGFP
pE3-pTRE3G	ATAGATGGTCTCAGACTGTACACGCCACCTCGACATACTC	ATAGATGGTCTCAGTCCTTTACGAGGGTAGGAAGTGGTACG	pTRE3G-BI-ZsGreen1
pE3-EF1a	CAATAGGTCTCAGACTGGCTCCGGTGCCCGTCAG	CATTAGGTCTCAGTCCCTGAGTCCGGTAGCGCTAGCG	YCe2741 HC_Kan_EF1ap_p18
pE3-pPGK	pPGK promoter sequence was synthetized by Genewiz **		
pE3-NiClear	ATAGATGGTCTCAGACTAAAAGCAGGCTGCTAGCCC	ATAGATGGTCTCAGTCCCTGGAAGTCGAGCTTCCATTATATACC	p_iii/A28 pDEST Niclear UbGFP
pE7-gLuc	ATAGATGGTCTCACAGCGGAGTCAAAGTTCTGTTTGCCCTGATCTG	ATAGATGGTCTCAGCCTGTCACCACCGGCCCCCTTG	pTK-Gaussia Luc
pE7-rtTA3G	ATAGATGGTCTCACAGCATGTCTAGACTGGACAAGAGCAAAGTC	ATAGATGGTCTCAGCCTGACCCGGGGAGCATGTCAAG	pCMV-Tet3G
pE7-CasRx(NLS)	ATAGATGGTCTCAGCCTGACCCGGGGAGCATGTCAAG	ATAGATGGTCTCAGCCTGAGCCGCTAGCAGCGTAATCTGG	pXR001: EF1a-CasRx-2A-EGFP
pE7-Fluc	ATAGATCGTCTCACAGCGAAGATGCCAAAAAC	ATAGATCGTCTCAGCCTGACACGGCGATCTTGCCGCC	HC_Amp_7BpMin_DDfLuc _Q5_mut
pE7-E2A(DBP)	ACAGTAGGTCTCACAGCGCCAGTCGGGAAGAGGAG	CCATTCGGTCTCAGCCTGAAAAATCAAAGGGGTTCTGCC	p884 Helper AD5
p7-E4(Orf6)*	CTCACACGTCTCAGCCTGACATGGGGGTAGAGTCATAATCG	TAACATCGTCTCACAGCACTACGTCCGGCGTTCC	p884 Helper AD5
pE8-T2A-mCherry	ATAGATGGTCTCAAGGCGAGGGCAGAGGAAGTCTGCTAAC	ATAGATGGTCTCAACGCTGGGCCAGGATTCTCCTCG	pXR001: EF1a-CasRx-2A-EGFP
pE9-mCherry	CACTGGGTCTCAGCGTCCGTGAGCAAGGGCGAGGAGGA	GATATGGTCTCAAGCATTACTTGTACAGCTCGTCCATGCCG	p_iii/A57 pAI274_pT2A-MCS-pPGK-H2B-mcherry-pA

pE9-E2A(DBP)	ATAGATGGTCTCAGCGTCCATGGCCAGTCGGGAAGAGGAG	ATAGATGGTCTCAAGCATTAAAAATCAAAGGGGTTCTGCC	p884 Helper AD5
p9-E4(Orf6)*	ATAGATCGTCTCAAGCATTACATGGGGGTAGAGTCATAATCG	ATAGATCGTCTCAGCGTCCATGACTACGTCCGGCGTTC	p884 Helper AD5
pE10-DR(Rfx)	ATAGATGGTCTCATGCTGCAAGTAAACCCCTACCAACTGGTCGG	ATAGATGGTCTCATACCGTTTCAAACCCCGACCAGTTGGTAGGGG	primer dimer
pE13-VARNAI	CAAAGAGGTCTCAGTAGAAAAAAGCTAGCGCAGCAGCC	CAAAGAGGTCTCATCACAAAGGAGAGCCTGTAAGCGGG	p884 Helper AD5
pE24-pCMV-RedFireFly	ATAGATGGTCTCACACGCGTTACATAACTTACGGTAAATGG	ATAGATGGTCTCACAGTTGATGCAATTTCCTCATTTTATTAGGAAAG	pCMV-Red Firefly Luc

* p7-E4(Orf6) and p9-E4(Orf6) used for Adenoviral μHelper plasmid assembly were not domesticated because of internal BsaI site, therefore an Esp3I flanked PCR amplicon was used instead of the domesticated part in the EMMA assembly reactions listed in Table 2.

** pPGK ordered sequence:
CTTATGGTCTCAGACTTACCGGGTAGGGGAGGCGCTTTTCCCAAGGCAGTCTGGAGCATGCGCTTTAGCAGCCCCGCTGGGCACTTGGCGCTACACAAGTGGCCTCTGGCCTCGCACACATTCCACATCCACCGGTAGGCGCCAACCGGCTCCGTTCTTTGGTGGCCCCTTCGCGCCACCTTCTACTCCTCCCCTAGTCAGGAAGTTCCCCCCCGCCCCGCAGCTCGCGTCGTGCAGGACGTGACAAATGGAAGTAGCACGTATCACTAGTCTCGTGCAGATGGACAGCACCGCTGAGCAATGGAAGCGGGTAGGCCTTTGGGGCAGCGGCCAATAGCAGCTTTGCTCCTTCGCTTTCTGGGCTCAGAGGCTGGGAAGGGGTGGGTCCGGGGGCGGGCTCAGGGGCGGGCTCAGGGGCGGGGCGGGCGCCCGAAGGTCCTCCGGAGGCCCGGCATTCTGCACGCTTCAAAAGCGCACGTCTGCCGCGCTGTTCTCCTCTTCCTCATCTCCGGGCCTTTCGGGACTGAGACCTAATG

Table 2. Assembled plasmid with EMMA Cloning

Plasmid name	Figure	EMMA assembly general scheme*
pCMV/TO-gLuc	3.2	[AC]-p3-[DF]-[p6-Kozak-ATG]-p7-[HK]-[p11-SV40pA]-[LY]-[YZ]
pCMV-rtTA3G	3.2, 3.4, 3.5, 3.7, 5.3	[AC]-p3-[DF]-[p6-Kozak-ATG]-p7-[HK]-[p11-SV40pA]-[LY]-[YZ]
pCMV-gLuc_pCMV-RedFirefly	3.3, 3.4	[AC]-p3-[DF]-[p6-Kozak-ATG]-p7-[HK]-[p11-SV40pA]-[LY]-p24
pCMV-gLuc-DR_pCMV-RedFirefly	3.3, 3.4	[AC]-p3-[DF]-[p6-Kozak-ATG]-p7-[HJ]-p10-[p11-SV40pA]-[LY]-[YZ]
pCMV/TO-CasRx(NLS)-T2A-mCherry	3.3, 3.4, 3.5, 3.7, 4.2, 4.4, 5.3	[AC]-p3-[DF]-[p6-Kozak-ATG]-p7-p8-p9-[JK]-[p11-SV40pA]-[LY]-YZ]

pTRE3G-gLuc_pCMV-RedFirefly	3.5, 3.7	[AC]-p3-[DF]-[p6-Kozak-ATG]-p7-[HK]-[p11-SV40pA]-[LY]-p24
pTRE3G-gLuc-DR_pCMV-RedFirefly	3.5, 3.7	[AC]-p3-[DF]-[p6-Kozak-ATG]-p7-[HJ]-p10-[p11-SV40pA]-[LY]-[YZ]
pEF1a-rtTA3G	3.7	[AC]-p3-[DF]-[p6-Kozak-ATG]-p7-[HK]-[p11-SV40pA]-[LY]-[YZ]
pPGK-rtTA3G	3.7	[AC]-p3-[DF]-[p6-Kozak-ATG]-p7-[HK]-[p11-SV40pA]-[LY]-[YZ]
pTRE3G-Fluc	4.2, 4.4	[AC]-p3-[DF]-[p6-Kozak-ATG]-p7-[HK]-[p11-SV40pA]-[LY]-[YZ]
pTRE3G-Fluc-DR	4.2, 4.4	[AC]-p3-[DF]-[p6-Kozak-ATG]-p7-[HJ]-p10-[p11-SV40pA]-[LY]-[YZ]
pNiClear-Fluc	4.4	[AC]-p3-[DF]-[p6-Kozak-ATG]-p7-[HK]-[p11-SV40pA]-[LY]-[YZ]
pNiClear-rtTA3G	4.4	[AC]-p3-[DF]-[p6-Kozak-ATG]-p7-[HK]-[p11-SV40pA]-[LY]-[YZ]
pCMV-E2A(DBP)-P2A-E4(Orf6)-VARNAI	AuH-1, 5.1	[AC]-p3-[DF]-[p6-Kozak-ATG]-p7-p8-p9-[JK]-[p11-SV40pA]-[LM]-p13-[NR]-[RW]-[WZ]
pCMV-E4(Orf6)-P2A-E2A(DBP)-VARNAI	AuH-2, 5.1	[AC]-p3-[DF]-[p6-Kozak-ATG]-p7-p8-p9-[JK]-[p11-SV40pA]-[LM]-p13-[NR]-[RW]-[WZ]
pCMV-E2A(DBP)-IRES-E4(Orf6)-VARNAI	AuH-3, 5.1	[AC]-p3-[DF]-[p6-Kozak-ATG]-p7-p8-p9-[JK]-[p11-SV40pA]-[LM]-p13-[NR]-[RW]-[WZ]
pCMV-E4(Orf6)-IRES-E2A(DBP)-VARNAI	AuH-4, 5.1	[AC]-p3-[DF]-[p6-Kozak-ATG]-p7-p8-p9-[JK]-[p11-SV40pA]-[LM]-p13-[NR]-[RW]-[WZ]
pCMV-E2A(DBP)-P2A-E4(Orf6)	AuH-5, 5.1	[AC]-p3-[DF]-[p6-Kozak-ATG]-p7-p8-p9-[JK]-[p11-SV40pA]-[LY]-YZ]
pCMV-E4(Orf6)-P2A-E2A(DBP)	AuH-6, 5.1	[AC]-p3-[DF]-[p6-Kozak-ATG]-p7-p8-p9-[JK]-[p11-SV40pA]-[LY]-YZ]
pCMV-E2A(DBP)-IRES-E4(Orf6)	AuH-7, 5.1	[AC]-p3-[DF]-[p6-Kozak-ATG]-p7-p8-p9-[JK]-[p11-SV40pA]-[LY]-YZ]
pCMV-E4(Orf6)-IRES-E2A(DBP)	AuH-8, 5.1	[AC]-p3-[DF]-[p6-Kozak-ATG]-p7-p8-p9-[JK]-[p11-SV40pA]-[LY]-YZ]

*EMMA Kit (Addgene) domesticated parts and connectors in square brackets. Custom domesticated parts are indicated by the plasmid name and their position in the assembly scheme is not in square brackets. All custom domesticated parts are listed in Table 1.

Table 3. Plasmid cloned with Custom Golden Gate

Plasmid	Inserts - Vector	**Forward primer**	**Reverse primer**	**Target**
pE7-E2A(DBP)-IRES-E4(Orf6)	E2A(DBP)-IRES-E4(Orf6)-partA	ATAGATGGTCTCACAGCATGGGCAGCGCCAGTCGG	ATAGATGGTCTCAGTTTCCTCAGCGATGATTCGCACCG	pCMV-E2A(DBP)-IRES-E4(Orf6)

	E2A(DBP)-IRES-E4(Orf6)-partB	ATAGATGGTCTCAAAACCACTGCCATGTTGTATTCC	ATAGATGGTCTCAGCCTGACATGGGGGTAGAGTCATAATCG	pCMV-E2A(DBP)-IRES-E4(Orf6)
	YCe1695 HC_Kan_RFP-p7 (EMMA Kit, Addgene)			
pTRE3G-Rep2-Cap2-DR	p7-Rep2/Cap2	ATAGATGGTCTCACAGCATGCCGGGGTTTTACGAGATTG	ATAGATGGTCTCAGCCTGACAGATTACGAGTCAGGTATCTGG	p887 pack 2/2
	pTRE3G-p7-DR	ATAGATGGTCTCAAGGCTAACCGGACTCTAGAACATCCCTACAGG	ATAGATGGTCTCTGCTGCCCATGGTGGCGG	pTRE3G-gLuc-DR
pTRE3G-E2A(DBP)-IRES-E4(Orf6)-DR	pE7-E2A(DBP)-IRES-E4(Orf6)			
	pTRE3G-p7-DR	ATAGATGGTCTCAAGGCTAACCGGACTCTAGAACATCCCTACAGG	ATAGATGGTCTCTGCTGCCCATGGTGGCGG	pTRE3G-gLuc-DR
pMRE-rtTA3G	pE7-rtTA3G			
	pMRE-p7	ATAGATCGTCTCAAGGCTAATTCTAGAGTCGGGGCGG	ATAGATCGTCTCAGCTGCCCATGGTACCAACAGTACCGG	pGL3-E1b-TATA-4MRE
pMRE-Fluc	pE7-Fluc			
	pMRE-p7	ATAGATCGTCTCAAGGCTAATTCTAGAGTCGGGGCGG	ATAGATCGTCTCAGCTGCCCATGGTACCAACAGTACCGG	pGL3-E1b-TATA-4MRE

References

Alon, U. (2007). Network motifs: Theory and experimental approaches. *Nature Reviews Genetics*, *8*(6), 450–461. https://doi.org/10.1038/nrg2102

Auslände, S., Ausländer, D., Müller, M., Wieland, M., & Fussenegger, M. (2012). Programmable single-cell mammalian biocomputers. *Nature*, *487*(7405), 123–127. https://doi.org/10.1038/nature11149

Ausländer, S., & Fussenegger, M. (2016). Engineering gene circuits for mammalian cell-based applications. *Cold Spring Harbor Perspectives in Biology*, *8*(7), 1–17. https://doi.org/10.1101/cshperspect.a023895

Balakrishnan, B., & Jayandharan, G. (2014). Basic Biology of Adeno-Associated Virus (AAV) Vectors Used in Gene Therapy. *Current Gene Therapy*, *14*(2), 86–100. https://doi.org/10.2174/1566523214666140302193709

Beitz, A. M., Oakes, C. G., & Galloway, K. E. (2022). Synthetic gene circuits as tools for drug discovery. *Trends in Biotechnology*, *40*(2), 210–225. https://doi.org/10.1016/j.tibtech.2021.06.007

Berthet, C., Raj, K., Saudan, P., & Beard, P. (2005). How adeno-associated virus Rep78 protein arrests cells completely in S phase. *Proceedings of the National Academy of Sciences of the United States of America*, *102*(38), 13634–13639. https://doi.org/10.1073/pnas.0504583102

Bonnet, J., Yin, P., Ortiz, M. E., Subsoontorn, P., & Endy, D. (2013). Amplifying genetic logic gates. *Science*, *340*(6132), 599–603. https://doi.org/10.1126/science.1232758

Brophy, J. A. N., & Voigt, C. A. (2014). Principles of genetic circuit design. *Nature Methods*, *11*(5), 508–520. https://doi.org/10.1038/nmeth.2926

Cameron, D. E., Bashor, C. J., & Collins, J. J. (2014). A brief history of synthetic biology. *Nature Reviews Microbiology*, *12*(5), 381–390. https://doi.org/10.1038/nrmicro3239

Cooper, M., Oldstone, M., Potter, M., Vogt, P. K., & Wilson, I. (1995). *The Molecular Repertoire of Adenoviruses III: Vol. 199/3* (W. Doerfler & P. Böhm (eds.)). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-642-79586-2

Cubillos-Ruiz, A., Guo, T., Sokolovska, A., Miller, P. F., Collins, J. J., Lu, T. K., & Lora, J. M. (2021). Engineering living therapeutics with synthetic biology. *Nature Reviews Drug Discovery*, *20*(12), 941–960. https://doi.org/10.1038/s41573-021-00285-3

Deans, T. L., Cantor, C. R., & Collins, J. J. (2007). A Tunable Genetic Switch Based on RNAi and Repressor Proteins for Regulating Gene Expression in Mammalian Cells. *Cell*, *130*(2), 363–372. https://doi.org/10.1016/j.cell.2007.05.045

Del Vecchio, D., Dy, A. J., & Qian, Y. (2016). Control theory meets synthetic biology. *Journal of the Royal Society Interface*, *13*(120). https://doi.org/10.1098/rsif.2016.0380

DiAndreth, B., Wauford, N., Hu, E., Palacios, S., & Weiss, R. (2022). PERSIST platform provides programmable RNA regulation using CRISPR endoRNases. *Nature Communications*, *13*(1), 1–11. https://doi.org/10.1038/s41467-022-30172-3

Donovan-Banfield, I., Turnell, A. S., Hiscox, J. A., Leppard, K. N., & Matthews, D. A. (2020). Deep splicing plasticity of the human adenovirus type 5 transcriptome drives virus evolution Supplementary. *Svenska Lakartidningen*, *54*(22), 1859–1861. http://www.ncbi.nlm.nih.gov/pubmed/13455520

Gardner, T. S., Cantor, C. R., & Collins, J. J. (2000). Construction of a genetic toggle switch in Escherichia coli. *Nature*, *403*(6767), 339–342. https://doi.org/10.1038/35002131

Geoffroy, M.-C., & Salvetti, A. (2005). Helper Functions Required for Wild Type and Recombinant Adeno- Associated Virus Growth. *Current Gene Therapy*, *5*(3), 265–271. https://doi.org/10.2174/1566523054064977

Gossen, M., & Bujard, H. (1992). Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proceedings of the National Academy of Sciences of the United States of America*, *89*(12), 5547–5551. https://doi.org/10.1073/pnas.89.12.5547

Gossen, M., Freundlieb, S., Bender, G., Müller, G., Hillen, W., & Bujard, H. (1995). Transcriptional Activation by Tetracyclines in Mammalian Cells. *Science*, *268*(5218), 1766–1769. https://doi.org/10.1126/science.7792603

Greco, F. V., Grierson, C. S., & Gorochowski, T. E. (2020). *Harnessing the central dogma for stringent multi-level control of gene expression*.

Grieger, J. C., Soltys, S. M., & Samulski, R. J. (2016). Production of recombinant adeno-associated virus vectors using suspension HEK293 cells and continuous harvest of vector from the culture media for GMP FIX and FLT1 clinical vector. *Molecular Therapy*, *24*(2), 287–297. https://doi.org/10.1038/mt.2015.187

Grimm, D., Kern, A., Rittner, K., & Kleinschmidt, J. A. (1998). Novel tools for production and purification of recombinant adenoassociated virus vectors. *Human Gene Therapy*, *9*(18), 2745–2760. https://doi.org/10.1089/hum.1998.9.18-2745

Hermonat, P. L., & Muzyczka, N. (1984). Use of adeno-associated virus as a mammalian DNA cloning vector: Transduction of neomycin resistance into mammalian tissue culture cells. *Proceedings of the National Academy of Sciences of the United States of America*, *81*(20 I), 6466–6470. https://doi.org/10.1073/pnas.81.20.6466

Jones, S., Martella, A., & Cai, Y. (2019). EMMA assembly explained: A step-by-step guide to assemble synthetic mammalian vectors. In *Methods in Enzymology* (1st ed., Vol. 617). Elsevier Inc. https://doi.org/10.1016/bs.mie.2018.12.017

Kondratov, O., Marsic, D., Crosson, S. M., Mendez-Gomez, H. R., Moskalenko, O., Mietzsch, M., Heilbronn, R., Allison, J. R., Green, K. B., Agbandje-McKenna, M., & Zolotukhin, S. (2017). Direct Head-to-Head Evaluation of Recombinant Adeno-associated Viral Vectors Manufactured in Human versus Insect Cells. *Molecular Therapy*, *25*(12), 2661–2675. https://doi.org/10.1016/j.ymthe.2017.08.003

Konermann, S., Lotfy, P., Brideau, N. J., Oki, J., Shokhirev, M. N., & Hsu, P. D. (2018). Transcriptome Engineering with RNA-Targeting Type VI-D CRISPR Effectors. *Cell*, *173*(3), 665-676.e14. https://doi.org/10.1016/j.cell.2018.02.033

Kwok, R. (2010). Five hard truths for synthetic biology. *Nature*, *463*(7279), 288–290. https://doi.org/10.1038/463288a

Li, J., Samulski, R. J., & Xiao, X. (1997). Role for highly regulated rep gene expression in adeno-associated virus vector production. *Journal of Virology*, *71*(7), 5236–5243. http://www.ncbi.nlm.nih.gov/pubmed/9188591%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC191759

Lim, J. W., Ha, D., Lee, J., Lee, S. K., & Kim, T. (2015). Review of micro/nanotechnologies for microbial biosensors. *Frontiers in Bioengineering and Biotechnology*, *3*(MAY), 1–13. https://doi.org/10.3389/fbioe.2015.00061

Litovco, P., Barger, N., Li, X., & Daniel, R. (2021). Topologies of synthetic gene circuit for optimal fold change activation. *Nucleic Acids Research*, *49*(9), 5393–5406. https://doi.org/10.1093/nar/gkab253

Loew, R., Heinz, N., Hampf, M., Bujard, H., & Gossen, M. (2010). Improved Tet-responsive promoters with minimized background expression. *BMC Biotechnology*, *10*, 1–13. https://doi.org/10.1186/1472-6750-10-81

Mangan, S., & Alon, U. (2003). Structure and function of the feed-forward loop network motif. *Proceedings of the National Academy of Sciences of the United States of America*, *100*(21), 11980–11985. https://doi.org/10.1073/pnas.2133841100

Martella, A., Matjusaitis, M., Auxillos, J., Pollard, S. M., & Cai, Y. (2017). EMMA: An Extensible Mammalian Modular Assembly Toolkit for the Rapid Design and Production of Diverse Expression Vectors. *ACS Synthetic Biology*, *6*(7), 1380–1392. https://doi.org/10.1021/acssynbio.7b00016

Matsushita, T., Elliger, S., Elliger, C., Podsakoff, G., Villarreal, L., Kurtzman, G. J., Iwaki, Y., & Colosi, P. (1998). Adeno-associated virus vectors can be efficiently produced without helper virus. *Gene Therapy*, *5*(7), 938–945. https://doi.org/10.1038/sj.gt.3300680

Meier, A. F., Fraefel, C., & Seyffert, M. (2020). The interplay between adeno-associated virus and its helper viruses. In *Viruses* (Vol. 12, Issue 6, pp. 8–12). https://doi.org/10.3390/v12060662

Merten, O.-W. (2016). AAV vector production: state of the art developments and remaining

challenges. *Cell and Gene Therapy Insights*, *2*(5), 521–551. https://doi.org/10.18609/cgti.2016.067

Michael B. Elowitz, & Stanislas Leibler. (2000). A synthetic oscillatory networkof transcriptional regulators. *Nature*, *403*, 335–338.

Nitzan, M., Rehani, R., & Margalit, H. (2017). Integration of Bacterial Small RNAs in Regulatory Networks. *Annual Review of Biophysics*, *46*, 131–148. https://doi.org/10.1146/annurev-biophys-070816-034058

Ogasawara, Y., Urabe, M., & Ozawa, K. (1998). The use of heterologous promoters for adeno-associated virus (AAV) protein expression in aav vector production. *Microbiology and Immunology*, *42*(3), 177–185. https://doi.org/10.1111/j.1348-0421.1998.tb02269.x

Origel Marmolejo, C. A., Bachhav, B., Patibandla, S. D., Yang, A. L., & Segatori, L. (2020). A gene signal amplifier platform for monitoring the unfolded protein response. *Nature Chemical Biology*, *16*(5), 520–528. https://doi.org/10.1038/s41589-020-0497-x

Penaud-Budloo, M., François, A., Clément, N., & Ayuso, E. (2018). Pharmacology of Recombinant Adeno-associated Virus Production. *Molecular Therapy - Methods and Clinical Development*, *8*(March), 166–180. https://doi.org/10.1016/j.omtm.2018.01.002

Pieters, P. A., Nathalia, B. L., Van Der Linden, A. J., Yin, P., Kim, J., Huck, W. T. S., & De Greef, T. F. A. (2021). Cell-Free Characterization of Coherent Feed-Forward Loop-Based Synthetic Genetic Circuits. *ACS Synthetic Biology*, *10*(6), 1406–1416. https://doi.org/10.1021/acssynbio.1c00024

Puertollano, R., Ferguson, S. M., Brugarolas, J., & Ballabio, A. (2018). The complex relationship between TFEB transcription factor phosphorylation and subcellular localization . *The EMBO Journal*, *37*(11), 1–12. https://doi.org/10.15252/embj.201798804

Purcell, O., & Lu, T. K. (2014). Synthetic analog and digital circuits for cellular computation and memory. *Current Opinion in Biotechnology*, *29*(1), 146–155. https://doi.org/10.1016/j.copbio.2014.04.009

Purnick, P. E. M., & Weiss, R. (2009). The second wave of synthetic biology: From modules to systems. *Nature Reviews Molecular Cell Biology*, *10*(6), 410–422. https://doi.org/10.1038/nrm2698

Qian, Y., McBride, C., & Del Vecchio, D. (2018). Programming Cells to Work for Us. *Annual Review of Control, Robotics, and Autonomous Systems*, *1*, 411–440. https://doi.org/10.1146/annurev-control-060117-105052

Qiao, C., Li, J., Skold, A., Zhang, X., & Xiao, X. (2002). Feasibility of Generating Adeno-Associated Virus Packaging Cell Lines Containing Inducible Adenovirus Helper Genes. *Journal of Virology*, *76*(4), 1904–1913. https://doi.org/10.1128/jvi.76.4.1904-1913.2002

Qiao, C., Wang, B., Zhu, X., Li, J., & Xiao, X. (2002). A Novel Gene Expression Control System and Its Use in Stable, High-Titer 293 Cell-Based Adeno-Associated Virus Packaging Cell Lines. *Journal of Virology*, *76*(24), 13015–13027. https://doi.org/10.1128/jvi.76.24.13015-13027.2002

Robert, M. A., Chahal, P. S., Audy, A., Kamen, A., Gilbert, R., & Gaillet, B. (2017). Manufacturing of recombinant adeno-associated viruses using mammalian expression platforms. *Biotechnology Journal*, *12*(3), 1–16. https://doi.org/10.1002/biot.201600193

Rubens, J. R., Selvaggio, G., & Lu, T. K. (2016). Synthetic mixed-signal computation in living cells. *Nature Communications*, *7*. https://doi.org/10.1038/ncomms11658

Sardiello, M., Palmieri, M., Alberto di Ronza, Diego Luis Medina, & Valenza, M. (2009). A Gene Network Regulating Lysosomal. *Science*, *July*, 473–478.

Schmidt, M., Afione, S., & Kotin, R. M. (2000). Adeno-Associated Virus Type 2 Rep78 Induces Apoptosis through Caspase Activation Independently of p53. *Journal of Virology*, *74*(20), 9441–9450. https://doi.org/10.1128/jvi.74.20.9441-9450.2000

Siddiqui, M., Tous, C., & Wong, W. W. (2022). Small molecule-inducible gene regulatory systems in mammalian cells: progress and design principles. *Current Opinion in Biotechnology*, *78*, 102823. https://doi.org/10.1016/j.copbio.2022.102823

Srivastava, A., Mallela, K. M. G., Deorkar, N., & Brophy, G. (2021). Manufacturing Challenges and Rational Formulation Development for AAV Viral Vectors. *Journal of Pharmaceutical Sciences*, *110*(7), 2609–2624. https://doi.org/10.1016/j.xphs.2021.03.024

T. Das, A., Tenenbaum, L., & Berkhout, B. (2016). Tet-On Systems For Doxycycline-inducible Gene Expression. *Current Gene Therapy*, *16*(3), 156–167. https://doi.org/10.2174/1566523216666160524144041

Täuber, B., & Dobner, T. (2001). Molecular regulation and biological function of adenovirus early genes: The E4 ORFs. *Gene*, *278*(1–2), 1–23. https://doi.org/10.1016/S0378-1119(01)00722-3

Teo, J. J. Y., Woo, S. S., & Sarpeshkar, R. (2015). Synthetic Biology: A Unifying View and Review Using Analog Circuits. *IEEE Transactions on Biomedical Circuits and Systems*, *9*(4), 453–474. https://doi.org/10.1109/TBCAS.2015.2461446

Thorne, B. A., Takeya, R. K., & Peluso, R. W. (2009). Manufacturing recombinant adeno-associated viral vectors from producer cell clones. *Human Gene Therapy*, *20*(7), 707–714. https://doi.org/10.1089/hum.2009.070

Tigges, M., Marquez-Lago, T. T., Stelling, J., & Fussenegger, M. (2009). A tunable synthetic mammalian oscillator. *Nature*, *457*(7227), 309–312. https://doi.org/10.1038/nature07616

Vachon, V. K., & Conn, G. L. (2016). Adenovirus VA RNA: An essential pro-viral non-coding RNA. *Virus Research*, *212*, 39–52. https://doi.org/10.1016/j.virusres.2015.06.018

Van Den Berghe, P. V. E., Folmer, D. E., Malingré, H. E. M., Van Beurden, E., Klomp, A. E. M., Van De Sluis, B., Merkx, M., Berger, R., & Klomp, L. W. J. (2007). Human copper transporter 2 is localized in late endosomes and lysosomes and facilitates cellular copper uptake. *Biochemical Journal*, *407*(1), 49–59. https://doi.org/10.1042/BJ20070705

Xiao, X., Li, J., & Samulski, R. J. (1998). Production of high-titer recombinant adeno-associated virus vectors in the absence of helper adenovirus. *Journal of Virology*, *72*(3), 2224–2232. http://www.ncbi.nlm.nih.gov/pubmed/9499080%0Ahttp://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC109519

Xie, M., & Fussenegger, M. (2018). Designing cell function: assembly of synthetic gene circuits for cell biology applications. *Nature Reviews Molecular Cell Biology*, *19*(8), 507–525. https://doi.org/10.1038/s41580-018-0024-z

Xie, Z., Wroblewska, L., Prochazka, L., Weiss, R., & Benenson, Y. (2011). Multi-input RNAi-based logic circuit for identification of specific cancer cells. *Science*, *333*(6047), 1307–1311. https://doi.org/10.1126/science.1205527

Yang, L., Nielsen, A. A. K., Fernandez-Rodriguez, J., McClune, C. J., Laub, M. T., Lu, T. K., & Voigt, C. A. (2014). Permanent genetic memory with >1 byte capacity. *Nature Methods*, *11*(12), 1261–1266. https://doi.org/10.1038/nmeth.3147

Yang, Q., Chen, F., & Trempe, J. P. (1994). Characterization of cell lines that inducibly express the adeno- associated virus rep proteins. *Journal of Virology*, *68*(8), 4847–4856.

Yeoh, J. W., Ng, K. B. I., Teh, A. Y., Zhang, J. Y., Chee, W. K. D., & Poh, C. L. (2019). An Automated Biomodel Selection System (BMSS) for Gene Circuit Designs. *ACS Synthetic Biology*, *8*(7), 1484–1497. https://doi.org/10.1021/acssynbio.8b00523

Zajchowski, D. A., Boeuf, H., & Kédinger, C. (1985). The adenovirus-2 early EIIa transcription unit possesses two overlapping promoters with different sequence requirements for EIa-dependent stimulation. *The EMBO Journal*, *4*(5), 1293–1300. https://doi.org/10.1002/j.1460-2075.1985.tb03775.x

Zhang, B., Georgiev, O., Hagmann, M., Günes, Ç., Cramer, M., Faller, P., Vasák, M., & Schaffner, W. (2003). Activity of Metal-Responsive Transcription Factor 1 by Toxic Heavy Metals and H 2 O 2 In Vitro Is Modulated by Metallothionein . *Molecular and Cellular Biology*, *23*(23), 8471–8485. https://doi.org/10.1128/mcb.23.23.8471-8485.2003

Zhang, C., Konermann, S., Brideau, N. J., Lotfy, P., Wu, X., Novick, S. J., Strutzenberg, T.,

Griffin, P. R., Hsu, P. D., & Lyumkis, D. (2018). Structural Basis for the RNA-Guided Ribonuclease Activity of CRISPR-Cas13d. *Cell*, *175*(1), 212-223.e17. https://doi.org/10.1016/j.cell.2018.09.001

Zhou, X., Vink, M., Klaver, B., Berkhout, B., & Das, A. T. (2006). Optimization of the Tet-On system for regulated gene expression through viral evolution. *Gene Therapy*, *13*(19), 1382–1390. https://doi.org/10.1038/sj.gt.3302780