



HAL
open science

Developing new tools and platforms for mammalian synthetic biology: From the assembly and chromosomal integration of complex DNA circuits to the engineering of artificial intercellular communication systems

Xavier Duportet

► To cite this version:

Xavier Duportet. Developing new tools and platforms for mammalian synthetic biology: From the assembly and chromosomal integration of complex DNA circuits to the engineering of artificial intercellular communication systems. Biotechnology. Université Paris Diderot (Paris 7), 2014. English. NNT: . tel-01108520

HAL Id: tel-01108520

<https://theses.hal.science/tel-01108520>

Submitted on 22 Jan 2015

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

UNIVERSITE PARIS DIDEROT (Paris7)

ECOLE DOCTORALE FRONTIERES DU VIVANT

Centre de Recherches Interdisciplinaires
8-10 rue Charles V
75004 Paris

DOCTORAT

Discipline: Biologie de Synthèse

XAVIER DUPORTET

**Developing new tools and platforms for mammalian synthetic biology:
From the assembly and chromosomal integration of complex DNA
circuits to the engineering of artificial intercellular communication
systems**

Thèse dirigée par Gregory Batt et Ron Weiss

Au sein des laboratoires

Lifeware

INRIA Rocquencourt
Domaine de Voluceau,
78150 Rocquencourt, France

Weiss Lab for Synthetic Biology

Massachusetts Institute of Technology
77 Massachusetts Ave, Cambridge, MA
02139, United States

Soutenue publiquement le 14 Novembre 2014

JURY

Reiner Veitia
Timothy Lu
Diego Di Bernardo
Didier Mazel
Franck Molina
Gregory Batt
Ron Weiss

Président
Rapporteur
Rapporteur
Examineur
Examineur
Co-directeur de thèse
Co-directeur de thèse

Abstract

Mammalian synthetic biology may provide novel therapeutic strategies, help decipher new paths for drug discovery and facilitate synthesis of valuable molecules. Yet, our capacity to program cells is currently hampered both by the lack of efficient approaches to streamline the design, construction and screening of synthetic gene networks, and also by the complexity of mammalian systems and our poor understanding of cellular processes context-dependencies. To address these problems, I proposed and validated a number of concepts and approaches during my PhD.

First, I created a framework for modular and combinatorial assembly of functional (multi)gene expression vectors and their efficient and specific targeted integration into a well-defined chromosomal context in mammalian cells. The potential of this framework was demonstrated by assembling and integrating different functional mammalian regulatory networks including the largest gene circuit built and chromosomally integrated to date (6 transcription units, 27kb) encoding an inducible memory device. Such a rapid and powerful prototyping platform is well suited for comparative studies of genetic regulatory elements, genes and multi-gene circuits as well as facile development of libraries of isogenic engineered cell lines.

Second, I developed a platform to identify and characterize new serine recombinase systems from Mycobacteriophage genomes in order to extend the toolbox of genome engineering tools available for mammalian cells programming. I validated the approach by identifying 26 new large serine recombinases from 400 Mycobacteriophage genomes, from which 4 were using new recombination sites. These recombinases could mediate site-specific recombination events in both *E. coli* and mammalian cells. Additionally, I demonstrated that a library of 6 orthogonal recombination site pairs could be engineered for each of these recombinases.

To overcome the apparent limitations in our single-cell rational engineering capacity, I also engineered two new artificial intercellular communication systems for mammalian cells, in order to facilitate the spatial decoupling of different modules of a synthetic circuit. The first one consists in synthetic sender/receiver modules that can be either integrated within the same cell population to create an autocrine-like system or integrated into two distinct populations of cell to create a paracrine-like system. To create the sender module, I assembled and stable integrated a synthetic metabolic pathway using different plants enzymes to produce a small diffusible molecule: phloretin. This small molecule, orthogonal to endogenous signaling

pathways, can be sensed by the receiver module I engineered, which relies on a de-novo synthetic inducible gene expression system combining bacterial and mammalian genetic parts. Based on previous development of Virus Like Particles (VLPs), I created a second intercellular communication system which enables the transfer of proteins from a sender cell population to a separate receiver cell population. I demonstrated that I could induce the budding of particles carrying recombinases (Cre and B3) from senders cell that could be delivered to the receiver cells and perform a targeted genomic rearrangement to activate transgene expression.

Even though we are still years away from therapies using engineered cells carrying synthetic circuits to repair damaged or non-functional organs or to create de-novo tissues, I believe the contributions developed during the course of my PhD could potentially be used to push mammalian synthetic biology forward, whether it is by helping fasten the development of therapeutically relevant synthetic circuits or by providing new means to better understand the underlying mechanisms of cellular processes.

Acknowledgments

First and foremost I wish to thank my advisors, Doctor **Gregory Batt**, research scientist at INRIA and Professor **Ron Weiss**, director of the Weiss Lab for Synthetic Biology at MIT.

Gregory Batt has been supportive since the day we met when I was a Master student in his lab back in 2009. Ever since, Gregory has supported me not only by providing a research assistantship over four years, but also academically and emotionally through the road to finish this thesis. He first helped me start our collaboration with Professor Ron Weiss and then guided me over the different projects I initiated. Importantly, he gave me the moral support and the freedom I needed to move on and to pursue various projects without objection. It was a privilege for me to work with Gregory and I am deeply thankful. I am also very grateful to **Ron Weiss** for his scientific advice and many insightful discussions and suggestions. Ron Weiss warmly welcomed me in his lab, let me explore many different topics and gave me the opportunity to present my work many times during this PhD. I would also like to thank **Pascal Hersen** who welcome me in his lab to allow me to perform experiments when I was in France. Without his warm support, I could not have achieved my work on synthetic intercellular communication.

I would like to thank my thesis committee **Reiner Veitia, Timothy Lu, Diego Di Bernardo, Didier Mazel and Franck Molina**, for taking the time to read my thesis, attend my PhD defense in paris and provide me with precious comments and advice.

I am deeply grateful to all my labmates for their continuous support and for their friendship throughout these 4 years of PhD. Lila Wroblewska was my closest friend during my stay at MIT and I would like to thank her for he patience, support, kindness and craziness which I will always remember. I would also like to thank Yinqing Li, Patrick Guye, Deepak Mishra and Adrian Sluzarczyk from the Weiss Lab, and Clément Vulin, Artémis Llamosi and Jean-Baptiste Lugagne from the Hersen lab. All these friends have contributed immensely to my personal and professional time. The group has been a source of friendships as well as good advice and collaboration. I also thank the three incredible interns I had the chance to work with, Julianne Rieders, Nikki Shah and Helena Shomar.

I thank Olivier Gandrillon and Reiner Veitia for their support and advices during the Thesis Advisory Committees of my PhD. I also thank all collaborators with whom we initiated many projects: Nadine Honoré from Institut Pasteur, Pr. Graham Hatfull from Pittsburg University, Sri

Kosuri from Harvard Medical School, Antoine Decrulle from INSERM and Justin Eyquem from Collectis.

I gratefully acknowledge the funding sources that made my Ph.D. work possible. I was funded by INRIA, DARPA CCM, DARPA Synbio BBN, NIGMS of the NIH; Syne2arti and Iceberg from the French National Research Agency.

Lastly, I would like to thank my family and friends for all their love and encouragement. For my parents who raised me with a love of science and supported me in all my pursuits. And most of all for my loving, supportive, encouraging, and patient girlfriend Julia whose faithful support during the final stages of this Ph.D. was so appreciated. Thank you.

Outcomes & Productions

• Publications

- **A platform for rapid prototyping of synthetic gene networks in mammalian cells.**

Duportet X, Wroblewska L, Guye P, Eyquem J, Rieders J, Batt G and Weiss R

Nucleic Acids Res, 2014 ; 42, 13440-13451

- **Development of sequence-specific antimicrobials using programmable CRISPR-Cas nucleases.**

Bikard D, Euler C, Nussenzweig P, Jiang W, **Duportet X**, A. Fischetti V, Marraffini L

Nature Biotechnology, 2014 ; 32, 1146-1150

- **Rapid, modular and reliable construction of complex mammalian gene circuits.**

Guye P, Li Y, Wroblewska L, **Duportet X**, Weiss R.

Nucleic Acids Res. 2013 Sep;41(16):e156.

- **Design and connection of robust genetic circuits.**

Randall A, Guye P, Gupta S, **Duportet X**, Weiss R.

Methods Enzymol. 2011;497:159-86.

- **The biological interpretation of metabolomic data can be misled by the extraction method used.**

Duportet X, B M Aggio R, Carneiro C and G Villas-Boas S.

Metabolomics, 2011, DOI: 10.1007/s11306-011-0324-1.

• Posters

- **Targeted and efficient integration of multi-unit genetic payloads in mammalian cells: a modular and robust platform for genetic engineering**

Precision Genome Engineering and Synthetic Biology: Designing Genomes and Pathways (C5), Keystone, USA, 2013

- **Synthetic and orthogonal intercellular communication systems for mammalian cells**

Synthetic Biology 6, London, UK, 2013

- **A landing pad platform for large circuit integration in mammalian cells**

Synthetic Biology 5, Stanford, USA, 2011

• Awards

- **Worldwide Innovation Challenge – Innovation 2030**

Laureate of the personalized medicine track with PhageX (France, 2014)

- **Concours National de Création d’Entreprises Innovantes**

Laureate of the « En Emergence » track with PhageX (France, 2014)

- **Rencontres Universités Entreprises AEF**

Best French « PhD student / Entrepreneur » Award with PhageX (France, 2014)

- **Gen9 Inaugural G-Prize Contest**

3rd Prize Winner – 2012 – Discovery of new recombinases (USA, 2012)

- **MISTI France grant** (MIT, 2012)

- **Associated Teams grant** (INRIA, 2012)

- **Young Entrepreneurs Initiative**

Laureate with Omeecs (USA, 2011)

- **Genopole National Biotech Startup Competition**

Finalist with Omeecs (France, 2011)

- **ESCP 300K Business Plan Competition**

« Coup de Cœur » prizewith Omeecs (France, 2011)

- **Synthetic Biology 5.0**

Selective travel award (USA, 2011)

• Talks

- **TEDx Paris Universités**

Ingredients for a vibrant tech-entrepreneurship ecosystem (Cité des Sciences, Paris, 2013)

- **New Technology Venture Accelerator**

La boîte à outils pour réussir aux Etats-Unis (Microsoft HQ, Paris, 2013)

- **VIVAGORA – Journée sur le Vivant Artificiel**

La biologie de synthèse, ça sert à quoi? Applications, perspectives, promesses (Paris, 2012)

• Teaching

- Synthetic Biology 101

Paris 7 Medical School (Paris, 2012-2013)

- Synthetic Biology: thinking out of the box

AgroParisTech (Paris, 2011-2013)

• Training

- Julianne Rieders

Master 2 student, 3 months internship, Center for Interdisciplinary Research

- Nikki Shah

MIT Undergrad Student, 6 months internship

- Tiffany Peng

MIT Undergrad Student, 4 months internship

- Helena Shomar

AgroParisTech Master 1 student, 6 months internship

• Not for profit activities

- Science, Innovation and Entrepreneurship Network / Hello Tomorrow Challenge

Co-founder & President (since 2011)

<http://www.hello-tomorrow.org>

• For profit activities

- PhageX SAS

Co-founder (France, 2014)

- XD Lab SARL

Synthetic Biology Consultant for Lung LLC (USA, 2014)

Vice-President [Osons La France](#) (France, 2014)

Table of contents

Acknowledgments	iii
Abstract	v
Outcomes & Productions	vii
Abstract.....	iii
Chapter 1 General Introduction	1
1. 1. Current state of the art of Synthetic Biology.....	3
1. 1. 1. When engineering is used to program the living: the promises of Synthetic Biology.....	3
1. 1. 2. An impressive development of functional genetic circuits in unicellular organisms.....	4
1. 1. 3. Extending synthetic biology to multi-cellular organisms: focus on mammalian cells	11
1. 2. Current limitations in the rational engineering of circuits for mammalian systems..	17
1. 2. 1. Lack of parts and circuits characterization platforms.....	17
1. 2. 2. The importance of the spatial decoupling of functions.....	18
1. 3. Approach: Developing new tools and platforms to move mammalian synthetic biology forward	20
1. 3. 1. A prototyping platform towards a better characterization of genetic parts and circuits..	20
1. 3. 2. Extending the site-specific recombination toolbox towards a streamlined manipulation of circuits and genomes	21
1. 3. 3. Orthogonal intercellular communication platforms towards a spatial decoupling of functions	21
1. 4. Contributions.....	23
1. 5. Outline of the thesis	25
PART 1 Developing tools and platforms to assemble, manipulate and integrate DNA circuits for synthetic biology applications	
Chapter 2 A platform for rapid prototyping of synthetic genetic networks in mammalian cells	29
2. 1. Introduction.....	31
2. 2. Design of the platform	33
2. 2. 1. Integration module: Chassis cell lines generation	33
2. 2. 2. Assembly module: Mammalian Modular Cloning (mMoClo).....	35
2. 3. Test of the platform	36
2. 3. 1. Site-specific integration into the Landing Pad.....	36
2. 3. 2. Expression homogeneity of multi-genic integrated constructs in a polyclonal cell population	38
2. 4. Experimental proof of concepts	39

2. 4. 1. Integration of a 7-gene regulatory network encoding an inducible memory device	39
2. 4. 2. Integration of a library of 18 circuits.....	42
2. 5. Conclusion and discussion	44
2. 6. Materials and methods.....	48

Chapter 3 Expanding the site-specific recombinases toolbox for mammalian synthetic biology 53

3. 1. Introduction.....	55
3. 2. Discovery of new mycobacteriophage large serine recombinases	59
3. 2. 1. Identification of large serine recombinases.....	59
3. 2. 2. Validation of the putative recombinase activity in <i>M. smegmatis</i>	60
3. 2. 2. 1. Creation of integrative vectors for <i>Mycobacterium</i> species.....	60
3. 2. 2. 2. Transformation of <i>Mycobacterium smegmatis</i> with the integrative vectors	62
3. 3. Discovery of new recombination sites.....	63
3. 3. 1. Identification of the integration site in <i>M. smegmatis</i> genome	63
3. 3. 2. Validation of functional attB/attP recombination sites.....	67
3. 4. Validation of recombinases activity in mammalian cells	69
3. 4. 1. Assessment of intermolecular recombination.....	69
3. 4. 2. Assessment of recombination crosstalk.....	71
3. 4. 3. Creation of a library of orthogonal sites.....	71
3. 5. Conclusion and discussion	73
3. 6. Materials and methods.....	75

PART 2 Engineering synthetic and orthogonal intercellular communication systems in mammalian cells

Chapter 4 Hijacking plant metabolic pathways and bacterial nuclear receptors to create a de-novo small molecule based intercellular communication system 82

4. 1. Introduction.....	84
4. 2. Engineering of the receiver module	86
4. 2. 1. Phloretin responsive mammalian promoter (PRMP).....	86
4. 2. 1. 1. Design.....	86
4. 2. 1. 2. Influence of TtgR's operator sequence on the induction profile	88
4. 2. 1. 3. Influence of transactivator vs. repressor ratio on induction profile	89
4. 2. 2. Creation and characterization of stable receiver cell lines	91
4. 2. 2. 1. Lentiviral mediated integration of the receiver module components.....	91
4. 2. 2. 2. Characterization of induction profile of different receiver cell lines	93
4. 3. Engineering of the sender module	95
4. 3. 1. Phloretin synthesis	95

4. 3. 2. Creation of a stable sender cell line	95
4. 4. Validation of functional cell-to-cell communication.....	96
4. 5. Conclusion and discussion	99
4. 6. Materials and methods.....	103
Chapter 5 Hijacking lentiviruses to create a multi-channel protein-based intercellular communication system	106
5. 1. Introduction.....	108
5. 2. Engineering of a protein-based intercellular communication	111
5. 2. 1. Design	111
5. 2. 2. Validation of sender and receiver modules proper functioning.....	113
5. 3. From single-channel to multichannel intercellular communication	114
5. 3. 1. Design	114
5. 3. 2. Validation of the AND gate functioning.....	116
5. 4. Conclusion and discussion	117
5. 5. Materials and methods.....	119
Chapter 6 General conclusion	122
6. 1. Summary	124
6. 2. Perspectives.....	126
Chapter 7 Bibliography	130
Chapter 8 Annexes	135
Annexes for Chapter 2 - A platform for rapid prototyping of synthetic genetic networks in mammalian cells.....	137
Annexes for Chapter 3 - Expanding the site-specific recombinases toolbox for mammalian synthetic biology.....	165
Annexes for Chapter 4 - Hijacking plant metabolic pathways and bacterial nuclear receptors to create a de-novo small molecule based intercellular communication	169
Annexes for Chapter 5 - Hijacking lentiviruses to create a multi-channel protein-based intercellular communication system.....	171
Chapter 9 Manuscripts, reports and presentations.....	175

Chapter 1 General Introduction

Contents

1.1 Current state of the art of Synthetic Biology

- 1.1.1 When engineering is used to program the living: the promises of Synthetic Biology
- 1.1.2 An impressive development of functional genetic circuits in unicellular organisms
- 1.1.3 Extending Synthetic Biology to multi-cellular organisms: focus on mammalian cells

1.2 Current limitations in the rational engineering of circuits for mammalian systems

- 1.2.1 Lack of parts and circuits characterization platforms
- 1.2.2 The importance of spatial decoupling of functions

1.3 Approach: Developing new tools and platforms to move mammalian Synthetic Biology forward

- 1.3.1 A prototyping platform towards a better characterization of genetic parts and circuits
- 1.3.2 Extending the site-specific recombination toolbox towards a streamlined manipulation of circuits and genomes
- 1.3.3 Orthogonal intercellular communication systems towards a spatial decoupling of functions

1.4 Contributions

1.5 Outline of the thesis

We will start this introduction by trying to define Synthetic Biology and depicting its state of the art in unicellular organisms. We first present a non-exhaustive list of examples demonstrating the successful advancements in the field. We will then illustrate the state of the art of Synthetic Biology in multicellular systems with a specific focus on mammalian cells. By listing most appealing works and applications of Synthetic Biology for mammalian systems, we will highlight the current major limitations that bioengineers will have to overcome to move the field forward. We will finally give an overview of the different approaches I have decided to implement towards this goal and list the contributions I have brought to the field through my PhD.

1. 1. Current state of the art of Synthetic Biology

1. 1. 1. When engineering is used to program the living: the promises of Synthetic Biology

Stéphane Leduc was the first scientist to use the term “Synthetic Biology”, in 1912, in his publication: “Théorie physique-chimique de la vie et générations spontanées” and then in “La Biologie Synthétique”.

One has to wait until 1974 for the terms to be used again. It was the Polish geneticist: Waclaw Szybalski who wrote about synthetic biology:

“Let me now comment on the question “what next”. Up to now we are working on the descriptive phase of molecular biology. ... But the real challenge will start when we enter the synthetic biology phase of research in our field. We will then devise new control elements and add these new modules to the existing genomes or build up wholly new genomes. This would be a field with the unlimited expansion potential and hardly any limitations to building “new better control circuits” and.... finally other “synthetic” organisms, like a “new better mouse”. ... I am not concerned that we will run out of exciting and novel ideas, ... in the synthetic biology, in general.”

Since then, many definitions of Synthetic Biology have been proposed given the large diversity of goals and ways to achieve them it encompasses. From biosensing of metabolites to the biomanufacturing of high added-value molecules, from the engineering of DNA based drugs to the creation of protocells and minimal genomes, Synthetic Biology has goals and methodologies which are specific to very different fields (Biochemistry & Molecular Biology, Biotechnology & Applied Microbiology, Chemistry, Life Sciences & Medicine, Computer Science, Engineering,

Mathematical & Computational Biology, Cell Biology) which is why an universal definition of synthetic biology is hardly possible.

Though, a consensus definition has recently been drafted by a group of European experts defining Synthetic Biology as follows: 'Synthetic biology is the engineering of biology: the synthesis of complex, biologically based (or inspired) systems, which display functions that do not exist in nature. This engineering perspective may be applied at all levels of the hierarchy of biological structures—from individual molecules to whole cells, tissues and organisms. In essence, synthetic biology will enable the design of 'biological systems' in a rational and systematic way'

(Synthetic Biology: Applying Engineering to Biology: Report of a NEST High Level Expert Group).

The promises of both fundamental and applied synthetic biology led to a significant and increasing global interest towards this emerging field in the last 10 years, as demonstrated by:

- the increase in the number of publications: from a handful in 2004 to more than 400 in 2012
- the increase in the number of authors: from a dozen in 2004 to more than 1400 in 2012
- the impressive budget allocated to this new field in major countries in 2013 (~\$90M in Europe, ~\$500M in the USA and more than \$2B in China).

1. 1. 2. An impressive development of functional genetic circuits in unicellular organisms

For practical applications or to offer fundamental insight into biological processes, scientists have managed to design, modify or redesign internal genetic circuitry of unicellular biological systems.

In order to be able to rationally design artificial gene regulatory networks with robust function, it is first critical to be able to identify primary genetic elements: building blocks. It is then necessary to have the means to manipulate them at will towards a good understanding of how they work together and how they interfere with each other.

Such operation can be perfectly illustrated by the work of Chan et al⁴, who engineered a well-studied biological system when refactoring the genome of the T7 bacteriophage (**Figure 1-1**). The T7 bacteriophage, sequenced in 1983, is a lytic phage that infects *Escherichia coli*. T7 is a challenging model for studying the functionality and the interconnection of the different genetic parts as most of its genome is based on overlapping promoter and gene sequences. Based on

previous (incomplete) knowledge and computational, quantitative models built to explore questions related to the organization of genetic elements on the T7 genome, Chan et al worked on the engineering of a surrogate genome encoding a functioning phage in order to better understand how all parts have to be organized to ensure functionalities.

By making over 600 simultaneous changes or addition to the wild-type genome DNA, Chan et al were the first to engineer on the genome scale a synthetic viable organism, insulating and manipulating known primary genetic elements. They managed to translate a natural small but complex and overlapping gene regulatory network into an artificial circuit based on independent and insulated primary genetic elements (Figure 1-1).

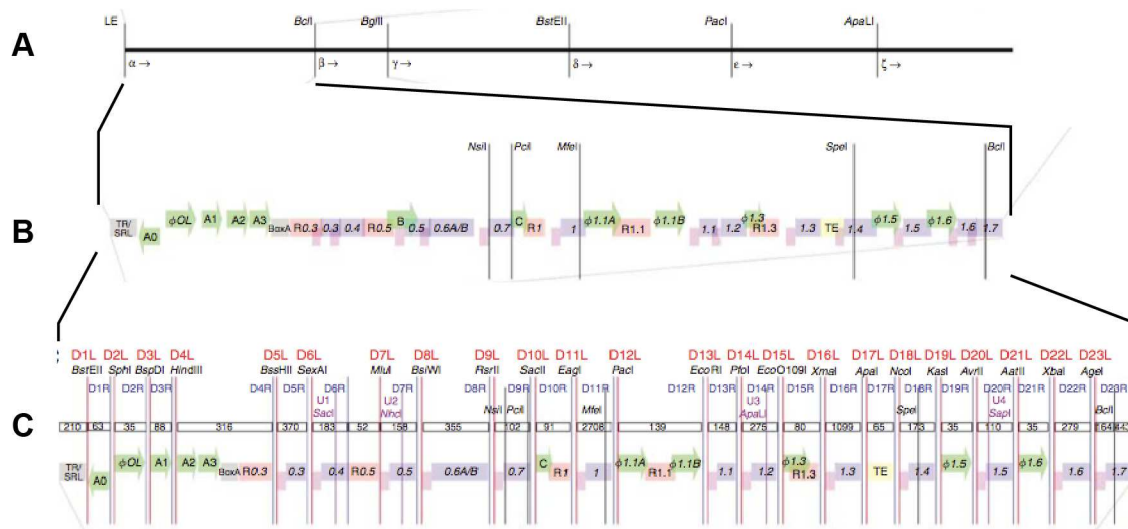


Figure 1-1 Refactoring of the T7 genome¹. (A) The natural T7 genome was split in 6 sections, alpha through zêta, using five restriction sites, unique across the natural sequence. (B) Details of the wild-type genetic elements composing the alpha section of the T7 genome: protein coding regions (blue) RBSs (purple), promoters (green), RNaseIII recognition sites (pink) and others (grey). (C) The same alpha section but refactored (T7.1) by removing all genetic overlaps.

Thanks to the recent and impressive development of modern biological tools (including DNA sequencing, DNA synthesis and DNA assembly), the ability to manipulate basic elements allowed scientists to not only characterize and standardize a constantly growing library of these parts, but also to go beyond refactoring the existing. They indeed created modules with artificial behaviors depending on the specific combination of the different parts (promoters, RBS, transcriptional repressors or activators, terminators, etc....). The largest library of these independent modules has been developed for bacterial systems: switches, time-delayed circuits, cascades, pulse generators, logic gates, spatial patterning and memory devices 2-7.

The first synthetic modules were developed in *Escherichia coli* in 2000. Gardner et al developed a genetic bi-stable toggle switch ² using two regulatory proteins: the Lac and Tet repressors, and based on a simple mathematical model (**Figure 1-2**). They demonstrated that the switch could be flipped from an off state to an on state and vice-versa upon induction with IPTG or aTc. The same year, Elowitz and Leibler developed a synthetic oscillatory network based on three transcriptional regulators ⁸. These works were the first to demonstrate the feasibility of designing, constructing and validating the proper functioning of artificial genetic networks built from simple genetic components from various backgrounds (genetic context or organism for instance).

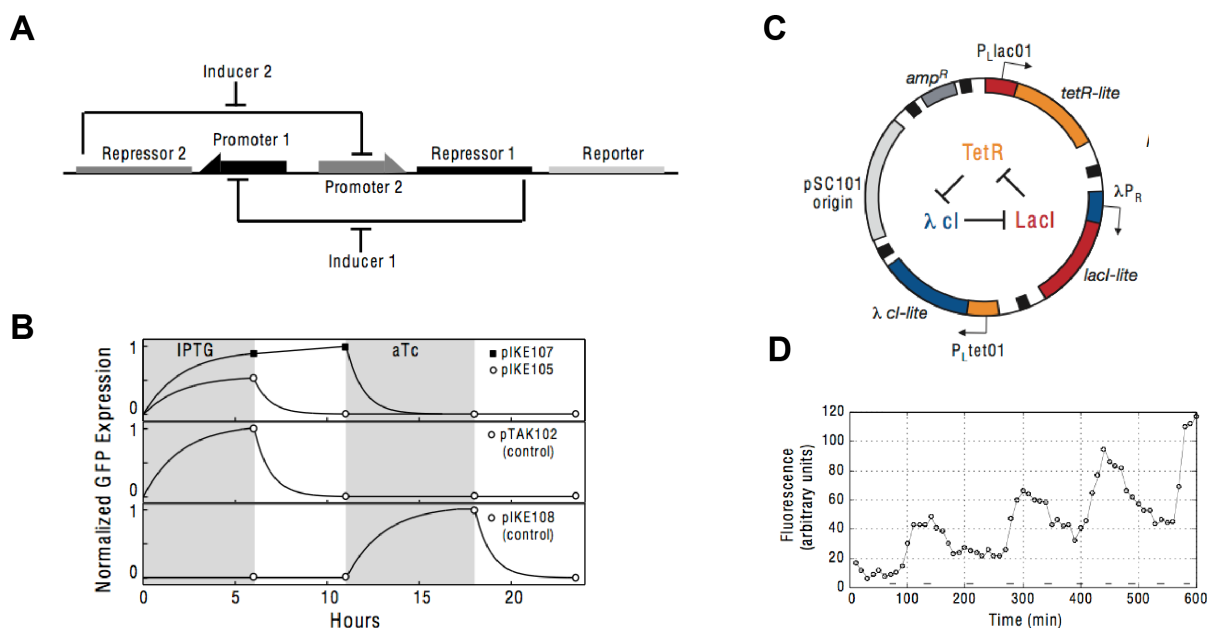


Figure 1-2 Toggle switch and repressilator designs and behaviors ^{2,8}. (A) Repressor 1 inhibits transcription from Promoter 1 and its expression is derepressed by Inducer 2. Repressor 2 inhibits transcription from Promoter 2 and is induced by Inducer 2. (B) Demonstration of bistability with different toggle switch designs. The grey shading indicates periods of chemical induction. (C) The repressilator network: it is a negative-feedback loop composed of three repressor genes and their corresponding promoters. (D) Synthetic oscillations induced by the repressilator circuit in living bacteria measured by GFP fluorescence over time.

Since these foundational works, a large number of modules encoding logic formulas have been designed involving a large panel of transcriptional or translational regulatory proteins ^{6,9}.

For instance, Moon et al ⁴ developed complex logic formula circuits based on the successful layering of single orthogonal logic gates, i.e. not interfering with each other. By having the output of a single gate being the input of the next gate, they created different programs using different gate permutations (**Figure 1-3**). One of their circuit is actually the largest one ever assembled,

consisting of 11 different regulatory proteins. To achieve a robust functioning of these programs, they had to do an extensive part mining, which proved valuable to expand the number of orthogonal gates they used.

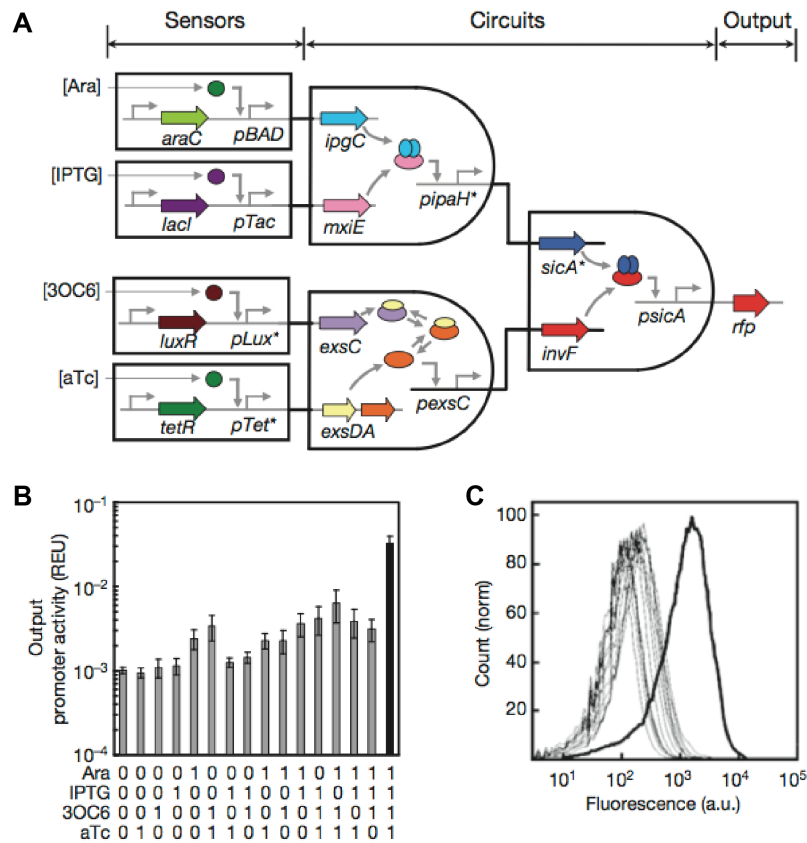


Figure 1-3 Genetic programs formed by layering AND gates⁴. (A) The 4-input AND gate consists of 4 sensors, an integrated circuit and a reporter gene. (B) The output fluorescence for different combinations of inputs: showing at least one order of magnitude difference for the output between the off states on the on state. (C) Raw cytometry data for all input states. The thick line is for [1111]

Beyond the design and construction of synthetic circuits to control single-cell behavior, scientists develop modules and circuits to program the dynamics at the population level. You et al¹⁰ created a cell-cell communication module in Escherichia coli and coupled it to cell survival and death mechanisms to autonomously regulate the density of the engineered cell population (Figure 1-4). They hijacked the well-characterized Lux system from the marine bacterium Vibrio fischeri to constitutively express the LuxI protein in Escherichia coli, and therefore the synthesis of the small and diffusible acyl-homoserine lactone (AHL) signaling molecule. The concentration of AHL in the growth media is therefore increasing with the density of engineered cells. By linking the subsequent binding and activation of the LuxR transcriptional regulator by

AHL to the induction of a killer protein, they demonstrated the autonomous regulation of the bacterial population at desired density levels, as predicted by mathematical models.

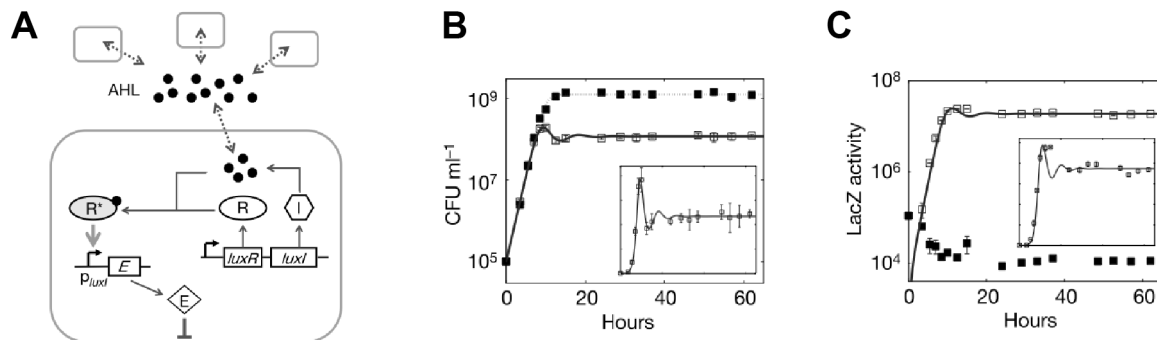


Figure 1-4 The population control circuit ¹⁰. (A) Schematic diagram of the circuit. E is a « killer » gene. I, R and R* represent LuxI, LuxR and active LuxR, respectively. Filled circles represent AHL. Constitutive expression of the LuxI protein results in increasing concentration of AHL in the media, which binds to LuxR and induces the expression of the killer gene E. (B) Experimentally measured growth curves and (C) corresponding levels of LacZ-CcdB when the population-control circuit is OFF (filled squares) and ON (open squares).

M-T. Chen et al extended the proof of concept of a rationally engineered coordinated cellular behavior at the cell population level in a more complex unicellular organism, the yeast *Saccharomyces cerevisiae* ¹¹. They indeed created an artificial and rationally designed intercellular communication system, which could mimic natural bacterial quorum-sensing behavior (**Figure 1-5**). By using parts originating from the plant *Arabidopsis thaliana*'s natural cytokine signaling pathway (synthesis enzyme and receptor component), they managed to render the system completely orthogonal to endogenous yeast pathways. To enable signal integration in yeast, they created a hybrid yeast phosphorylation-signaling pathway responsive to the engineered-receptor activation.

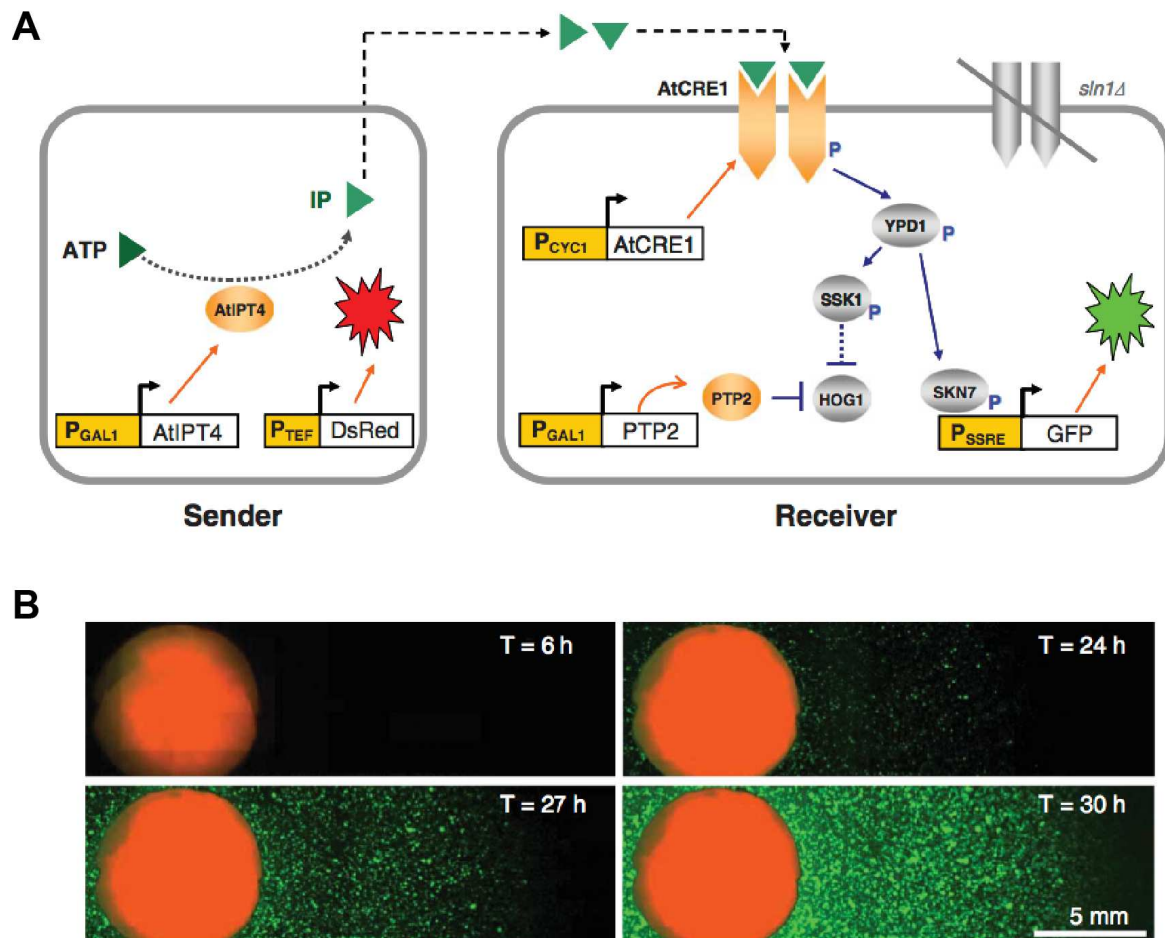


Figure 1-5 Engineered sender-receiver communication in yeast¹¹. **(A)** Diagram of the synthetic signaling pathways for sender and receiver cells. The sender module expresses recombinant *A. thaliana* enzyme AtIPT4 to synthesize and secrete IP to nearby receiver cells. The receiver module expresses *A. thaliana* AtCRE1 cytokinin receptor and yeast YPD1 and SKN7 signaling proteins. When IP signal binds AtCRE1, AtCRE1-YPD1-SKN7 phosphorylation activates GFP expression from the SSRE promoter in receiver cells. **(B)** Microscopy observations of sender-receiver communication: sender cells are fluorescing in red while the receiver cells are gradually fluorescing in green over time. This experiment demonstrates the spatial diffusion of the signaling molecule IP secreted by the sender cells.

Based on an increasing list of validated parts and modules, bioengineers have managed to create complex but reliable artificial systems built upon the connection of many of these modules. One of the best examples is the “sensing array of radically coupled genetic biopixels” by Prindle et al.¹², which has been designed based on previously characterized modules (**Figure 1-6**). By coupling two different intercellular communication systems to single-cell oscillators, they managed to synchronize single-cell oscillations across the whole population, while significantly reducing the intercellular variability. The authors demonstrated not only the impressive robustness of the circuit functioning in a noisy cellular environment, but also proposed an

extension of the usage of their complex though robust genetic program to real world applications such as the detection of heavy metals and pathogens ¹².

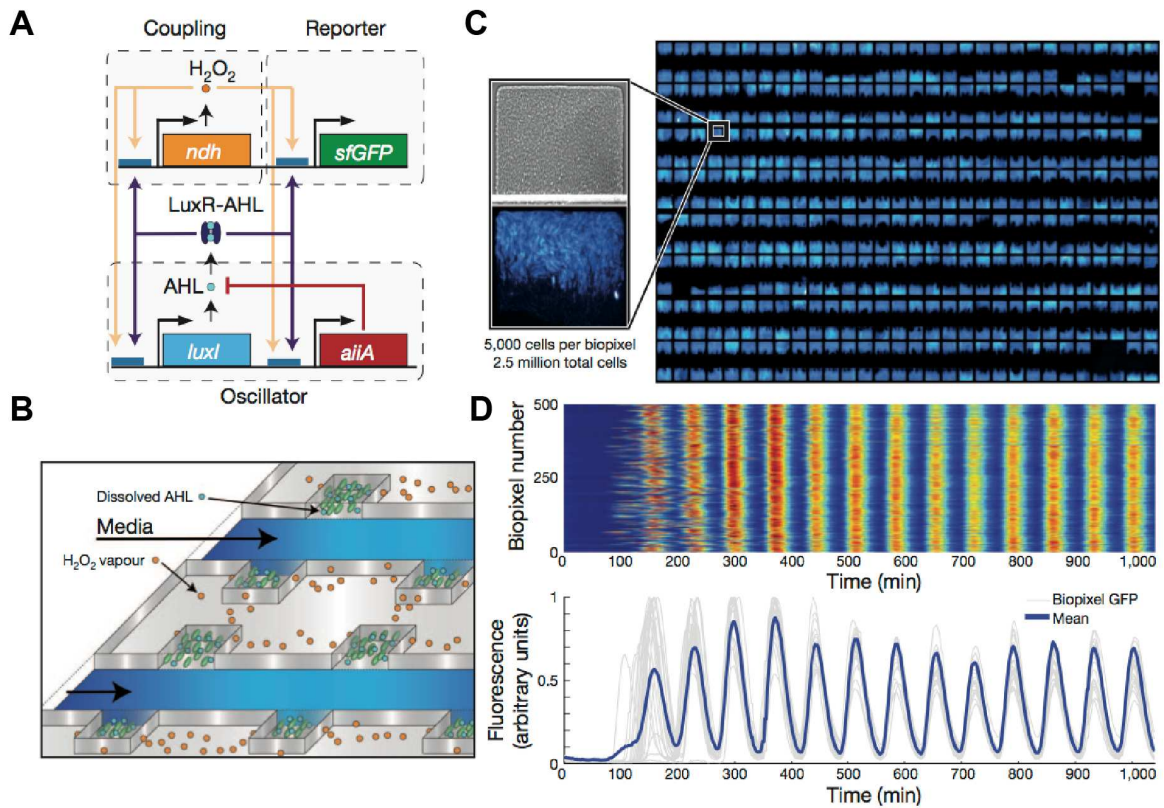


Figure 1-6 Sensing array of radically coupled genetic biopixels ¹². (A) Network diagram. The *luxI* promoter drives expression of *luxI*, *aiiA*, *ndh* and *sfGFP* (superfolder variant of GFP) in four identical transcription modules. The quorum-sensing genes *luxI* and *aiiA* generate synchronized oscillations within a colony via AHL. The *ndh* gene codes for NDH-2, an enzyme that generates H_2O_2 vapor, which is capable of migrating between colonies and synchronizing them by activating *luxI* promoter. (B) Conceptual design of the sensing array. AHL diffuses within colonies while H_2O_2 migrates between adjacent colonies through the PDMS. (C) Fluorescent image of an array of 500 *E. coli* biopixels containing about 2.5 million cells. Inset, bright-field and fluorescent images display a biopixel of 5 000 cells. (D) Heat map and trajectories depicting time-lapse output of 500 individual biopixels undergoing rapid synchronization. Sampling time is 2 min.

1. 1. 3. Extending synthetic biology to multi-cellular organisms: focus on mammalian cells

Most of cell behavior engineering has relied on bacteria as a testing ground. We have seen that it has been relatively successful, even when implementing complex artificial circuits.

In 2004, the first synthetic mammalian genetic switch was engineered by Kramer et al.¹³, four years after the creation of its counterpart in bacteria by Gardner et al.². Kramer et al created a switch that could be tuned to turn ON or to turn OFF the secretion of the enzyme alkaline phosphatase in Chinese hamster ovary cells. Reaching this first milestone for mammalian synthetic biology led to an increased interest in the field. Since then, efforts have been focused on transposing and refactoring the previously engineered bacterial modules for use in mammalian cells: logic circuits, memory devices and time-delay systems. (see Auslander et al for a review)¹⁴.

The motivation behind these studies is based on the fact that extending our engineering ability to multi-cellular organisms, especially mammalian systems, could provide solutions for extremely appealing applications, especially in human health: from the fundamental deciphering of cell biology to the design of next-generation therapeutic treatments.

We will review here a few systems that have been developed for therapeutic purposes, which will help us define current limitations and critical milestones that have yet to be achieved to move synthetic biology closer to real world applications.

For instance, Weber et al develop a synthetic drug-discovery platform based on a mammalian synthetic DNA circuit¹⁵ that they used to screen small molecules that could potentiate a *Mycobacterium tuberculosis* antibiotic. They demonstrated that once a target had been identified, synthetic genetic modules could be designed and constructed for whole-cell screening assays. For multidrug-resistant tuberculosis bacteria to be efficiently killed, the only treatment currently available is the use of ethionamide, which has to be activated by the endogenous enzyme EthA in order to be potent. However, due to transcriptional repression of *ethA* by the protein EthR, ethionamide-based therapy is often rendered ineffective. To address this critical issue, the researchers built a synthetic gene circuit in mammalian cells to screen for and identify EthR inhibitors that could reduce the resistance to ethionamide (**Figure 1-7**). The circuit consisted of reporter gene that could be activated by a chimeric EthR-based transcriptional activator.

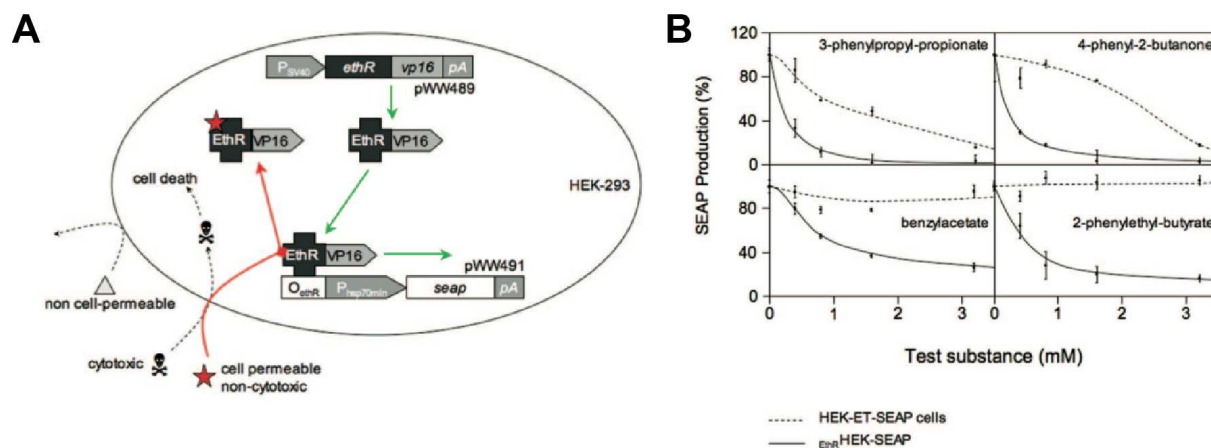


Figure 1-7. EthR-based drug discovery platform in mammalian cells¹⁵. (A) A gene fusion of ethR with the Herpes simplex-derived vp16 transactivation domain is expressed under the control of the simian virus 40 promoter in HEK-293. The chimeric transactivator EthR-VP16 binds to its operator O_{ethR} thereby activating transcription from the minimal *Drosophila* heat shock 70 promoter, driving expression of human placental secreted alkaline phosphatase (SEAP), which concentration is easily quantifiable. In the presence of a cell-permeable and non-cytotoxic molecule that can induce the release of EthR-VP16 from the promoter, the transcription of SEAP is then turned OFF (red lines). Non-cell-permeable or cytotoxic compounds are therefore automatically excluded from the hit list. (B) Effect of different rationally designed compounds on SEAP production, using the EthR-based gene network and illustrating the ability of these compounds to induce EthR-VP16 release from its promoter with different dynamics.

By creating and testing such circuits derived from bacterial components directly within mammalian cells, the candidate drugs are enriched in non-toxic and membrane-permeable molecules, which are critical drug criteria for all intracellular pathogens. Such framework could easily be extended to other diseases for which pathogenic mechanisms can be highlighted by intracellular engineered circuits.

Beyond using synthetic biology methodologies to discover new drugs, researchers have developed synthetic networks which act like a drug themselves. For instance, Xie et al developed a diagnostic biocomputer network in mammalian cells¹⁶ to detect cancer specific molecular factors and trigger apoptosis depending on the diagnostic.

The circuit they developed is composed of three essential modules: a sensor module, a computational module and an actuation module (Figure 1-8). After being transiently delivered to the cell, the sensor circuit of the genetic circuit samples up to five miRNAs (which have been identified as being indicative of the presence of cancer). The sensor signals are fed in the second module, the computational module. This portion integrates the different input signals and compute AND or NOT operations, functioning as a “classifier” of a specific cell type. Depending on whether the levels of the sensors, looked upon as a group, are actually indicative of the cancer, the computational core causes the production of the hBax protein in the actuation module, which ends up killing the cell or leaving it live.

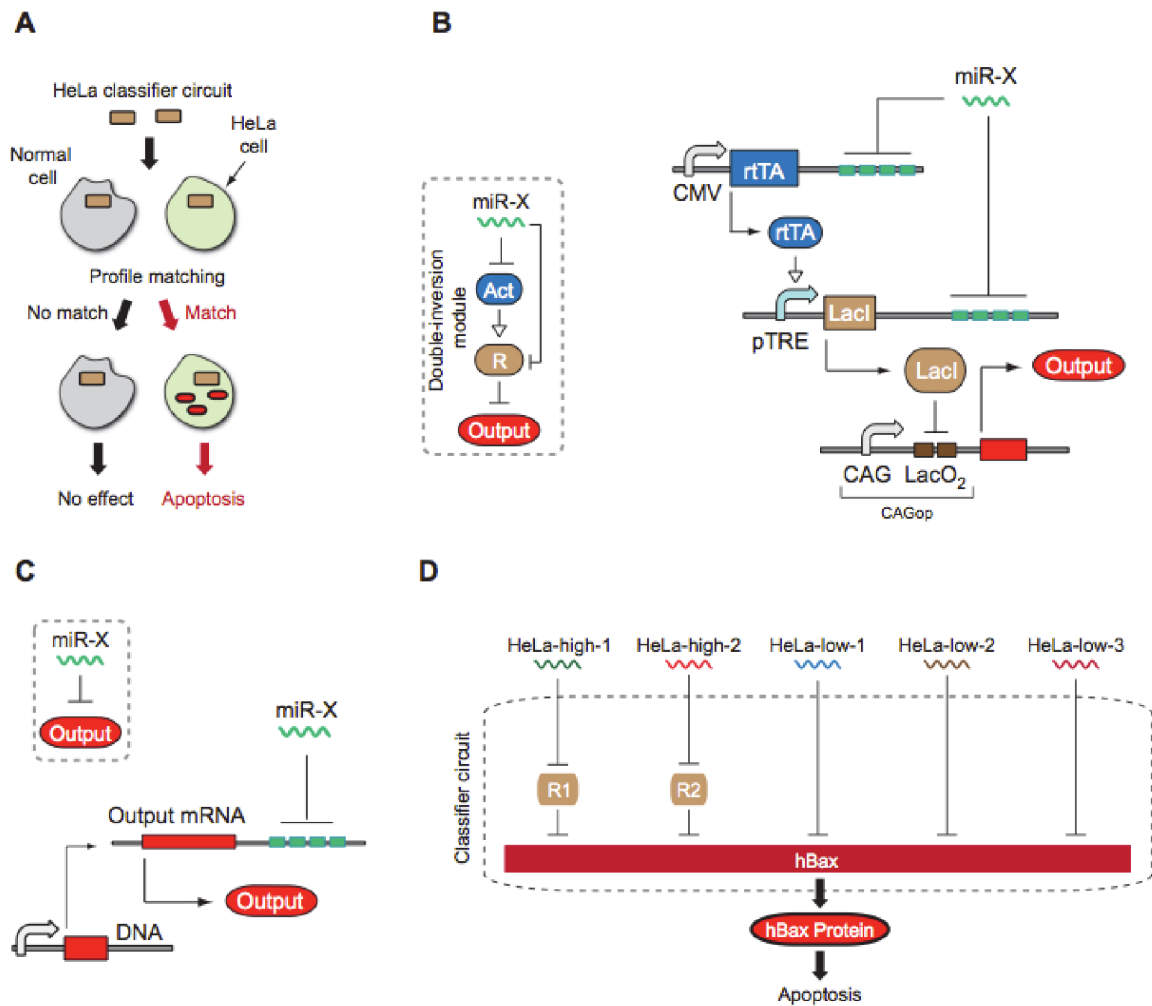


Figure 1-8 High-level architecture of a cell classifier ¹⁶. (A) Schematic representation of a HeLa-specific classifier circuit operation. Gray circle, healthy cells; light green, HeLa cells. (B) High-level and detailed description of double-inversion module for sensing HeLa- high miRNAs. Act, activator; R, repressor. (C) High-level and detailed description of HeLa- low miRNA sensor. (D) Schematic representation of an integrated multi-input classifier. The entire network implements a multi-input AND-like logic function for identification and selective killing of HeLa cells through regulated expression of hBax.

This work brought together two important concepts to create a system with high potential for therapeutic applications. The first is tissue-specific signalling, which has previously been used to restrict, or at least partially restrict, a therapeutic agent's action specifically to specific tissues/cell types, such as cancer cells. The second is the demonstration that multi-input information processing can work in mammalian cells. Such synthetic circuits that can both sense and modulate biological cell functions may offer the possibility to control the expression of genes of interests in targeted population of cells.

This work is therefore an extremely good proof of concept, it is however important to note that all results were obtained from *in-vitro* experiments based on transient transfections,

demonstrating a robust behavior of the system only on the short term. To apply such a concept for real world applications, the authors suggest that the DNA circuit would have to be stably integrated to enable long term expression of the transgenes. Given the high number of modules of the circuit and the fine tuning required for its robust behavior, integration of such system would likely be a very tedious and time-consuming process with current technologies.

In the same endeavour to create synthetic circuits for therapeutic applications, Kemmer et al engineered a system to restore uric acid homeostasis in the bloodstream ¹⁷ of mice which were subject to symptoms that could be caused by tumor lysis syndrome. The circuit is coupling a sensor module made of a modified *Deinococcus radiodurans* derived protein sensing uric acid levels to a dose-dependent actuation module that express an engineered urate oxidase eliminating uric acid (**Figure 1-9**).

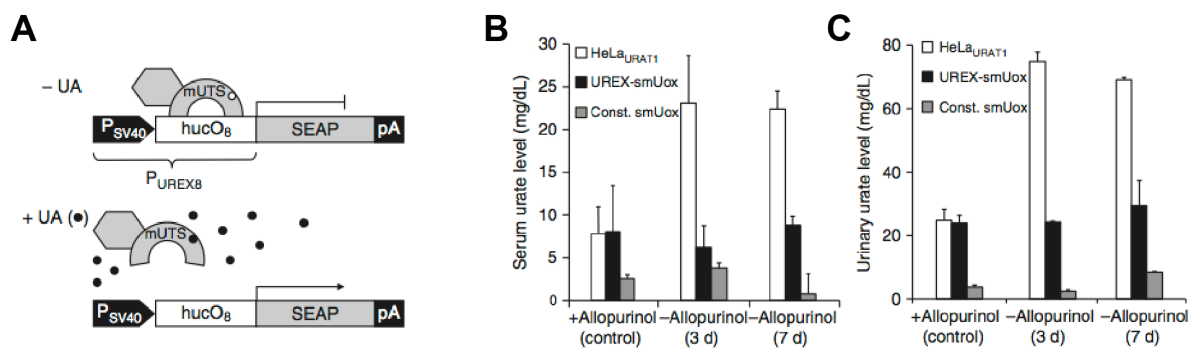


Figure 1-9 Synthetic uric acid-responsive mammalian sensor circuit ¹⁷. (A) Diagram of Uric acid Responsive Transgene Expression (UREX) in action. In the absence of uric acid (-UA), the fusion protein KRAB-HucR binds to huc operons and repress downstream transgene expression. In the presence of uric acid (+UA), KRAB-HucR is released from huc operons which derepresses UREX promoter and results in downstream transgene expression. (B) & (C) UREX-controlled reduction of pathologic urate levels in mice. In this case, SEAP transgene has been replaced by an urate reductase so as to be able to rapidly metabolise urate when its levels are too high. Urate levels were profiled in serum (B) and urine (C) of the animals 3 and 7 days after implantation of the encapsulated engineered cells. Urate levels were maintained constant both in the serum and in the urine when engineered cells were present, while without any engineered cells, the urate levels attained pathological levels rapidly once allopurinol (urate level stabilizer) was not fed to the mice anymore.

The researchers demonstrated that when implanting bioengineered cells inside a urate oxidase-deficient mouse, the circuit could read high uric acid levels in the bloodstream and trigger the secretion in a dose dependent manner the enzyme urate oxidase to neutralize the acid until blood levels return to normal. In order to work, the engineered cells were sealed within porous alginate capsules to protect them from the immune system and injected into the mice's body cavities. Even though they managed to stably integrate this relatively small circuit (2

transcription units) in mice cells, it is also important to mention that they had to isolate and screen a large number of clonal populations of engineered cells in order to find one population with the expected behavior.

Another yet more complex approach for developing next-generation tissue engineering strategies is the design of synthetic intercellular communication systems in which mammalian cells that can receive, process and transfer signals of choice. Although synthetic intercellular communication systems have been constructed in prokaryotes and yeast, transferring these systems into mammalian cells is not trivial due to their much greater complexity. Towards this goal, Bacchus et al engineered a synthetic multicellular communication system as a proof of concept for next-generation tissue engineering strategies ¹⁸ (**Figure 1-10**). They stably integrated a sender module and a receiver module in separate cell populations and created a device to program temporal permeability in vascular endothelial cell layers, mimicking natural control systems in the development of vertebrate extremities and vasculature.

Despite these promising advances, one can observe that we don't currently witness an increase in the complexity of synthetic circuits, which is one of the key to the development of therapeutically-relevant applications. The most probable explanation seems to be that it is currently hardly possible to develop reliable means for connecting together small functional circuits in order to realize larger networks with predictable behaviors.

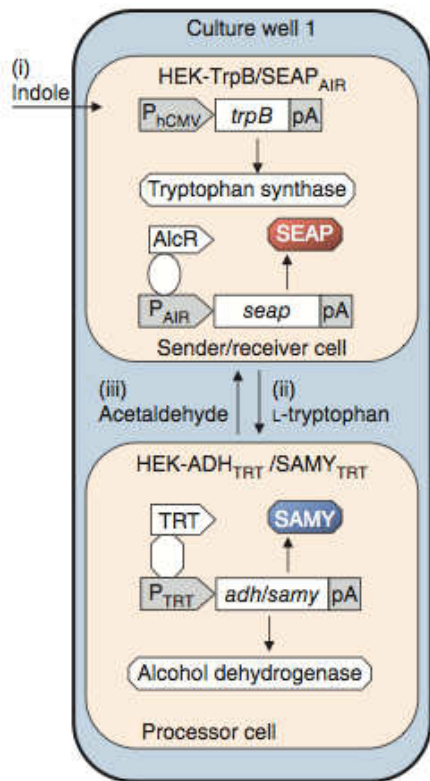


Figure 1-10 Two-way communication device¹⁹. Indole (signal (i)) is converted into L-tryptophan (signal (ii)) by HEK-TrpB/SEAP_{AIR} (sender/receiver cells), which is transmitted to HEK-ADH_{TRT}/SAMY_{TRT} (processor cells) where it triggers expression of SAMY as acknowledgment of receipt of signal (ii). Alcohol dehydrogenase is produced, converting spiked ethanol into acetaldehyde (signal (iii)), which is broadcast back to the sender/receiver cells where it triggers SEAP expression as acknowledgment of signal (iii) receipt.

1. 2. Current limitations in the rational engineering of circuits for mammalian systems

Cambray et al defined rational engineering of biology as “a rational and transparent design process wherein systems are built from understandable components whose interconnected, composite behavior is predictable”. To achieve such a goal, modularity of the parts, orthogonality of the modules with the endogenous system and with other modules, and robustness of their input/output functioning are required to be able to predict and control biological processes. This, of course, encompasses a complete characterization of the different components used to bound uncertainties effects. Unfortunately, the biological complexity of mammalian cells, the unpredictable changes in their performance, as well as the inefficiency of current methods to modify their genome them and the lack of clear standardization protocols are still major challenges for their rational engineering and therefore for unleashing the potential of mammalian synthetic biology.

1. 2. 1. Lack of parts and circuits characterization platforms

System biologists have extensively observed numerous intra/extra cellular interactions in mammalian cells revealing critical context-dependency effects²⁰⁻²², which actually affect most of the parts used by genetic engineers to assemble artificial circuits. For instance, promoters currently used in synthetic circuits show different levels of activity depending on the cell lines they are tested in, the locus in which they are integrated and the differentiation states of the cells ²³⁻²⁵. Taking this into consideration, it could become tedious and time-consuming to streamline the transposition of circuits such as the cell-line classifier to a wide variety of cell lines given the fine-tuning it requires.

Although significant, the effect of cis/trans sequences on gene expression levels is currently underestimated when building synthetic circuits^{20,21}. Within a single transcription unit, inter-part regions, such as 5' and 3' UTRs play a critical role in regulating gene expression levels and its mRNA half-life and translation efficiency. There is however no clear standardization protocol of both parts and inter-parts sequences within the community. It could therefore become extremely hard to compare experimental results between experiments within a lab if circuits are for instance not assembled with the same assembly method. Discrepancies can be even stronger between labs given that the ends of parts like promoters or polyA signals are not specifically defined among the community.

Another layer of complexity has been revealed in the past years, as increasing evidence has been accumulated suggesting that cis-acting regulatory DNA sequences may not be the only determinants of gene expression, but that DNA transcription also depends on trans-acting regulatory sequences²⁶, the genomic position of genes²⁷, the sub-nuclear localization of DNA sequences, and a complex interplay of the genome with specific features of nuclear architecture²⁸.

To overcome some of these limitations, strategies have been recently developed to insulate the synthetic transcription units from cell-context dependencies²⁹⁻³¹. Maldini et al have for instance tried to tailor a gene cassette design for sustainable gene transfer. They demonstrated that specific insulator sequences could significantly reduce promoter interference between the integrated transgene's promoter and the surrounding endogenous promoters from the integration locus.

Because they are still poorly understood and hardly engineerable, these context-dependencies drastically limit the characterization and therefore the predictability of parts, transcription units and synthetic circuits. In order to be able to rationally design and streamline the construction of functional genetic circuits, it is therefore critical to go beyond the simple observation of endogenous phenomena and create platforms to rapidly prototype and test circuits in a standardized manner.

1. 2. 2. The importance of the spatial decoupling of functions

Evolution of multi-cellular organisms heavily relied on the development of intercellular communication mechanisms. These mechanisms allowed them to organize themselves into consortia of specialized cells, which can execute and coordinate specific activities, spatially distributing the workload to increase overall fitness³². A broad range of intercellular communication mechanisms have been evolved across the multi-cellular organism kingdom^{19,33}. In humans, specialized cells from the immune system for instance, can communicate using signaling molecules that circulate in the bloodstream to regulate crucial biological processes³⁴. Intercellular communication can also be achieved by direct contact signaling, such as the delta-notch signaling pathway³⁵, which is highly conserved in most multicellular organisms. Among other actions, the notch signaling promotes in humans proliferative signaling during neurogenesis through transmembrane protein interaction.

When facing difficulties to engineer complex circuits, a spatial decoupling of the different modules between multiple engineered cell-types could be envisioned as an effective solution. For instance, Koseska et al and Prindle et al, demonstrated that synthetic multicellular consortia of communicating cell populations showed increased control precision and reliability ^{12,36}. Engineering specialized and interconnected mammalian cell populations may allow us to overcome the apparent limitations in our single-cell rational engineering capacity and therefore enable the design of increasingly complex multicellular gene networks ^{18,37,38}.

Such advance would have tremendous applications in tissue engineering, for instance the assembly of complex cellular patterns with new and useful functions or the design of synthetic hormone or immune systems. Towards this goal, several systems have been engineered in mammalian cells that respond to a variety of signaling mechanisms ^{18,39-43}. For instance, distinct sender and receiver cell populations have been implemented to create intercellular communication systems either by using direct cell to cell contact by engineered the delta-notch signaling pathway or by creating specialized cell populations with the ability to either produce or sense a variety of diffusible signaling molecules such as biotin, acetaldehyde, nitric oxide, L-arginine and L-tryptophan. Despite the demonstrated efficacy of these engineered systems as proof-of-concepts, they all present a major drawback for the development of elaborate spatially decoupled systems. Indeed, all the signaling molecules used in these approaches are either essential to our metabolism or significantly cross-talking with endogenous signaling pathways. For instance, Biotin is necessary for cell growth, the production of fatty acids, and the metabolism of fats and amino acids. Biotin assists in various metabolic reactions involving the transfer of carbon dioxide. It may also be helpful in maintaining a steady blood sugar level. L-arginine is one of the common 20 natural amino acids, found in almost any proteins and can be synthesized by our body. L-tryptophan, is also one of the essential amino acids in the human diet, and is a natural precursor of serotonin, a critical neurotransmitter in the brain. In mammals including humans, NO is an important cellular signaling molecule involved in many physiological and pathological processes. It is a powerful vasodilator with a short half-life of a few seconds in the blood. Finally, acetaldehyde is one of the most important aldehydes, occurring widely in nature and in common food such as coffee, bread, ripe fruit and plants. It is also produced by the partial oxidation of ethanol and is a contributing factor to hangovers from alcohol consumption. Although using such molecule may enable the circuits to be directly integrated with the patient's metabolic networks, the signal levels depend on environmental cues or metabolic activity, and can therefore deregulate the predicted functioning of synthetic intercellular networks or even interfere with endogenous regulatory mechanisms.

1. 3. Approach: Developing new tools and platforms to move mammalian synthetic biology forward

Major limitations and roadblocks are slowing down our advances in mammalian synthetic biology and the design of more therapeutically relevant circuits. Deeply motivated by the development of technologies to solve real world problems, I focused my PhD towards the overcoming of some of these major limitations in order to pave the way for impactful applications.

1. 3. 1. A prototyping platform towards a better characterization of genetic parts and circuits

When I started my PhD, I wanted to implement a very large and complex gene regulatory network to create a synthetic tissue which could autonomously regulate its density. This circuit was designed to be one of the numerous modules to create a beta-like synthetic stem cell tissue to treat type 1 diabetes. However, I rapidly realized that such a circuit would require a long-term robust and fine-tuned expression of dozens of transgenes. At that time, there were 1) no precise characterization of the parts I would need to use, and 2) no method to easily assemble, study and compare (large) circuits and their behavior over an extended period of time.

When looking at previously cited works, circuits were either transiently transfected, and therefore subjected to plasmid dilution over time, or randomly integrated with lentiviruses, resulting in very different expression levels in the stable cell lines created. Moreover, given the fact that integration sites were totally random and in various copy numbers, it would be almost impossible to compare different circuits' behavior as context dependency effects would be significantly different between cell lines.

I therefore decided to address these issues in order to have a long-term contribution for the field. Towards this goal, I developed a platform that couples 1) a modular assembly method to support a standardized architecture for transcription units and multi-gene circuits with 2) a recombinase-based targeted and stable integration strategy to enable robust expression of transgenes over time.

Such platform not only provides a mean to characterize parts, transcription units and circuit outputs over time in a standardized way across different cell lines but it also enables side by side comparison and high-throughput screening of large libraries of circuits. As a whole, such ability to easily prototype combinatorial designs may allow to bypass the current lack of predictability of parts and circuits, and fasten the pace of robust circuits development.

1. 3. 2. Extending the site-specific recombination toolbox towards a streamlined manipulation of circuits and genomes

In the last decades, a wide range of technologies have been developed to stably integrate transgenes in the genome of mammalian cells (viruses, transposons, nucleases). However, site-specific recombinases are the only that combine both the specificity in the integration target site and the ability to integrate extremely large pieces of DNA. There are therefore an extremely interesting strategy to use for a targeted integration of large synthetic circuits. However, only a handful of these enzymes have yet been discovered and we are therefore limited in the sequences that can be used to promote site-specific recombination. Natural recombination sites are indeed very specific and can not yet be engineered to resemble specific sequences from target loci in mammalian genomes. They therefore need to be preintegrated with another integration strategy (nuclease for instance) beforehand. Moreover, the recombinases used for the manipulation of mammalian genomes are extremely hard to engineer given the complexity of the site-specific recombination mechanism. I wanted to discover new recombinases to extend the currently available toolbox and pave the way for a better understanding of their structure/function relation. I specially focused on the large serine recombinases as they have shown to be the most promising ones in terms of efficiency and specificity for the manipulation of mammalian genomes.

Towards this goal, I set up a protocol to identify new large serine recombinases from phage genomes and find their attachment sites both on the phage genomes and on their bacterial counterpart. Finally, to validate their potential interest for mammalian applications, I tested their function in mammalian cells and tested libraries of orthogonal sites for all the recombinases I characterized.

1. 3. 3. Orthogonal intercellular communication platforms towards a spatial decoupling of functions

The synthetic intercellular communication systems that have been developed up to date use signaling molecules cross-talking with endogenous signaling pathways. I therefore worked on the development of orthogonal intercellular signalling systems that would not interfere with endogenous signalling pathways, in order to enable a rational engineering of complex functions relying on the spatial coordination of engineered cells.

Towards this goal, I worked in parallel on the development of two completely different but complementary systems. The first system I developed is inspired from existing communication systems relying on the use of small molecules as signaling molecules : bacterial quorum-sensing or mammalian hormonal systems. The second system I developed allows the encapsulation and the transfer of a variety of protein-based messages by a synthetic signal-packaging system. The system hijacks the properties of viral particles synthesis, release, transfer and import mechanisms. It allows each viral particle to be packed up with thousands of signaling molecules that can be carried over a long distance without being diluted in the extra cellular media. Combined together, these communication systems could enable the design of synthetic tissues containing multiple consortia of specialized cells spatially organized and interacting with each other.

1. 4. Contributions

Through these different projects, I made several contributions, relevant to both basic and applied research.

First, I engineered a system to rapidly construct and test complex genetic circuits in human and mouse cells. The method involves the generation of “chassis” cell lines, which contain a single copy of a landing pad for reliable site-specific genomic integration. I coupled it with a modular assembly method, mMoClo, which uses type IIS restriction enzymes and unique overhangs, enabling one-pot assembly of complete transcriptional units. As proof-of-principle, I created 6 different chassis cell lines, compared the integration efficiency of different versions of the BxB1 recombinase, assembled, integrated and validated the robust function over time of the largest synthetic circuit to date and a small library of variant expression constructs. The most significant advantage of this platform is that the integration strategy enables direct comparison of variants by providing an isogenic environment. This is a very useful property for characterizing and optimizing complex synthetic circuits in mammalian cells. To disseminate this platform and to fasten the pace of parts characterization across the community, more than 150 parts and vectors have been submitted to Addgene.

Second, I developed a platform to identify and characterize new serine recombinase systems from Mycobacteriophage genomes. It consists of a script to scan phage genomes and identify putative large serine recombinases. I also created a suite of plasmids to enable the discovery of their recombination sites in their natural host and to transpose the system in *E. coli* and mammalian cells. I validated the approach by identifying 26 new large serine recombinases from 400 Mycobacteriophage genomes, from which 4 were using new recombination sites. These recombinases could mediate site-specific recombination events in both *E. coli* and mammalian cells. Finally, I demonstrated that a library of 6 orthogonal recombination site pairs could be engineered for each of these recombinases.

Third, I engineered two new artificial intercellular communication systems for mammalian cells. The first one consists in synthetic sender/receiver modules that can be either integrated within the same cell population to create an autocrine-like system or integrated into two distinct populations of cell to create a paracrine-like system. To create the sender module, I assembled and stable integrated a synthetic metabolic pathway using different plants enzymes to produce a small diffusable molecule: phloretin. This small molecule, orthogonal to endogenous signaling pathways, can be sensed by the receiver module I engineered, which relies on a de-novo synthetic inducible gene expression system combining bacterial and mammalian genetic

parts. I demonstrated that in the presence of an inducer, co-culture experiments (sender with receiver) cells would result in the induction of transgene activation within the receiver cells.

Based on previous development of Virus Like Particles (VLPs), I created a second intercellular communication system which enables the transfer of proteins from a sender cell population to a separate receiver cell population. I demonstrated that I could induce the budding of particles carrying recombinases (Cre and B3) from senders cell that could be delivered to the receiver cells and perform a targeted genomic rearrangment to activate transgene expression.

As a whole, I believe these contributions could potentially be used to push mammalian synthetic biology foward and help fasten the development of therapeutically relevant synthetic circuits and systems.

1. 5. Outline of the thesis

In this thesis, I will present the different approaches described above with detailed results and discussions.

The first part of the thesis will be dedicated to the development and validation of new tools and platforms to assemble, integrate, manipulate and characterize DNA circuits in mammalian cells. I will first start by introducing the platform I developed for rapid prototyping of synthetic genetic networks in mammalian cells. After a detailed analysis of current technologies to stably integrate transgenes in mammalian genomes (and their limits), I will explain the rationale behind the design of the site-specific recombinase based landing pad platform. I will demonstrate its efficacy for circuit integration and its advantages for robust part and circuits characterization over extended period of time. In the second chapter, I will review the different site-specific recombinase technologies to illustrate the need to discover new large serine recombinases to both better understand their functioning and extend the toolbox for circuits integration and manipulations. I will then propose and validate a protocol to identify new recombinases from mycobacteriophage genomes. After demonstrating their functioning in mammalian cells, I will show that libraries of orthogonal recombination sites can be design for these recombinases.

The second part of the thesis will focus on the engineering of new intercellular communication systems based on signals orthogonal to endogenous signaling mechanisms, and which could therefore be used to spatially decouple functions encoded by synthetic circuits. After reviewing systems that have been developed to date, I will introduce a new system based on the secretion of a diffusable plant metabolite, which can activate a receiver module in engineered cells in a dose-dependant manner. Finally, I will demonstrate the ability to hijack lentiviruses to create a multi-channel intercellular communication system based on the transfer of proteins as signaling molecules.

The last part of this thesis consists in the synthesis of the overall work, putting it into the context of other approaches, followed by a discussion and numerous perspectives suggested by this work.

PART 1

Developing tools and platforms to assemble, manipulate and integrate DNA circuits for synthetic biology applications

Contents

Chapter 2

A platform for rapid prototyping of synthetic genetic networks in mammalian cells

Chapter 3

Expanding the site-specific recombinases toolbox for mammalian synthetic biology

Chapter 2 A platform for rapid prototyping of synthetic genetic networks in mammalian cells

Contents

- 2.1 Introduction**
 - 2.2 Design of the platform**
 - 2.2.1 Integration module: chassis cell lines generation
 - 2.2.2 Assembly module: Mammalian Modular Cloning (mMoClo)
 - 2.3 Test of the platform**
 - 2.3.1 Site-specific integration into the landing pad
 - 2.3.2 Expression homogeneity of multi-genic integrated constructs
 - 2.4 Experimental proof of concepts**
 - 2.4.1 Integration of a 7-gene regulatory network encoding an inducible memory device
 - 2.4.2 Integration of a library of 18 circuits
 - 2.5 Conclusion and discussion**
 - 2.6 Materials and methods**
-

2. 1. Introduction

Programming mammalian cells with large synthetic gene networks is expected to play a central role in helping elucidate complex regulatory cellular mechanisms ^{13,44-46}, implementing new useful biological functions ^{14,47,48} and accelerating the design of novel tailor-made therapeutic treatments ^{17,49-54}. However, our limited ability to precisely engineer and predict the behavior of these genetic programs in mammalian cells remains a major challenge ⁴⁹. Towards systematic and rational engineering of mammalian cells, new tools and methods are required that enable rapid prototyping and validation of genetic circuits in a standardized manner.

Stable chromosomal integration of genetic payloads can help achieve long-term expression of transgenes. Given the pleiotropic effect of the integration locus on transgene expression, it is critical to be able to study and compare the function of the integrated genetic components, genes or networks in the same genomic context ²⁷. Gene transfer methods such as retroviruses, lentiviruses and transposons are therefore not well suited because they result in random integration and the copy number of the integrated payload is not controlled well. Moreover, such techniques often limit the size of the payload to a few kilobases and do not tolerate the presence of repetitive sequences, which is often essential for genetic circuits comprising multiple transcription units. Several approaches have been developed that focus on targeted integration of foreign DNA into a transcriptionally active locus. Recent engineering of Meganucleases, Zinc Finger Nucleases (ZFN), TALENs and CRISP/Cas9 systems enable efficient integration of small DNA fragments at the locus of choice in mammalian chromosomes ⁵⁵⁻⁵⁸. However, such strategies involve double-strand break repair by homologous recombination or non-homologous end joining, which can lead to frequent head to tail concatemer integrations ²⁷, partial integration of the DNA fragments (**Supplementary Fig. 1-1**) or sequence alteration close to the target site ⁵⁹ and are therefore not well suited for single copy integration of large multi-gene payloads. Moreover, time-consuming clonal expansion and insert verification are almost always required due to the high frequency of off-target and multi-copy integrations ³⁰. Alternatively, precise integration of intact constructs can be achieved using site-specific recombination technologies ⁶⁰⁻⁶³, although the use of these techniques for integration of genetic networks in mammalian cells has not been demonstrated yet.

To address these challenges, we developed a comprehensive framework for simple and efficient generation of engineered cell lines that stably express multi-component genetic networks in the same chromosomal context (Fig.1). Our strategy consists of three main

components: (a) engineering of monoclonal chassis (landing pad) cell lines, (b) fast and modular assembly of large synthetic circuits and (c) targeted integration of the assembled circuits into the landing pad of the chassis cell lines with an efficient Bxb1 site-specific recombinase. Once the chassis cell line is generated, our method allows us to proceed from genetic parts (genes, promoters of choice) to functional assays of assembled and integrated circuits in mammalian cells in as little as 20 days. We demonstrate that the unique combination of very high integration efficiency, specificity and integrity (intact, functional payload) provided with our method enables rapid generation of nearly isogenic polyclonal cell populations characterized by highly homogenous and correlated transgene expression. We show scalability of the approach by construction, targeted chromosomal integration and functional validation of the largest mammalian genetic circuit that has been integrated to date (42 parts, 6 transcription units, 27kb). Finally, we performed a one-flask integration of a multi-gene payload library to demonstrate the suitability of our method for high-throughput screening of circuit libraries. As a whole, our method paves the way for robust and fast prototyping of synthetic genetic networks and provides a modular platform to streamline the generation of engineered isogenic stable cell lines, useful for a broad research community.

2. 2. Design of the platform

2. 2. 1. Integration module: Chassis cell lines generation

To generate the chassis cell lines (**Figure 2-1**), we created a landing pad vector for integration into the chromosomal locus of choice using engineered zinc finger nucleases.

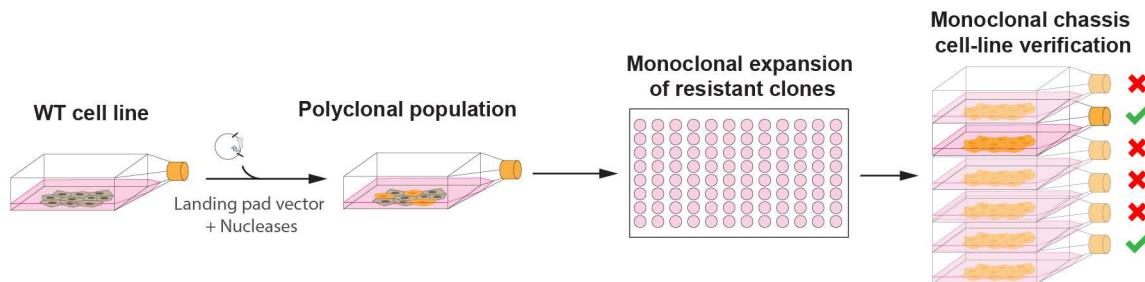


Figure 2-1 Generation of a chassis cell-line. A landing pad is integrated into the locus of a WT cell-line of choice using a zinc-finger nuclease, TALEN or CRISPR-Cas system. Integration events are then selected with Hygromycin and resistant clones are expanded. Finally, Southern-Blots and PCRs are performed on different clones to verify the insertion locus and integrity of the landing pad.

The landing pad contains a constitutive promoter driving co-expression of a fluorescent protein (EYFP) and a selection marker (Hygromycin) (**Figure 2-2A**). Between the promoter and the coding sequence we inserted an attP BxB1 phage attachment site ⁶⁴ to enable site-specific recombination in the landing pad. Co-transfection in the chassis cell line of a vector carrying the corresponding attB BxB1 attachment site with a vector expressing the BxB1 recombinase results in attB/attP site-specific recombination and insertion of the complete vector in the landing pad. To limit interference with surrounding host genes, we placed insulator sequences on both sides of the transcription unit ³¹. We integrated the landing pad into the AAVS1 locus ^{30,65} of one simian (Cos) and five human model cell lines (HEK293FT, HEK293, Hela, human Embryonic Stem Cells and human Induced Pluripotent Stem Cells). This locus was chosen because it promotes sustained expression of transgenes ⁶⁶. Similarly, we targeted the Rosa26 locus of a mouse Embryonic Stem cell line, as it also supports robust gene expression ⁵⁸. In this study we focus primarily on characterization and testing of the HEK293FT chassis cell line. After selection and expansion of clonal populations resistant to hygromycin, we isolated a HEK293FT-Landing Pad (LP) monoclonal cell line that showed correct mono-allelic integration of the landing pad, as confirmed with Southern blot and PCRs (**Supplementary Fig. 1-2 & 1-3**). EYFP expression in the confirmed chassis cell lines was sustained at high levels for over 40 passages (maximal duration of culture) without further antibiotic selection. While time-consuming (5 to 6 weeks), monoclonal selection, expansion and verification of the landing pad cell line is required only

once with our method, as the cell line can be stored and reused for all subsequent BxB1-mediated integrations (**Figure 2-2B**).

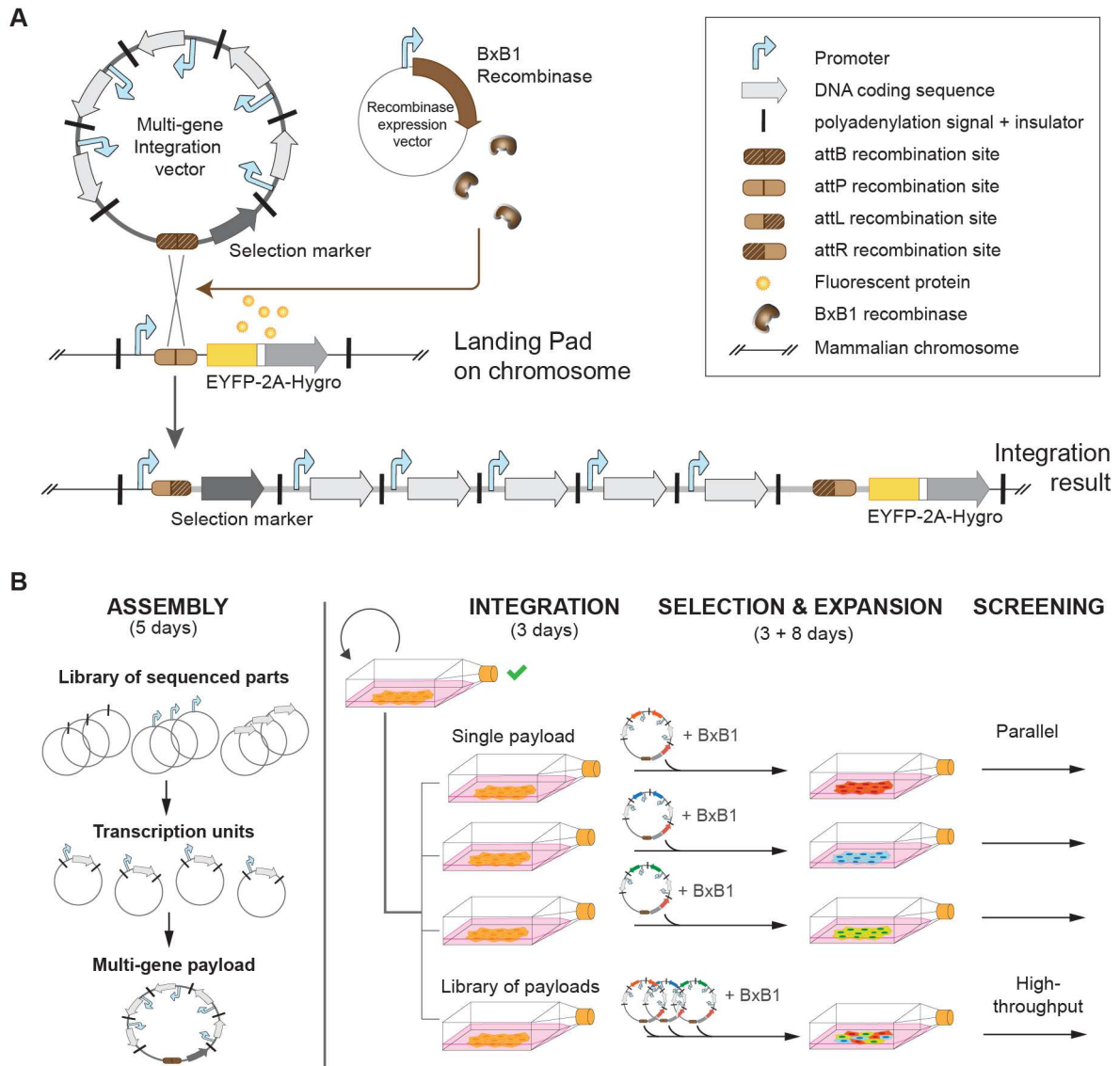


Figure 2-2. Mutligene circuit integration, assembly and screening

(A) Multi-gene circuit integration details. Co-transfection of the integration vector and the recombinase expression plasmid results in site-specific recombination between the BxB1 attB site (from the integration vector) and the attP site (pre-integrated in the mammalian chromosome within the landing pad). As a result, expression of the promoterless selection marker from the integration plasmid is triggered since it is inserted after the constitutive promoter of the landing pad. (B) Second phase: Multi-gene circuit assembly and integration in the landing pad. The monoclonal chassis cell line can be used to integrate multi-gene circuits constructed with the mMoclo assembly method, which requires only 2 cloning steps from sequenced-verified basic genetic parts to complete multi-gene circuits. Depending on the application, either a single circuit or a library of circuits can be co-transfected with BxB1 recombinase expression plasmid. Cell populations are ready to be screened in less than 2 weeks after transfection.

2. 2. 2. Assembly module: Mammalian Modular Cloning (mMoClo)

To create a framework for modular multi-gene circuits construction and integration into the landing pad of our chassis cell-lines, we extended the previously developed Modular Cloning strategy ⁶⁷ (**Figure 2-3 & Supplementary Fig. 1-4**). This highly efficient assembly method (95-100% correctly assembled constructs) ⁶⁸ relies on flanking the different circuit basic elements with Type IIs restriction enzyme sites that create unique 4bp-overhangs in order to obtain predetermined positioning of the parts within the overall circuit. The mammalian Modular Cloning (mMoClo) we introduce here includes 6 parts positioning vectors, 9 transcription unit-positioning vectors, 9 linker vectors and 1 destination vector (**Figure 2-3A**).

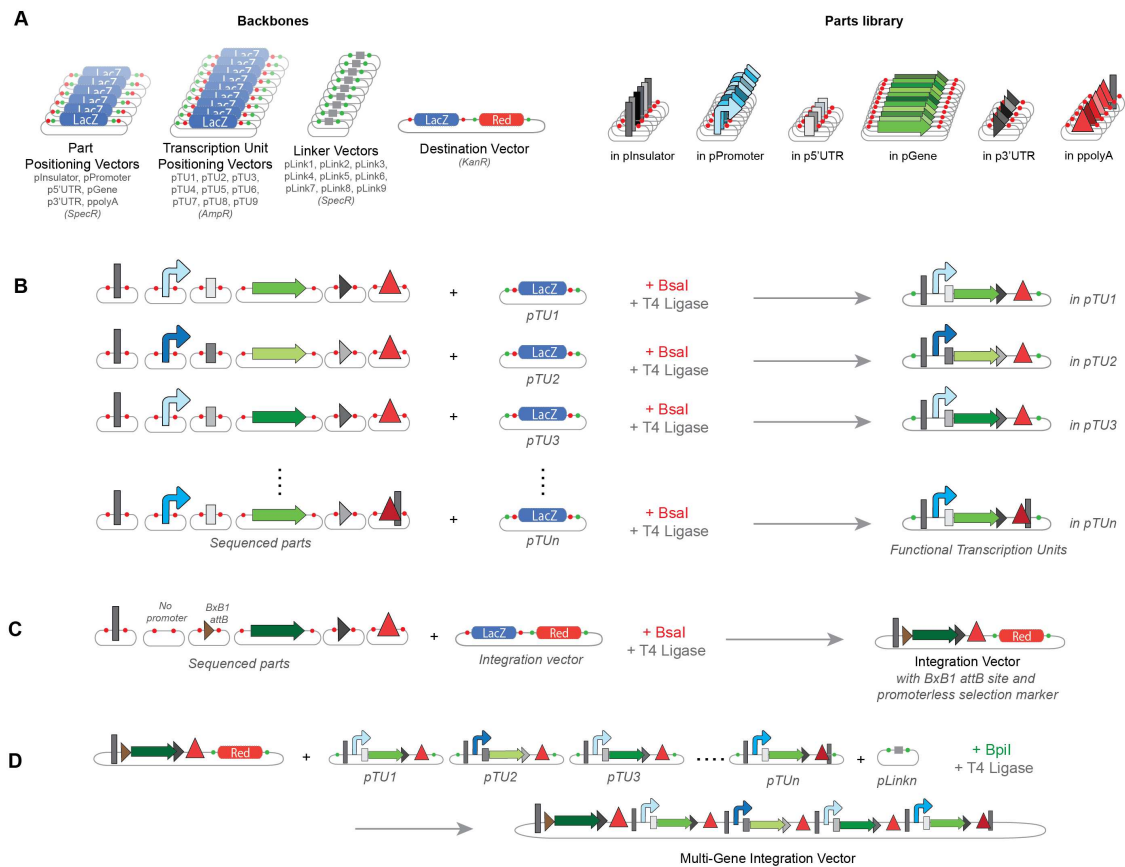


Figure 2-3. Overview of mammalian circuit assembly (mMoClo).

(A) mMoClo is based on a library of empty backbones used to assemble parts (Parts Positioning Vectors), transcription units (Transcription Unit Positioning Vectors) and complex multi-unit circuits (Destination Vector) to be integrated into the landing pad. The Parts Library is made of sequence-verified basic genetic parts (Insulators, Promoters, 5'UTR, Genes, 3'UTR and polyA sequences). (B) One-step creation of an integration vector carrying BxB1 attB site and promoterless selection marker. (C) One-step assembly of different functional transcription units in specific position vectors. (D) One-step assembly of a multi-unit genetic circuit by combining the desired transcription units, the corresponding linker vector and the previously built integration vector carrying the promoter-trap cassette.

The parts positioning vectors are used to create a sequence verified library of all the basic components necessary to construct an insulated mammalian transcription unit: insulators (in pInsulator), promoters (in pPromoter), 5' UTR sequences (in p5'UTR), gene coding sequences (in pGene), 3'UTR sequences (in p3'UTR) and polyA signals (in ppolyA). These parts are used to assemble either (1) functional transcription units into the “transcription unit positioning vectors” (**Figure 2-3B**) or (2) a promoter trap cassette carrying attB BxB1 recombination site followed by a promoterless selection marker of choice into the destination vector, to create an integration vector (**Figure 2-3C**). A complex multi-unit circuit can then be assembled into this integration vector by combining transcription units with the corresponding linker vector chosen according to the number of transcription units of the circuit (**Figure 2-3D**). The vector backbone used for the assembly of large circuits carries a pBR322 origin of replication and can be used to amplify circuits up to 31kb. While transformed bacteria were grown at 37°C for circuits up to 20kb, we reduced the growth temperature to 30°C for circuits above this size in order to minimize deletions in the circuits.

2. 3. Test of the platform

2. 3. 1. Site-specific integration into the Landing Pad

To assess integration efficiency, we created an integration vector that contains mKate2 fluorescent reporter fused to a promoterless resistance marker such that integration events can be monitored easily with flow cytometry (**Figure 2-4A**). Constitutively expressed Cerulean fluorescent reporter gene was also included in the integration plasmid to determine the transfection efficiency. We co-transfected our HEK293FT-LP chassis cell line with the integration plasmid and a plasmid expressing BxB1 recombinase (wild-type or codon optimized versions), and monitored expression of the fluorescent reporters. Three days after transfection, 10% of the transfected cells expressed mKate2 (8% of the entire population) (**Figure 2-4B**).

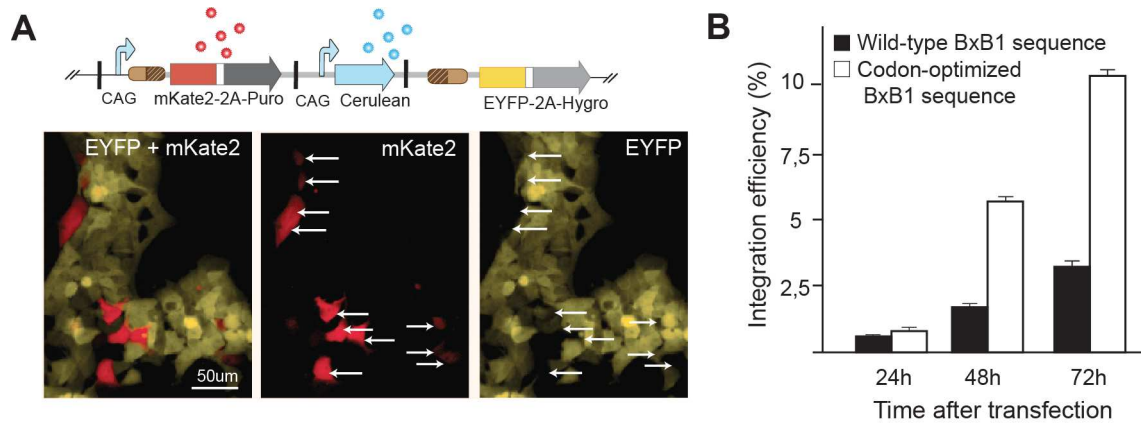


Figure 2-4. Integration efficiency

(A) Fluorescence microscopy images of HEK293FT Landing Pad (HEK293FT-LP) chassis cell line 3 days after co-transfection of the integration vector and the recombinase expression plasmid. Site-specific integration events can be monitored by expression of mKate2 and concomitant decrease in EYFP levels (integrated circuits shown on top). Cells that undergo the yellow to red switch, indicating successful integration, are denoted with arrows. (B) Integration efficiency assayed by flow cytometry defined as the percentage of cells expressing mKate2 among the population of transfected cells, without any selection.

BxB1 codon optimization yielded a 3-fold improvement in integration efficiency over wild-type BxB1 coding sequence (Figure 2-5).

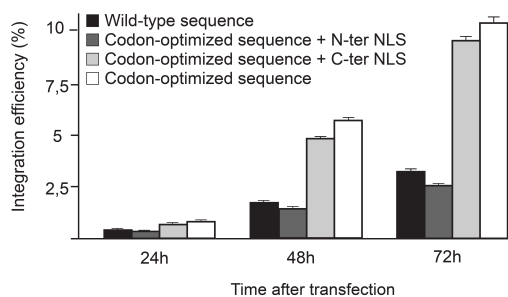


Figure 2-5. Effect of nucleotide sequence optimization and nuclear localization signal on the integration efficiency of BxB1 recombinase.

A 12kb integrative plasmid comprising a promoterless Puromycin-2A-mKate2 followed by constitutively expressed Cerulean was co-transfected with BxB1 expressing plasmid into the HEK293FT-LP chassis cell line. Integration efficiency was measured as percentage of mKate2 positive cells within the transfected cell population. The number of cells expressing mKate2 increased with time and reached a maximum on day 3-post transfection (we monitored the expression of mKate2 up to 10 days post-transfection, data not shown). NLS fusion to BxB1 did not improve the integration efficiency. When NLS was added at the N-terminus of the recombinase, significant negative impact on integration efficiency was observed. Our hypothesis is that in this case the NLS may sterically interfere with the N-terminal domain necessary for catalytic cleavage and ligation of DNA.

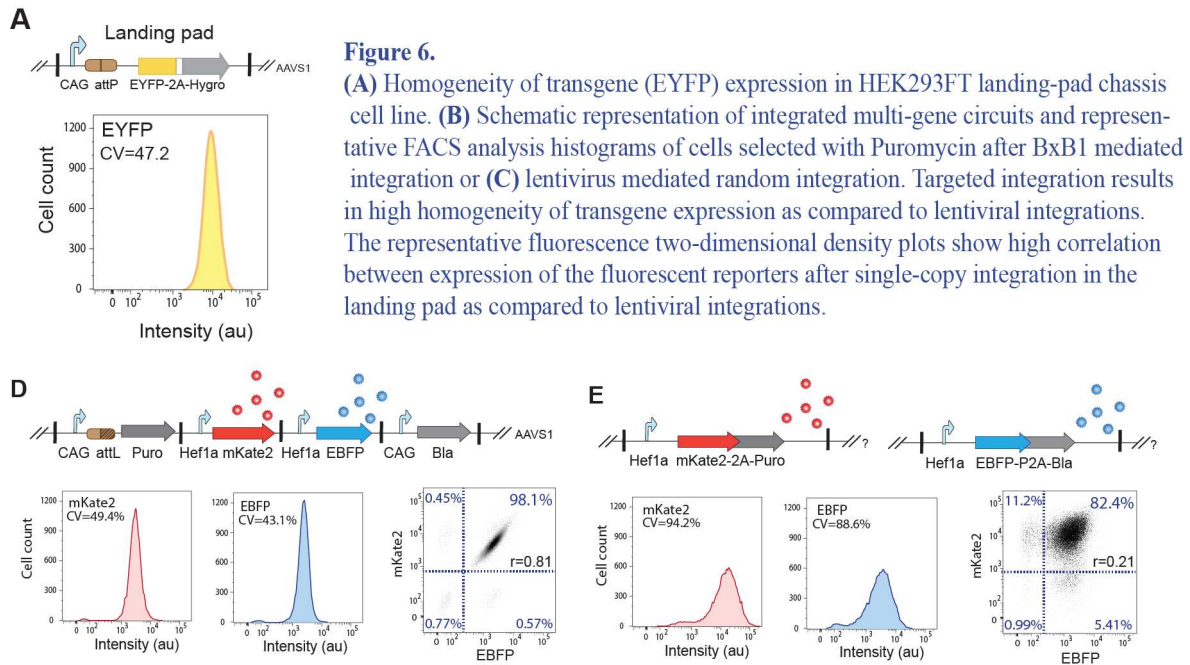
No integration events were observed when the integration plasmid was transfected without the BxB1 expressing plasmid. Addition of a Nuclear Localization Signal (NLS) to the recombinase did not increase the integration efficiency (Supplementary Fig. 1-5). We also tested different ratios

of the integration plasmid vs. the BxB1 expression plasmid to determine whether vector titration would influence the integration efficiency (**Supplementary Fig. 1-6**) and found that a 1:1 ratio resulted in the highest integration efficiency. After confirming the proper function of the platform in all our chassis cell lines (**Supplementary Fig. 1-7**), we assessed the integration efficiency in two of them (CHO and HeLa). We found that similar integration efficiency was obtained in these cell lines (**Supplementary Fig. 1-8**). To assess the integration specificity of the promoter-trap strategy, we monitored expression of mKate2 and EYFP among the resistant cells after puromycin selection. More than 99% of the resistant cells carried a single targeted integration of the full vector into the landing pad (mKate2 positive, EYFP negative) while less than 1% of resistant cells expressed both mKate2 and EYFP (0,23% for HEK293FT, 0,43% for HeLa and 0,86% for CHO, **Supplementary Fig. 1-9**), accounting for rare integration events into cryptic acceptor sites placed after an endogenous constitutive promoter. To validate these flow cytometry statistics and further demonstrate the platform's ability to generate nearly isogenic-engineered cell populations, we randomly isolated and expanded 42 resistant clones derived from the HEK293FT-LP cell line and analyzed their genomic DNA with Southern blots. Correct integration in the landing pad was confirmed for all 42 clones (**Supplementary Fig. 1-10**). We also sequenced the flanking sequences of the insert for these 42 isolated clones and confirmed that the integration process did not trigger any small insertions or deletions close to the recombination site (**Supplementary Fig. 1-11**). To assess the ability to transfer the platform to other cell lines, we co-transfected both the integration plasmid and the BxB1 expressing plasmid in all our chassis cell lines.

2. 3. 2. Expression homogeneity of multi-genic integrated constructs in a polyclonal cell population

The high integration efficiency and specificity of the recombinase-based method, combined with its ability to preserve intact payload, should result in homogenous expression of transgenes from a multi-genic construct even in polyclonal cell populations. To validate this hypothesis, we created a more complex integration vector containing a promoterless puromycin resistance marker followed by three transcription units constitutively expressing mKate2, EBFP2 and Blasticidin each using a separate promoter. To monitor fluorescence levels, we co-transfected this circuit together with the BxB1 expression vector, initiated a 5-day puromycin selection 3 days after the transfection and grew the surviving cells for 6 additional days without any selection before FACS analysis. As desired, we obtained highly homogenous expression of the two fluorescent proteins. We observed that the variability of transgene expression in the

polyclonal population after integration measured by the coefficient of variation (CV) (mKate2: =49.4%; EBFP: CV=43.1%) is similar to that of the monoclonal chassis cell line population (EYFP: CV=47.2%) (**Figure 2-6A**). We also observed that expression of transgenes placed on the same payload is highly correlated (coefficient of correlation $r=0.81$, **Figure 2-6B**).



Similar results were confirmed with other chassis cell-lines (**Supplementary Fig. 1-12**). In comparison, polyclonal populations generated by widely used viral-based integration strategies exhibited a stronger signal due to a high multiplicity of infection but larger variances and a markedly lower correlation in transgene expression (mKate2: CV=94.2%; EBFP: CV=88.6%; $r=0.21$) (**Figure 2-6C**). To generate these polyclonal populations, we infected the cells with two separate lentiviruses, one encoding the Hef1a_EBFP-P2A-Blasticidin transcription unit and the other encoding the Hef1a_mKate2-P2A-Puromycin. To select for integrations, we first applied puromycin for 5 days followed by Blasticidin for 7 days before FACS analysis.

2. 4. Experimental proof of concepts

2. 4. 1. Integration of a 7-gene regulatory network encoding an inducible memory device

To test the scalability of our method with respect to payload size and number of transgenes, we constructed and integrated a large gene circuit (42 basic parts, 6 transcription units, 27kb, **Supplementary Fig. 13a**) encoding an inducible memory device (**Figure 2-7A**). At the initial

state, prior to doxycycline induction, the circuit constitutively expresses the Cerulean fluorescent protein. Upon induction, expression of both EYFP and B3 site-specific recombinase (orthogonal to BxB1) are triggered, which results in precise genomic rearrangement within the circuit. Specifically, B3 recombinase recognizes the B3 recombination sites flanking the Cerulean coding sequence, which is then excised. The promoterless downstream mKate2 coding sequence is thus placed in frame with the constitutive promoter previously activating the Cerulean gene expression. The circuit is therefore locked into a final state in which the output has been switched from constitutive Cerulean expression to constitutive mKate2 expression. Following transfection, integration and selection with puromycin for 5 days, we monitored expression levels of the different fluorescent reporters before, during and after induction of genomic rearrangement with doxycycline (**Figure 2-7A**). At the end of the selection, after circuit transfection but before doxycycline induction, about 40% of the selected cells had already switched from Cerulean to mKate2 expression (**Supplementary Fig 13b**). As this ratio was stable over time and essentially no cells had an intermediate level of either Cerulean or mKate2, we hypothesized the B3 mediated switch had occurred right after transfection since leaky TRE promoter expression from the transfected circuits in the presence of high levels of constitutively expressed rtTA3 from the same plasmids was sufficient to trigger expression of B3 integrase at low levels. This would then result in moderate excision of the Cerulean expression cassette in the pool of plasmids, whether or not they had been already integrated in the landing pad.

To gain improved understanding, we performed a doxycycline induction experiment (**Figure 2-7B**) with cells harboring a non-rearranged payload only (mKate2-negative sorted cells). We monitored expression levels of the different fluorescent reporters before, during and after induction of genomic rearrangement (**Figure 2-7B**). Before induction with doxycycline, the circuit remained in its stable initial state with a highly homogenous expression of Cerulean in the population. Efficient transition between the device's states could be successfully monitored 3 days after induction of the circuit with doxycycline at the same time through the strong activation of EYFP and mKate2 expression, and the slow reduction in Cerulean levels. The B3-induced genetic rearrangement therefore locks the circuit into a new state with constitutive expression of mKate2. After withdrawal of doxycycline, mKate2 is still constitutively expressed while EYFP expression diminishes, as the induced B3-mediated genetic rearrangement is irreversible.

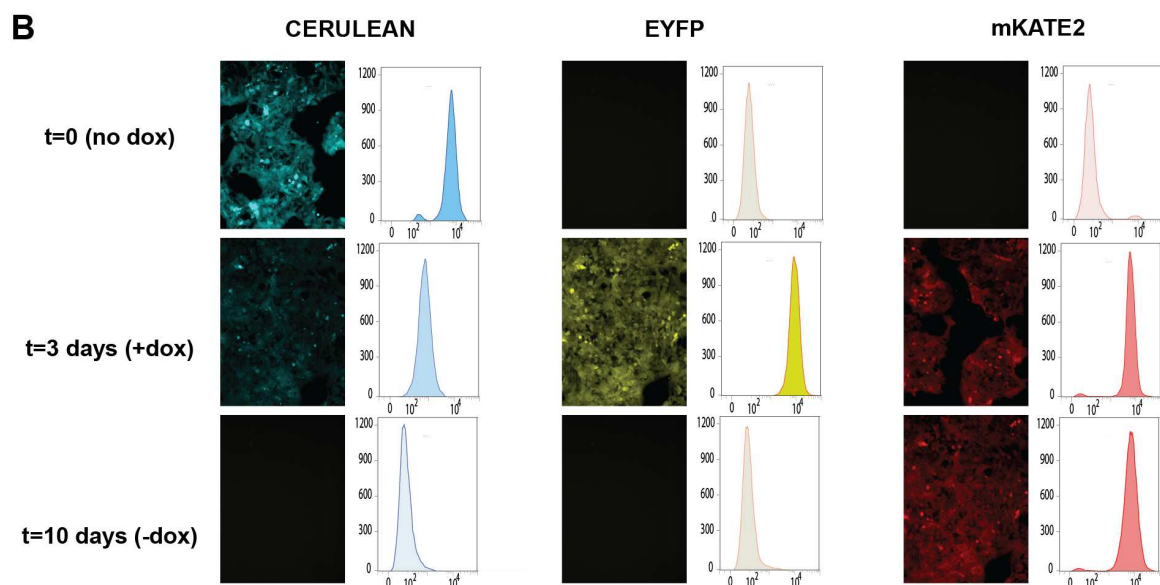
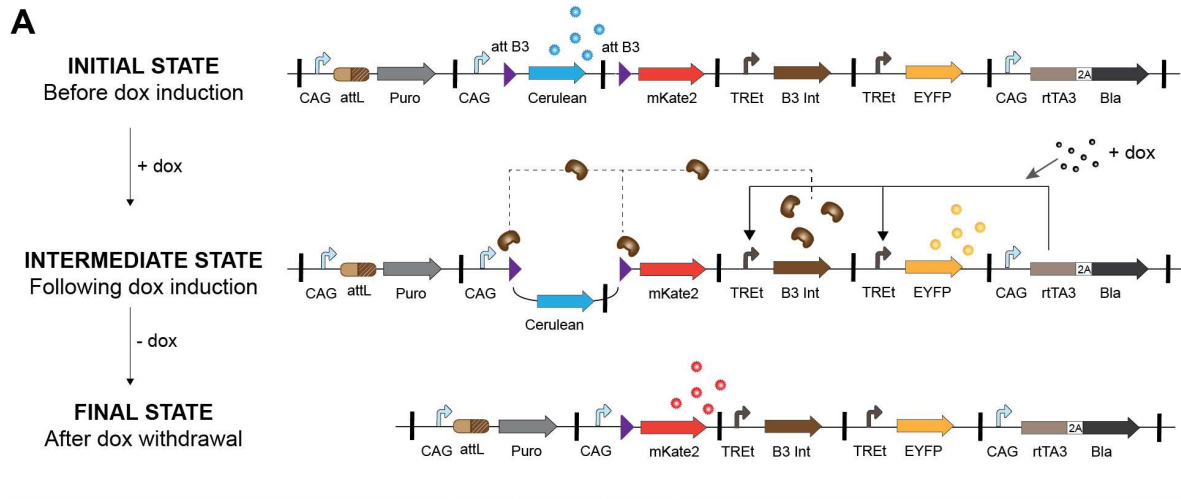


Figure 2-7. Transcriptional regulation and genomic rearrangement for a 7-gene circuit integrated into the landing pad of HEK293FT chassis cell line.

(A) Schematic representation of the integrated circuit. In its initial state, the circuit expresses Puromycin resistance gene, Cerulean, rtTA3 and Blasticidin resistance gene. After dox induction, rtTA3 activates TREt promoter and triggers expression of EYFP and B3 recombinase (Step 1). This leads to the intermediate state, during which B3 integrase mediates excision of the Cerulean cassette by recombination of the two attB3 sites. Subsequently, mKate2 gene is expressed. When dox is removed, EYFP is no longer expressed. In the final state, mKate2 is the only reporter that is expressed. (B) Fluorescent microscopy images and flow cytometry results confirmed the correct function of the circuit and shows the homogeneity of transgene expression in the polyclonal cell population (Intermediate state: 3 days after dox induction; Final state: 7 days after dox removal).

Flow cytometry analysis of the polyclonal population showed homogenous expression in the population for all reporters at the different time points (**Figure 2-7B**). These results suggest that our framework preserves the genetic integrity of such large payloads and provides not only homogenous expression of all transgenes without monoclonal expansion but also reliable dynamic behavior.

2. 4. 2. Integration of a library of 18 circuits

Finally we examined the suitability of our platform for high-throughput screening of multi-genic payload libraries. As a proof of concept, we performed three replicate co-transfections of the HEK293FT-LP chassis cell line with an equimolar mix of 18 different plasmids (each containing one distinct circuit) together with the BxB1 expression plasmid (**Figure 2-8A**). Each circuit encoded a fluorescent reporter fused to a nuclear localization signal and a different fluorescent reporter with variable cellular localization tags such that each circuit generates a distinct fluorescence phenotype (**Figure 2-8B** and **Supplementary Fig. 1-14**). 14 days after transfection, we evaluated the phenotype of 1500 cells from each of the 3 polyclonal populations and observed that all 18 phenotypes were represented well (**Figure 2-8E**).

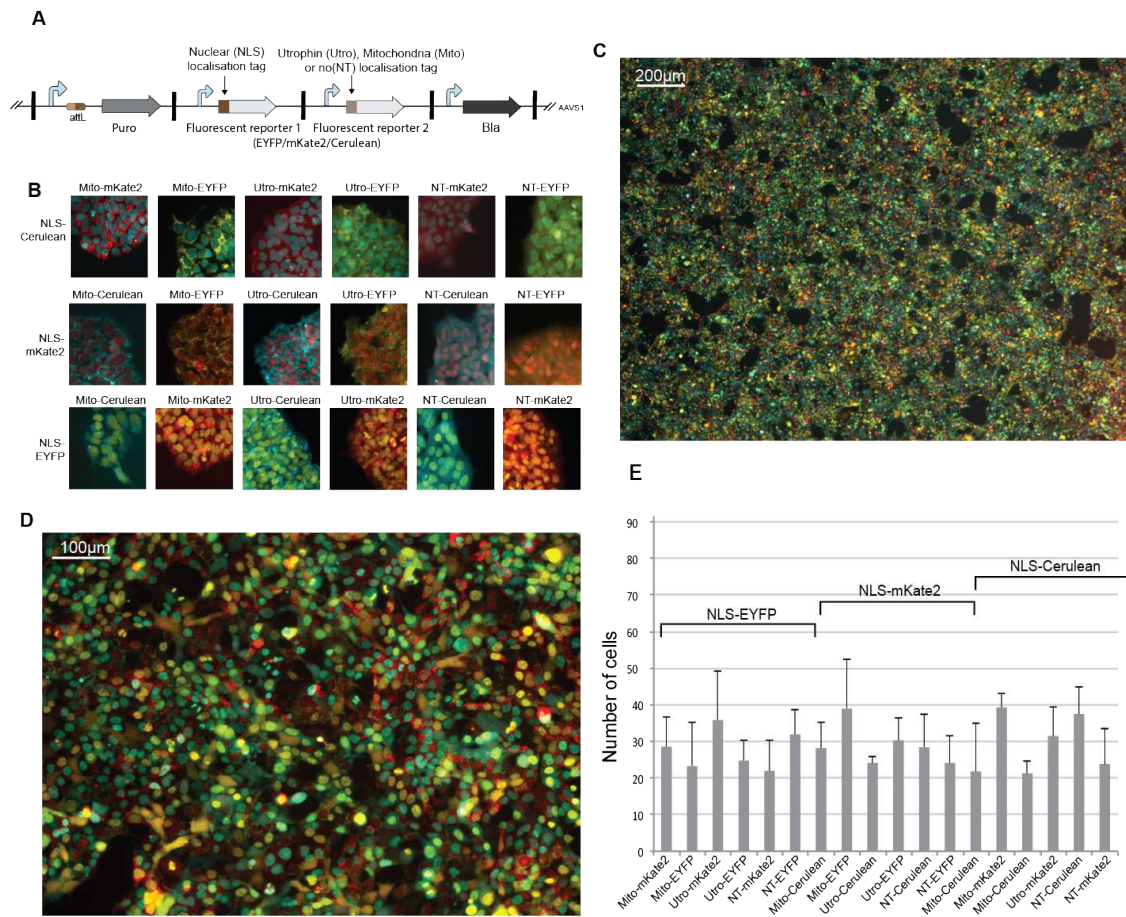


Figure 2-8. One pot integration of payload library.

(A) Architecture of the payload library. The first fluorescent reporter (EYFP, mKate2 or Cerulean) is fused to a nuclear localization tag. The second fluorescent reporter, different from the first one, is tagged either with an utrophin localization signal, a mitochondrial localization signal or with no localization signal. All possible combinations result in a library of 18 different payloads. (B) Fluorescent microscopy images of microcolonies 7 days after selection with puromycin (10 days after transfection of equimolar amount of the 18 plasmids and BxB1 expression plasmid in HEK293FTLP chassis cell line) and no cell passaging. All 18 phenotypes were observed and are shown. (C-D) Fluorescence microscopy images of polyclonal population 14 days after transfection, with 10x (C) and 40x (D) magnification, after cell passaging. (E) Average number of each cell phenotype appearance in 500 classified cells. Error bars represent the standard deviation from these classifications performed over 3 different fields of views from 3 independently repeated experiments (total of 4,500 classified cells).

2. 5. Conclusion and discussion

In this study, we introduce and validate a comprehensive framework for rapid prototyping of mammalian multi-gene synthetic networks. The framework combines a modular and rapid DNA assembly method to create the circuits together with a platform for stable integration of large synthetic gene networks in a predefined location in a mammalian chromosome. We demonstrate that our method provides an efficient and reliable strategy for correct and sustained execution of complex genetic programs in mammalian cells.

Engineering of cellular behavior remains a challenge and is still strongly dependent on the ability to screen many variants of a designed genetic program in a manner that allows direct comparison of the results. To accelerate the development of such libraries, we developed mMoClo, an extension of the MoClo assembly method, specifically tailored for construction of mammalian genetic circuits. mMoClo enables modular and combinatorial assembly of sequence-verified basic genetic parts into mammalian transcription units and then multi-unit gene circuits in only two cloning steps. Our extension confers an important advantage compared to other assembly methods previously developed for mammalian systems ⁶⁹⁻⁷¹. In particular, the design of the library reflects the structure of a functional mammalian transcription unit and allows combinatorial generation of a vast diversity of transcription units and circuits from basic sequence-verified genetic parts, from short 3' and 5' UTR sequences to large insulator sequences. We believe such a method could provide a new and broadly used standard for assembly of mammalian transcription units and complex circuits. To disseminate our platform, we deposited in Addgene a library of more than 120 parts and 20 assembly vectors (2 different insulators, 8 promoters, 20 5'UTRs, 90 Coding Sequences, 8 3'UTRs and 6 polyadenylation signals). We assembled circuits up to 31kb, although larger plasmids exhibited deletions or rearrangements during the bacterial cloning phase, likely due to the limited stability conferred by the pBR322 origin of replication for large plasmids. As an alternative more suitable for larger circuits, one could potentially replace the origin of replication of our destination vector with a BAC origin of replication, often used for the amplification of large circular DNA fragments ⁷².

To study and compare the different circuits we assembled with the mMoClo method, we created a stable integration platform, the landing pad, which can be inserted in specific loci of different cell types using engineered nucleases. While we used Zinc-Finger Nuclease in this study, other efficient systems such as CRISPR/Cas or TALEN could also be chosen for chromosomal insertion of the landing pad. Even though variation in DNA integration efficiency will differ according to both the nuclease used and the cell line, it is important to note that only a very limited number of candidate clones that survive selection are necessary in order to isolate and validate one functional chassis cell line. In this case, any of the available nucleases are therefore good

candidates to generate chassis cell-lines. The landing pad platform we created carries a BxB1 recombination site located downstream of a constitutive promoter and therefore acts as a synthetic promoter trap to allow for quick selection of targeted integration events. Such a strategy significantly reduces the survival likelihood of cells with circuit integration in cryptic pseudo att sites. Once a monoclonal landing pad chassis cell line is created, it can then be repeatedly used for single copy targeted integration of synthetic circuits mediated by our efficient codon-optimized BxB1 site-specific recombinase. While this is not the first study to use recombination sites for the creation of engineered cell lines, we demonstrate here a comprehensive platform based on the targeted insertion of a tailor-made landing pad in a wide range of cell lines, which can then be used to efficiently generate nearly isogenic cell populations carrying single and targeted copies of modularly assembled complex genetic circuits. This combination of mMoClo and our landing pad platform therefore becomes a convenient framework to significantly reduce the time from transfection of an integration vector to assay (from 2 months to 2 weeks) and to avoid the monoclonal expansion typically required when engineering cells with (multi)gene circuits.

We validated the approach by assembling the largest known functional mammalian synthetic circuit to date, integrating it into the landing pad of a HEK293FT monoclonal chassis cell line, and assessing its behavior within the polyclonal population of selected cells two weeks after transfection. Due to the unique combination of high integration efficiency, little to no off-target effect, and robust preservation of payload integrity, we obtained sustained and homogeneous expression of the genetic program. The desired genomic rearrangement was triggered with the same dynamics in the selected polyclonal cell population after induction of specific transgenes with doxycycline. For various applications, it would be important to integrate vectors without any bacterial sequences, as the presence of CpG nucleotides in these sequences has been shown to potentially lead to post-integrative silencing of the transgene⁷³. One possibility is to use the Minicircle⁷⁴ strategy to excise such sequences before transfection of the vectors, or use efficient excisionases such as Cre or Dre⁷⁵ after integration of the circuit.

To validate the suitability of our framework for the combinatorial assembly and screening of large circuit libraries, we also created a proof-of-concept library of 18 different circuits. We co-transfected all of them at the same time together with the BxB1 recombinase expression vector in our landing pad chassis cell line. Since there were no significant differences between the occurrences of the 18 different phenotypes within the polyclonal selected cell population, we expect that our framework could readily support the screening of a significantly higher number of circuits. In comparison to transient transfections and random stable transfections which lead to heterogeneous and often inconsistent gene expression, our approach enables the evaluation

of libraries of genetic constructs in the same genomic context, and therefore the side-by-side comparison of circuit behavior over an extended period of time. At the part level, our platform could be useful to test the activity of libraries of synthetic DNA promoters or enhancers in order to better understand the architecture of gene regulatory regions since gene activity is modulated by a complex interplay between these *cis*- and *trans*- acting DNA elements. At the circuit level, creating and stably integrating libraries of circuits in which the position and orientation of specific parts or transcription units are shuffled could be used to evaluate promoter interference. Preliminary results indeed indicate that part positioning within a circuit can influence transgene expression levels and overall dynamics. Libraries of therapeutically relevant circuits such as prosthetic networks¹⁷ or T cell proliferation controllers⁷⁶ could be rapidly evaluated in parallel to tune their parameters and optimize their sensitivity. In summary, our framework could accelerate the development of a well-characterized and standardized parts libraries for genetic engineering and also help create rules for the design of synthetic networks with better prediction accuracy, useful for a broad range of applications from systems and synthetic biology to biotechnology and medicine.

List of Supplementary Data (found in Annexes)

- Supp Figure 2-1.** Non-homogeneity of transgene expression after zinc-finger nuclease mediated integration of a 4-gene circuit.
- Supp Figure 2-2.** Landing pad chassis cell line construction: Southern blot genomic analysis of landing pad integration with zinc-finger nuclease in wild-type cells.
- Supp Figure 2-3.** Landing pad chassis cell line construction: PCR confirmations of mono-allelic landing pad integration.
- Supp Figure 2-4.** Details of the mMoClo backbones used in the assembly workflow depicted in Figure 2.
- Supp Figure 2-5.** Effect of nucleotide sequence optimization and nuclear localization signal on the integration efficiency of BxB1 recombinase.
- Supp Figure 2-6.** Influence of vector ratios (BxB1 expression vector vs. integration vector) on the integration efficiency.
- Supp Figure 2-7.** Microscopy images of different model chassis cell lines after integration of a simple circuit.
- Supp Figure 2-8.** Comparison of circuit integration efficiency in three different chassis cell lines.
- Supp Figure 2-9.** Flow cytometry analysis of chassis cell line fluorescence before and after integration of circuits and selection with puromycin.
- Supp Figure 2-10.** Specificity of integration into the landing pad: Southern blot genomic analysis assessing the specificity of BxB1 mediated integration.
- Supp Figure 2-11.** Sequence alignment of insert flanking sequences of the 42 isolated clones.
- Supp Figure 2-12.** Circuit performance in HeLa and CHO chassis cell-lines.
- Supp Figure 2-13.** Early genomic rearrangement of the 7-gene payload following transfection.
- Supp Figure 2-14.** Microscopy images of isolated colonies representing the 18 different phenotypes of the circuit library.
- Supp Figure 2-15.** Annotated maps of vectors used in this study.
- Supp Table 2-1.** Comparison of integration efficiencies of different recombinases.
- Supp Text 2-1.** DNA sequences of the probes used for the Southern Blots.
- Supp Text 2-2.** DNA sequences of the primers used for genomic PCRs.
- Supp Text 2-3.** DNA sequences of the genetic components from our mammalian parts library used in this study.

2. 6. Materials and methods

Golden Gate reactions. For all Golden Gate assembly reactions we used: 0.4 μ l of TypeIIIS enzyme (either BsaI from NEB, or BpiI from Fermentas), 0.2 μ l of T4 Ligase HC + 1 μ l of T4 Ligase HC buffer (Promega), 1 μ l of 10xBSA (NEB), 40 fmol for all vectors used in the reaction, ddH₂O up to a final total volume of 10 μ l. The thermocycler program used for all assemblies included: 1 step of 15 min at 37°C; then 50 cycles of [2 min at 37°C followed by 5 min at 16°C]; 1 step of 15 min at 37°C, 1 step of 5 min at 50°C and 1 final step of 5 min at 80°C.

Bacterial cell cultures. Liquid cultures of *E. coli* MG1655 were grown in LB Medium (Difco) at 37°C for plasmids up to 20 kb, and at 30°C for larger plasmids. When appropriate, antibiotics were added as follows: spectinomycin (100 μ g/mL), ampicillin (100 μ g/mL) and kanamycin (25 μ g/mL). For blue/white screening, we used X-gal at a final concentration of 40 μ g/mL.

Cell cultures, transfections and nucleofections. HEK293FT and HEK293 cell lines were purchased from Invitrogen. HeLa (CCL.2), CHO, COS, hESC cell lines were obtained from ATCC. HEK293FT, HEK293, HeLa and CHO cells were maintained in Dulbecco's modified Eagle medium (DMEM, Cellgro) supplemented with 10% FBS (PAA), 0.045 g/mL penicillin/streptomycin and non-essential amino acids (HyClone) at 37 °C, 100% humidity and 5% CO₂. mESC were grown in DMEM supplemented with 15% FCS, 10 ng/mL LIF/ESGRO (Millipore), 0.1 mM MEM non-essential amino acids, 100 U/mL penicillin, 100 μ g/mL streptomycin and 100 μ M Mercaptoethanol. COS cells were grown in DMEM supplemented with 10% FCS and 0.045 g/mL penicillin/streptomycin. hESC (CHB8, gift of George Daley, Harvard Medical School) were grown on Matrigel-coated plates in mTeSR-1 (Stem Cell Technologies, Vancouver CA). HEK293FT and HEK293 transfections were carried out in 24-well plates using Attractene reagent (Qiagen), 200000 cells and 200-300 ng total DNA per well. Media were changed 24 hours after transfection. mESC were transfected using Metafectene Pro (Biontex, Germany). Transfection was performed in 6 well plates using 800,000 cells and 2 μ g of DNA. COS cells were transfected using Metafectene Pro. Transfection was performed in 6 well plates using 600,000 cells and 2 μ g of DNA. For HeLa, CHO, hESC we used the 4D Nucleofector (Lonza, Switzerland) to electroporate the vectors. For hESC we used 600000 cells, 800 ng total of DNA and the nucleofection program CA-137 (Buffer P2). For CHO we used 600000 cells, 600 ng total of DNA and the nucleofection program DT-133 (Buffer SF). For HeLa we used 800000 cells, 800 ng total of DNA and the nucleofection program CN-114 (Buffer SE).

Landing Pad integration using zinc finger nucleases. To create the pLanding_Pad vector, an 800bp sequence homologous to the AAVS1 sequence on the left of the ZFN cut site, was cloned into the p_TU1 position vector. Similarly, an 800bp sequence homologous to the AAVS1 sequence on the right of the ZFN cut site, was cloned into the p_TU3 position vector. The following transcription unit was assembled into the p_TU2 position vector with a golden gate reaction: double cHS4 core insulator from a p_Insulator, CAG promoter from a p_Promoter, attP BxB1 from a p_5'UTR, EYFP-2A-Hygromycin from a p_Gene, inert 3' UTR from a p_3'UTR and rb glob polyadenylation signal + double cHS4 core insulator from a p_polyA. The three verified position vectors were then assembled altogether into the Shuttle Vector, deleting the crt red operon cassette. To create the ZFN expressing vector, pLV_CAG_CN-2A-CN, the cDNA encoding the two zinc finger nucleases for the AAVS1 locus, separated by a 2A tag was synthesized (GeneArt, Regensburg, Germany) and PCR-amplified with Gateway attB1/attB2 tags. Upon gel-extraction, the PCR-product was recombined into a pENTR_L1_L2 vector using BP clonase (Life Technologies, Carlsbad CA), yielding the pENTR_L1_CN-2A-CN_L2 vector. In a next step, pENTR_L1_CN-2A-CN_L2 and pENTR_L4_CAG_R1 were recombined into pLV_R4R2_GTW using LR clonase II plus (Life Technologies, Carlsbad CA), resulting in pLV_CAG_CN-2A-CN. To integrate the landing pad into the AAVS1 locus, we co-transfected cells with equimolar amount of the ZFN vector and the pLanding_Pad vector. 72 hours post transfection cell culture medium was supplemented with 200 µg/mL Hygromycin B (Invivogen) and the selection was maintained over a period of 2 weeks. Clonal cell lines were generated by serial dilutions of the surviving population.

Southern blots. Genomic DNA was extracted with the Quick-gDNA MidiPrep kit (ZYMO RESEARCH). 4µg of gDNA was digested over night with the EcoNI and XbaI (Landing Pad integration test) or with AseI (circuit integration test), separated on 0.8% agarose gel, transferred to a nylon membrane and probed with the indicated 32P-radiolabeled probe (Supplementary Text 1). Biomax MS film (KODAK) were stored two days in an exposition cassette with the membrane and revealed with a darkroom X-Ray processor (Velopex MD2000).

BxB1-mediated integration of circuits. To integrate circuits into the landing pad of HEK293FT chassis cell-line, we co-transfected 150 ng of the appropriate multi-gene integration vector with 150 ng of BxB1 integrase expression vector using attractene (Qiagen) in 24-well format. 72 hours post transfection cells were transferred to 6-well plates and culture medium supplemented with 1µg/mL puromycin (InvivoGen). Unless otherwise noted, selection was maintained for 5 days. Cells were typically assayed 14 days post transfection. For induction of the TRE promoter, 2 µg/mL doxycycline was used.

Integration efficiency test. Cells were collected by trypsinisation from 24 well plates 24h, 48h and 72h after transfection. Cells were then pelleted (1600 g; 3 min) and resuspended in 300 µl of PSB.

Lentiviral particle production and infection. We used Gateway ⁷⁷ cloning to construct the integration vectors. Our lentiviral Gateway destination vectors contain pFUGW ²⁵ (Addgene plasmid 14883) backbone and Gateway cassette (comprising chloramphenicol resistance and ccdB genes flanked by attR4 and attR2 recombination sites) followed by blasticidin or puromycin resistance markers expressed constitutively. LR reaction of the destination vectors with entry vectors carrying human elongation factor 1 alpha (hEF1a) promoter and either mKate2 or EBFP2 fluorescent proteins was used to create the following expression vectors: pLV-hEF1a-mKate2-P2A-Puromycin and pLV-hEF1a-EBFP2-P2A-Blasticidin. For production of lentiviral particles ~ 2×10⁶ HEK293FT cells (Invitrogen) in 3 mL of DMEM complete media were plated into gelatin-coated 60 mm dishes (Corning Incorporated). Three hours later the ~80% confluent cells were co-transfected with 0.5 mg of the pLV-hEF1a-mKate2-P2A-Puromycin expression vector, 1.1 mg packaging plasmid pCMV-dR8.2 (Addgene plasmid 8455) and 0.55 mg envelope plasmid pCMV-VSV-G (Addgene plasmid 8454) ⁷⁸ using Attractene reagent (Qiagen) and following manufacturer's protocol. Media containing viral particles produced from transfected HEK293FT cells were harvested ~ 48 h post-transfection and filtered through a 0.45 mm syringe filter. 1.5 mL of the filtrate was added to ~ 20% confluent HEK293FT cells in 12-well plate seeded immediately before infection. After 48 h, media were changed and supplemented with 1 µg/mL puromycin (InvivoGen). Cells were maintained under selection for 5 days. After selection and expansion, cells were infected again with lentiviral particles produced using pLV-hEF1a-EBFP2-P2A-Blasticidin expression vector, following the same protocol. Blasticidin selection (10 µg/mL) (InvivoGen) was applied for 7 days. Cells were subsequently analyzed by flow cytometry.

Flow cytometry and data analysis. Cells were analyzed with LSRFortessa flow cytometer, equipped with 405, 488 and 561 nm lasers (BD Biosciences). We collected 30000–50000 events, using a forward scatter threshold of 5000. Fluorescence data were acquired with the following cytometer settings: 488 nm laser and 530/30 nm bandpass filter for EYFP, 561-nm laser and 610/20 nm filter for mKate2, 405 laser and 525/50 filter for AmCyan and 405 nm laser, 450/50 filter for Pacific Blue. Data analysis was performed with FACSDiva software (BD Biosciences) and FlowJo (<http://www.flowjo.com/>). For histogram analysis, flow cytometry data in .FCS format were exported into text format using FCS Extract 1.02 software (E.F. Glynn, Stowers Institute for Medical Research) and analyzed in Microsoft Excel. We used bi-exponential scales

for visualizing FACS data. For cell sorting, cells were collected directly into an 8-well micro-slide (Ibidi) by a FACSAria cell sorter.

Microscope measurements and image processing. Fluorescence microscopy images of live cells were taken in glass-bottom dishes or 12-well plates using Zeiss Axiovert 200 microscope and Plan-Neofluar 10x/0.30 Ph1 objective. The imaging settings for the fluorophores were S430/25x (excitation) and S470/30m (emission) filters for AmCyan, and S565/25x (excitation) and S650/70m (emission) for mKate2. Data collection and processing were performed using AxioVision software (Zeiss). For the circuit library experiment, we evaluated 500 cells from 3 different fields of views for each replicate transfection. We manually marked all cells with a specific tag corresponding to its observed phenotype and used an Adobe Illustrator automated script to sum the number of cell instances for each cell type and each field of view.

Chapter 3 Expanding the site-specific recombinases toolbox for mammalian synthetic biology

Contents

- 3.1 Introduction**
 - 3.2 Discovery of new mycobacteriophage large serine recombinases**
 - 3.2.1 Identification of large serine recombinases
 - 3.2.2 Validation of the putative recombinase activity in *M. smegmatis*
 - 3.2.2.1 *Creation of integrative vector for Mycobacterium species*
 - 3.2.2.2 *Transformation of M. smegmatis with the integrative vectors*
 - 3.3 Discovery of new recombination sites**
 - 2.3.1 Identification of the integration sites in *M. smegmatis* genome
 - 2.3.2 Validation of functional attB/attP recombination sites
 - 3.4 Validation of recombinases activity in mammalian cells**
 - 2.4.1 Assessment of intermolecular recombination
 - 2.4.2 Assessment of recombination crosstalk
 - 2.4.3 Creation of a library of orthogonal sites
 - 3.5 Conclusion and discussion**
 - 3.6 Materials and methods**
-

3. 1. Introduction

Site-specific recombination systems mediate recombination reactions between two specific DNA sequences termed recombination sites. While most site-specific recombination systems occur in bacteria and their viruses, many of the characterized site-specific recombinases function in heterologous systems, which can be extremely useful to manipulate or engineer their genome via integration, excision or inversion events as demonstrated in the previous chapter and in many other studies ⁷⁹⁻⁸⁵. To streamline the engineering of mammalian cells, one would benefit from having a programmable recombinase that could be directly targeted to the endogenous sequence of choice in the mammalian genome. This would eliminate the requirement to pre-integrate a natural recombination site (cf Chapter 2). To be useful, the integration reaction would still have to be highly specific and efficient.

Based on amino acid sequence homology, site-specific recombinases fall into one of two mechanistically distinct families: the tyrosine and the serine recombinases^{86,87}. The names come from the conserved nucleophilic amino acid residue that they use to attack the DNA and which becomes covalently linked to it during strand exchange.

The better characterized members of the **tyrosine recombinase** family are the integrases from coliphage I and from prophage lambda ^{88,89} which are used either to integrate or excise the phage genome from their bacterial host. The integrase recognize a specific attachment on the phage genome, attP, and its counterpart on the bacterial genome, attB, between which it catalyzes a single DNA crossover. The circular phage genome is therefore linearized upon integration and the hybrid attachment sites formed that flank it are called attL and attR (**Figure 3-1**). The integration reaction also requires an accessory protein, the Integration Host Factor (IHF), which is specific to bacteria and helps folding the DNA molecule in the appropriate way for recombination to happen ⁹⁰. When and only when an additional protein (Xis) is expressed together with Int and IHF, the excision reaction is triggered.

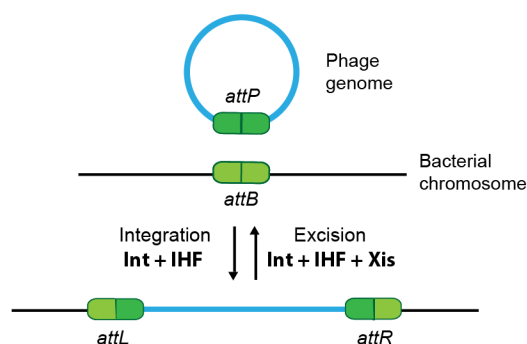


Figure 3-1 Lambda phage integration and excision

Integration of the circular phage genome into the bacterial genome is mediated by the Lambda phage integrase Int, which interacts with endogenous Integration Host Factor. Site-specific recombination between attP and attB results in the creation of attL and attR sites flanking the linearized phage genome within the bacterial genome. The reverse reaction happens when Int and Xis are expressed at the same time.

This system ensures that the integration events are directional and therefore not reversible in the absence of the Xis protein.

A major drawback of these systems that limits their use for mammalian genome engineering is the required intervention of the IHF for recombination to happen. IHF is indeed not naturally expressed in mammalian cells and even though scientists have engineered a mammalian version of IHF, the efficiency of both intra and intermolecular reactions remained extremely low in mammalian cell ^{75,79}.

The tyrosine recombinase family also includes many members such as Cre⁸⁶, Flp^{88,91} and Dre^{75,92} that catalyze non-directional and therefore reversible site-specific recombination reactions between two identical sites of approximately 35bp in length in the absence of accessory proteins. These recombination systems, in particular the widely used Cre-loxP and FLP-frt, demonstrated highly efficient deletion of DNA located between two directly repeated substrates inserted into the genomes of many higher eukaryotes^{93,94}. As the recombination sites are regenerated after Cre-loxP or FLP-frt recombination, efficient and stable integration of DNA is greatly compromised by the tendency of the integrated DNA to immediately excise, therefore favoring the deletion reaction (**Figure 3-2**). To overcome the reversibility of these systems, some recombination with a few base pair mutation sites have been created that prevent results in hybrid recombination sites after the integration events, not recognized anymore by the recombinase^{91,92}. Although these tricks improved the integration efficiency of DNA in the genomes of heterologous hosts, it stayed significantly lower than what has been achieved with serine recombinases.

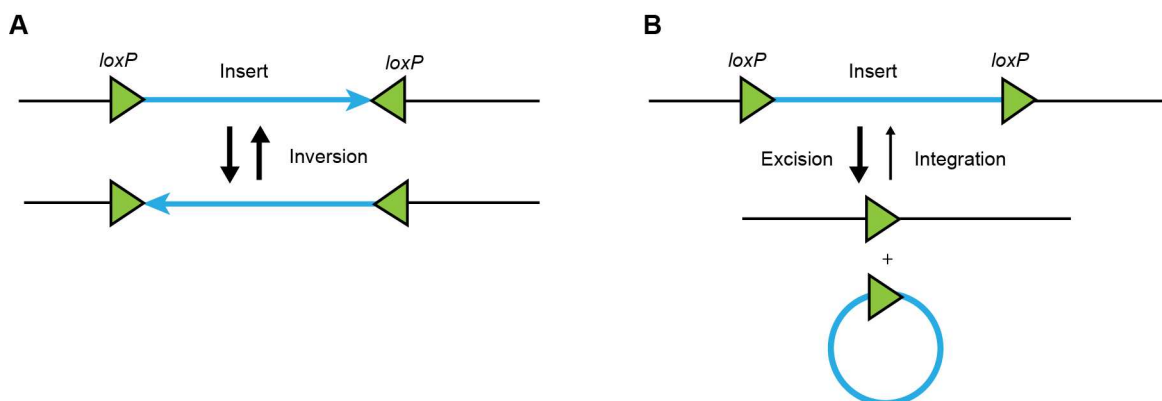


Figure 3-2. Recombination events mediated by Cre recombinase

(A) When both loxP sites are in opposite orientation on the same DNA strand, the sequence located between the sites is constantly inverted by the Cre recombinase. (B) When both loxP sites are in the same orientation, the Cre recombinase mediates the excision of the insert. Because the reaction is non-directional, the circular excised insert can be relinearized, though the intermolecular reaction is less favored compared to intramolecular recombination.

The **serine recombinase** family is composed of multiple structural groups represented by the resolvase/invertases, the large serine recombinases and the transposases (**Figure 3-3**)⁹⁵. They all mediate site-specific recombination reactions with strictly controlled directionality, in the absence of an accessory protein (Xis or recombination directionality factor, RDF), and are therefore of great interest for mammalian genome engineering^{92,95,96}

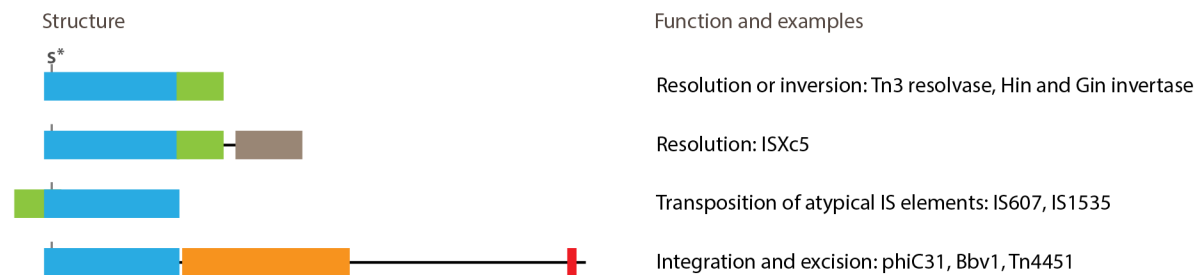


Figure 3-3 Overview of the structural and functional variation in the serine recombinases. This figure shows the catalytic domain and E-helix (blue), with S* showing the position of the serine nucleophile; the DNA-binding domain (green) containing a recognizable helix-turn-helix (H-T-H) motif; and conserved domains of unknown function found in subsets of recombinases (grey, orange, and red).

In their native contexts, serine resolvases and invertases selectively recombine target sites within the same DNA molecule. Although directional, they are poorly suited for accurate genomic recombination because the recognition motifs of their DNA binding domain (DBD) are short (4–6 bp) and degenerate. Recent studies have demonstrated that the DBD of a serine resolvase can be replaced with a zinc finger protein of higher affinity and specificity^{92,97}. This substitution retargets recombination to sequences flanked by zinc finger binding sites (ZFBS). However, these zinc finger-recombinases (RecZFs) retain a second, complementary specificity. The serine catalytic domain indeed imposes its own sequence requirements on the interior of the RecZF target site (20-bp core). Functional RecZF recombination sites must then contain sequences compatible with both the zinc finger DNA-binding protein and recombinase catalytic domain. Even though such hybrid recombinases can offer broader targeting capabilities, their low efficacy for DNA integration is still a bottleneck.

To date, the most efficient, directional and specific recombinases are the large serine recombinases, widely used for DNA integration in mammalian cells⁹⁶. The integrases from the *Streptomyces* temperate phage PhiC31^{97,98} or from the *Mycobacteriophage* BxB1^{98,99} were shown to mediate unidirectional and highly specific recombination between relatively small recombination substrates, attP and attB (< 50 bp). The domains that compose large serine

recombinases are however not as modular as with the resolvases. It is currently impossible to predict how to retarget these recombinases towards new recombination site sequences. To date, we have characterized a very limited number of highly efficient large serine recombinases. As they have been discovered in phage genomes, it is very probable that they are yet many other ones to discover given the large diversity of phages that exist on the planet.

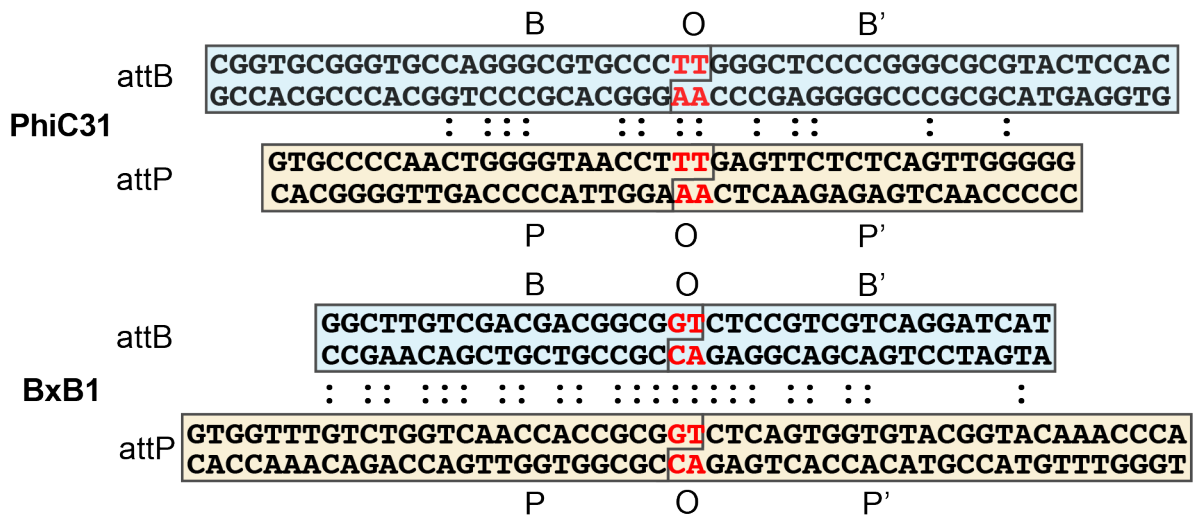


Figure 3-4 Sequences of PhiC31 and BxB1 attachment sites

For each recombinase, the attB and attP attachment sites share a common central dinucleotide sequence (in red) that is used for the strand exchange reaction during site-specific recombination. This central dinucleotide sequence ensures the directionality of the reaction. Homologies between the sites are shown with dots between the sequences.

The discovery characterization of new large serine recombinases would not only help us better understand the relation between their structure and their function and therefore give us more insight on how to reprogram them, but it would also expand the current toolbox scientists have at their disposal for the engineering of heterologous hosts genomes and for the manipulation of synthetic genetic circuits.

In this chapter, I will describe how I discovered new large serine recombinases from recently sequenced Mycobacteriophages and characterized their respective attachment sites. After having identified putative large serine recombinase from genome sequences, I validated their function in their natural host, *Mycobacterium smegmatis*, and used a plasmid rescue strategy to infer their respective attachment sites, which I then validated in *E. coli*. I then demonstrated that these systems could be used in mammalian cells and I created libraries of orthogonal site-pairs for each of these recombinases.

3. 2. Discovery of new mycobacteriophage large serine recombinases

3. 2. 1. Identification of large serine recombinases

Among the few large serine recombinases already identified, 3 of them have been discovered from Mycobacteriophage genomes: BxB1^{98,100}, Peaches and BxZ2^{99,101,102}. Up to now, BxB1 recombinase has been demonstrated to be the most efficient to integrate DNA into the genome of mammalian cells⁹⁶.

Peaches and BxZ2 have yet never been tested in eukaryotic cells. To extend the limited toolbox of large serine recombinases available for the manipulation of mammalian genomes, I decided to look for putative recombinase sequences among the very recently sequenced 400 new Mycobacteriophages genomes^{100,103}.

To do so, I created an algorithm that scans through all putative protein sequences from these genomes and extracts the ones that contains conserved motifs, typical of large serine recombinases. Compared to other serine recombinases, many features and motifs are specific to large serine recombinases (**Figure 3-3**). Not only all known large serine recombinases are constituted of more than 400 amino acids but they also carry very specific conserved residues in their catalytic and dimerization domains (**Figure 3-5**).

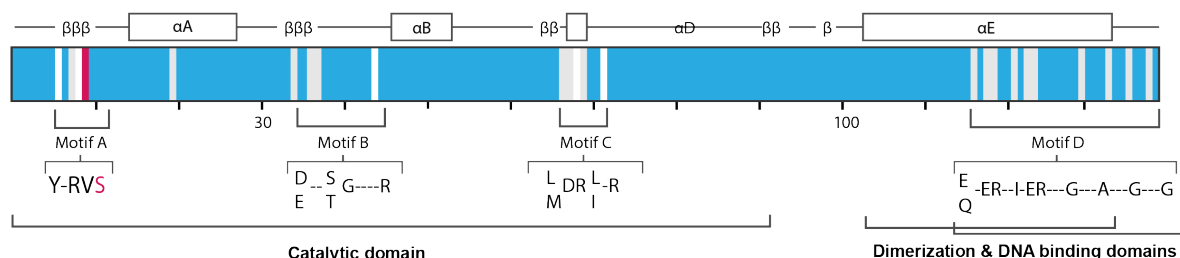


Figure 3-5 Conserved motifs within the catalytic domain and dimerization helix (αE) of large serine recombinases. Motifs A and C contain the critical active site residues of the recombinase. Motif D, contained within the C-terminal portion of the E-helix plus a few residues beyond, is mostly involved with binding the DNA in the region abutting the cleavage site. Motif B forms a rather mobile loop whose function remains mysterious despite the remarkable conservation of the Ser-39, Gly-40, and Arg-45 residues.

Based on rules taking into account these specificities, the algorithm identified that 26 genomes carried a large serine recombinases (out of the 400 genomes). The 374 other genomes carried a tyrosine recombinase. By comparing the amino-acid sequences of the 26 identified serine recombinases to the ones of BxB1, Peaches and BxZ2, two new clusters emerged, significantly distant from the Peaches and BxZ2 cluster and the BxB1 cluster (Figure 3-6). While

none of the 26 recombinases was grouped with Peaches and BxZ2, one third of them shared a high degree of homology with the well-characterized BxB1 recombinase: KSSJEB, PattyP, Doom, Scowl, Lockley, Switzer, Bob3, Troube, Abrogate, Anglerfish, Sarfire, SkiPole, ConceptII, Museum and Severus. The first new cluster contains both Rey and Bongo recombinases; and the second one contains Airmid, Benedict, Theia, Hinder, ICleared, Sheen, Mundrea, Veracruz and Rebeuca.

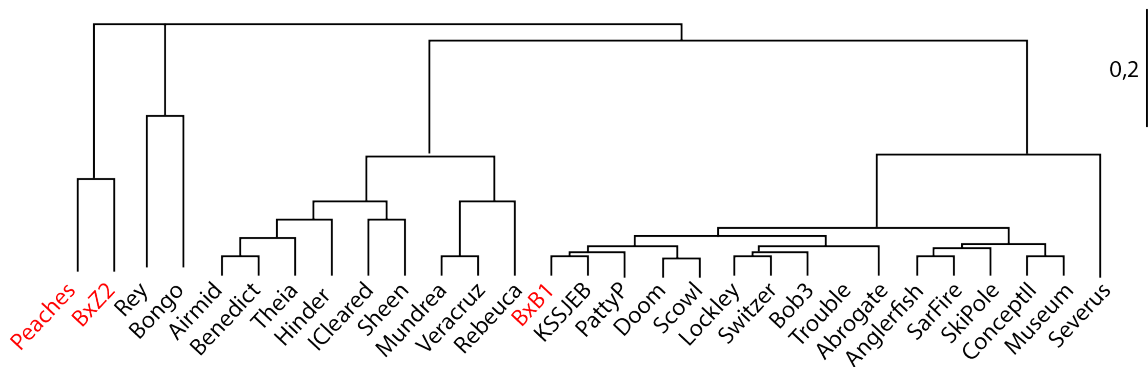


Figure 3-6 Phylogenetic tree of identified large serine recombinase sequences. The 26 recombinases together with Peaches and BxZ2. The scale bar indicates a difference of 2 amino acids

Such results were encouraging as different amino acid sequences likely results in different tertiary protein structure and therefore different DNA recognition and binding activities.

3. 2. 2. Validation of the putative recombinase activity in *M. smegmatis*

3. 2. 2. 1. Creation of integrative vectors for *Mycobacterium* species

When a mycobacteriophage infects *M. smegmatis*, its recombinase is expressed to mediate the integration of the phage genome into the *M. smegmatis* via site-specific recombination between the phage attachment site (attP) and the bacterial chromosomal attachment site (attB). Therefore, in order to validate the integration capacity of the 26 putative recombinases identified with the algorithm, I created for each of them a synthetic vector mimicking the mycobacteriophage genome and designed in a way such that integration events could be easily monitored

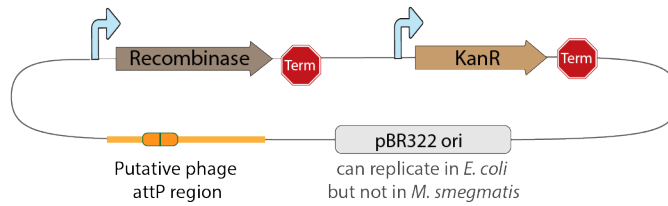


Figure 3-7).

This vector carried three essential components: an attP site, a gene cassette to constitutively express the recombinase and a selection marker to enable selection of integrants. While it is almost impossible to predict the exact sequence of the attP site, it is however possible to predict its approximate location on the phage genome. When looking into phage biology, it appears that the attP site is most often located close to the start codon of the recombinase coding sequence. This evolutionary strategy allows the recombinase expression to be under the control of bacterial promoter once the phage has integrated into its host genome. Given that I did not have the physical DNA of the 26 mycobacteriophages for which I had identified the presence of a large serine recombinase, I had to synthesize the DNA region that was potentially carrying the attP site (from 750bp before to 50bp after the start codon of the putative recombinase).

For the recombinase cassette, I used the Golden Gate assembly method to assemble the strong Hsp70 constitutive promoter together with a codon-optimized coding sequence of the recombinase followed by an efficient transcription terminator. Again, I had to synthesize the coding sequence of the recombinase given that I did not have the physical DNA of the corresponding phage. I decided to codon-optimize the coding sequence for an efficient expression in mammalian cells since my end goal was to test these recombinases in mammalian cells.

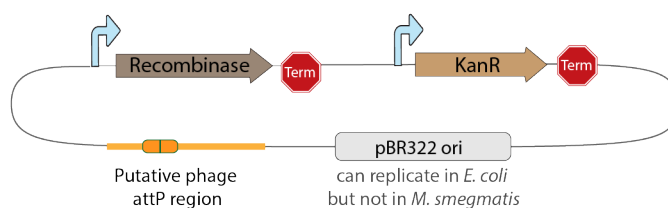


Figure 3-7 Details of the integrative vector architecture. The vector is composed of three main components: the recombinase, the resistant gene cassette and a sequence carrying the attP sequence. The pBR322 ori has been chosen so that the vector can replicate in *E. coli* but not in *M. smegmatis*.

To be able to screen for integration events of the synthetic vector into the chromosome of *M. smegmatis*, I added a selection marker constitutively expressed both in *E. coli* and *M. smegmatis*.

Finally, the origin of replication pBR322 was used to allow for efficient replication of the plasmid in *E. coli* for cloning purpose. However, pBR322 cannot be replicated in *M. smegmatis*.

Thanks to this specificity, transformation of *M. smegmatis* with the integrative vector would result in resistant cells only if the expressed recombinase would mediate site-specific recombination between the plasmidic attP and the chromosomal endogenous attB (Figure 3-8). This would indeed lead to the stable integration of the vector and therefore to the stable expression of the resistance marker. If the vector cannot self integrate into *M. smegmatis* genome (either because the recombinase is not functional or because the attP sequence is not present on the vector), then the non-replicative vector would be diluted among the population of cell.

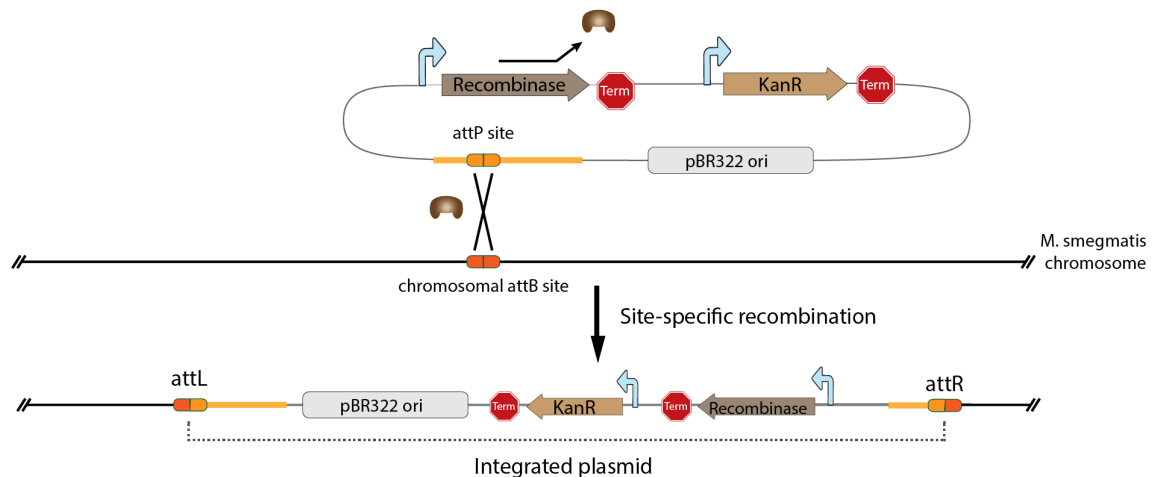


Figure 3-8 Integration of the integrative vector into *M. smegmatis* genome. After transformation of the integrative vector in *M. smegmatis*, the putative serine recombinase is expressed and will mediate site-specific recombination between the attP site located on the integrative vector and its counterpart attB site on the *M. smegmatis* genome. This will result in the linearization of the integrative vector and its stable integration into the genome of *M. smegmatis*, which will then become resistant to Kanamycin.

3. 2. 2. 2. Transformation of *Mycobacterium smegmatis* with the integrative vectors

Together with my collaborator Nadine Honoré from Pasteur Institute, we transformed each of the 27 integrative vectors (26 new recombinases + BxB1 as a positive control) in wild type *M. smegmatis* and plated the transformed bacteria on a selective media. Out of the 27 integrative vectors tested, 23 led to the growth of resistant clones, which we picked 4 days after transformation and expanded for further testing.

To confirm the stable integration of the vector into the chromosome of the resistant clones, we extracted their chromosomal DNA and performed a PCR with primers flanking the resistance cassette (Figure 3-9). The PCRs were positive for each of the 3 clones picked for the 23 functional

recombinases.

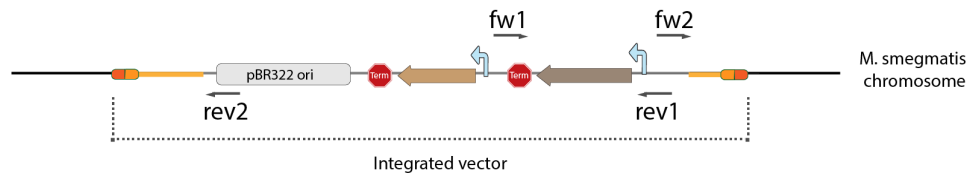


Figure 3-9 Representation of the primers used to check chromosomal integration of the vector. Fw1 and rev1 primer pair was used to confirm the presence of the Kanamycin resistance gene. Fw2 and rev2 primer pair was used to confirm that integration of the vector has been mediated by site-specific recombination with an attP site within the predicted attP region sequence carried by the vector.

To verify that stable integration into the chromosome was the result of the plasmid linearization via site-specific recombination from a sequence carried by the putative attP region on the plasmid, we tried to amplify this attP region with primers flanking the sequence carried by the original integration vector. While the amplification worked on all 23 integration vectors, it did not work when using the extracted chromosomal DNA as a template.

These first results were encouraging as they meant that both the identified recombinases were functional and that the attP sequences were indeed located in the phage genome region I predicted, somewhere close to the start codon of these recombinases.

Now that the integrative vectors had been integrated into M.smegmatis via site-specific recombination, I had to discover the exact location of their insertion.

Recombinases	Rey	Bongo	Airmid	Benedict	Theia	Hinder	ICleared	Sheen	Mundrea	Veracruz	Rebeuca	BxB1	K5SJE8	PattyP	Doom	Scowl	Lockley	Switzer	Bob3	Trouble	Abrogate	Anglerfish	Saffire	Skipole	Conceptil	Museum	Severus
Resistant clones	X	✓	✓	✓	✓	X	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	X	X	✓	✓	✓	✓	✓	✓	✓	✓
PCR verified	-	✓	✓	✓	✓	-	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓	-	-	✓	✓	✓	✓	✓	✓	✓	✓

Figure 3-10 Summary of functional recombinases in M.smegmatis

3. 3. Discovery of new recombination sites

3. 3. 1. Identification of the integration site in M. smegmatis genome

To discover the exact location of the insertion site of the integrative vector in M.smegmatis genome, I used the plasmid rescue method (Figure 3-11). This method allowed me to identify the genomic sequences flanking the integrative vectors insertion site.

First, I digested the extracted genomic DNA from the resistant clones with EcoRI restriction enzyme. This restriction digest reaction would normally result in 1199 independent fragments. However, I intentionally inserted one and only one EcoRI site in the integrative vector sequence, in between the putative attP region and the recombinase cassette. Restriction digest of the genomic DNA from the resistant clones would therefore result in 1200 independent fragments. I then ligated all these fragments together, transformed E.coli with the whole reaction mix and plated on selective media. While the majority of the ligation products would not be able to replicate in E.coli, this method allows the ligated fragment that contains both the resistance gene cassette and the origin of replication from our integrative vector to replicate in E.coli, conferring the resistance to Kanamycin to the bacteria. Because the integrative vector contained a single EcoRI restriction site, the recircularized product also contains some chromosomal sequences from *M.smegmatis* flanking the insertion site. Therefore, I expanded the resistant clones that grew on selective media, miniprepmed them and sequenced the extracted vectors. The sequencing results unveiled the sequence flanking the insertion site and allowed me to infer the sequence of the attB site.

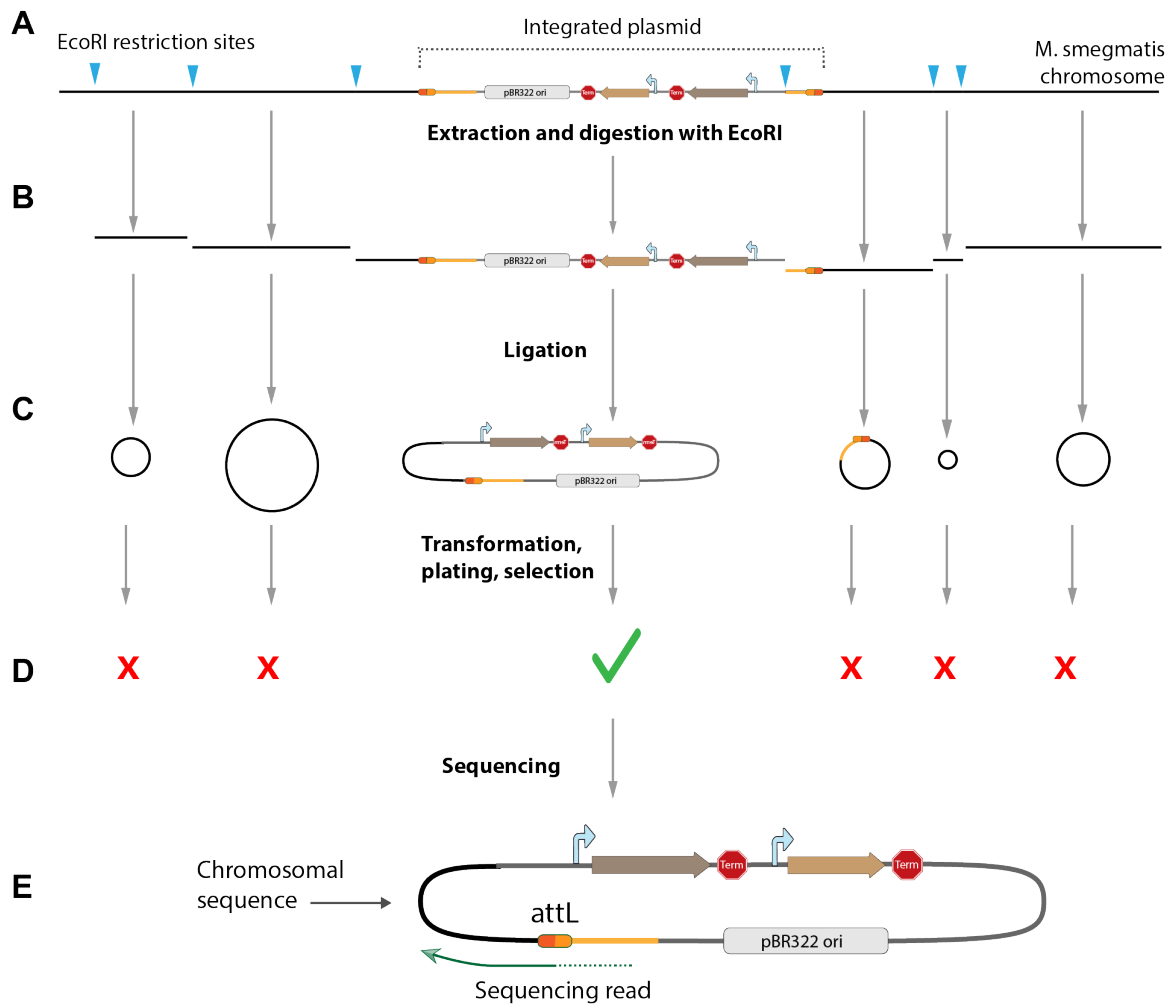


Figure 3-11 Plasmid-rescue workflow. (A) The genome sequence of *M. smegmatis* carries a large number of EcoRI restriction site (blue arrows) while the linearized integrated vector carries a single EcoRI restriction site. (B) Genomic DNA is extracted from a resistant clone and digested with EcoRI, leading to the presence of thousands of linearized DNA pieces among which one contains both part of the *M. smegmatis* genome and part of the integration vector with the origin of replication and the resistance cassette. (C) Ligation of these parts results in the circularization of the linearized DNA sequences. (D) When transforming all these parts into competent *E. coli* cells, only the vector containing the pBR322 origin of replication and the resistance cassette will be able to propagate and allow *E. coli* to grow on selective media. (E) Sequencing of the ligated vector reveals the sequence of the insertion site.

Out of the 23 putative recombinases I tested, 15 of them integrated into the same genomic attB site as BxB1 (*groEL* CDS), two integrated in the TmrH RNA methyltransferase CDS (Rebeuca and Veracruz) and two others integrated in the 2-nitropropane dioxygenase CDS (Figure 3-12). I could not find any information concerning the importance of these genes in the growth regulation or metabolism of *M. smegmatis*. However, the resistant clones that I expanded to prepare the plasmid-rescue workflow grew at a normal rate and with a normal phenotype. Unfortunately, I did not have the time to finish the whole plasmid-rescue workflow for the last four recombinases.

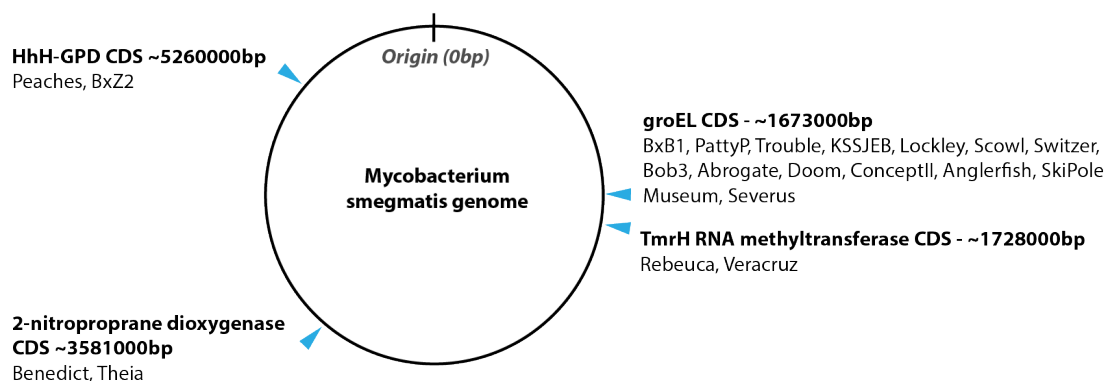


Figure 3-12 Localization of the attB sites for different large site-specific recombinases from Mycobacteriophages

Based on the sequencing results, I could infer the point of strand exchange of the site-specific recombination reaction and then predict the approximate attB and attP sites sequences.

For the recombinases that integrated into the groEL CDS, the attP and attB sites are totally homologs to the one recognized by BxB1 recombinase.

It is interesting to see that for Rebeuca and Veracruz, even though integration happens in the same attB site, the attP site recognized by Rebeuca is slightly different from the attP site recognized by Veracruz (**Figure 3-13**). This case is actually similar for Theia and Benedict recombinases and such differences had already been observed with the BxZ2/Peaches pair ⁹⁹.

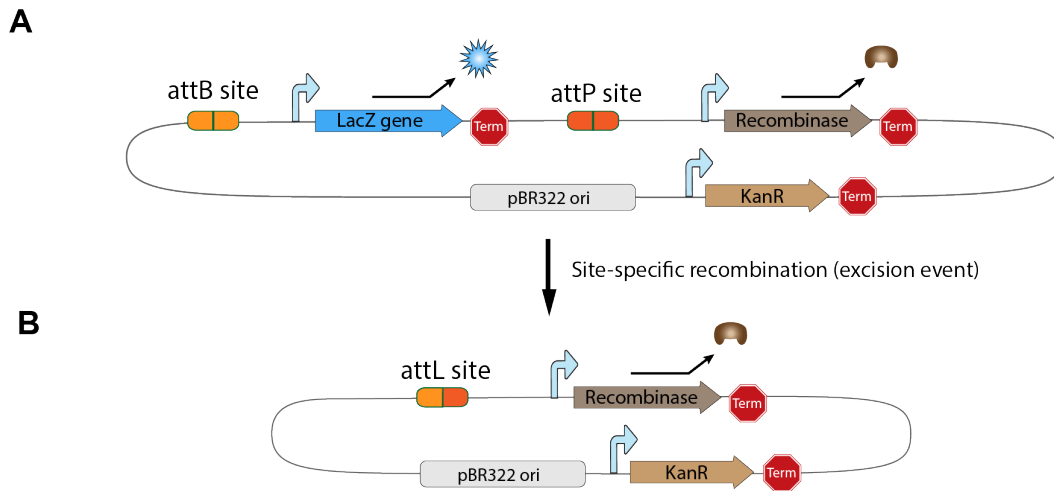


Figure 3-14 Intramolecular site-specific recombination reporter system

Given that my goal was to discover new recombinases and recombination sites, I created the reporter system only for the recombinases that did integrate in sites that were different from previously characterized ones, i.e. Theia, Veracruz, Rebeuca and Benedict.

I transformed the vectors in *E. coli*, plated on selective media and I incubated the plates overnight. Given that pBR322 is a relatively low copy plasmid, excision of the lacZ gene from all the copies within a cell should be achieved relatively rapidly. I screened for blue or white colonies the next morning and for all 4 recombinase, more than 99% of the obtained resistant clones were LacZ negative, validating the exactitude of the attP and attB sites.

To further validate that the non-expression of LacZ was due to the excision of the LacZ cassette and not a consequence of instability of the construct, I sequenced the plasmids miniprepmed from white clones. The results were positive as the sequencing showed the presence of the expected deletion and the attL site created by site-specific recombination between attB and attP site.

To determine the exact minimal sequence required for the att sites to enable site-specific recombination, further tests should be performed. For instance, one could synthesize a library of att sites with different lengths and repeat the experiment described above. When the att site tested will be too short, resistant clones should all express the LacZ gene even after an extended period of time.

3. 4. Validation of recombinases activity in mammalian cells

3. 4. 1. Assessment of intermolecular recombination

Now that I had confirmed both the activity of these recombinases for intramolecular site-specific recombination in *E.coli* and the correct sequences of their attB and attP sites, I wanted to further test whether they could also be used for the manipulation of DNA sequences in mammalian cells. I therefore created a reporter system based on two separate plasmids to test their intermolecular site-specific recombination capacity in mammalian cells (Figure 3-15). The first plasmid carried a gene cassette allowing the constitutive expression of EYFP. The attB site was placed in between the Hef1a constitutive promoter and the EYFP coding sequence. The second plasmid carried a promoterless gene cassette composed of the attP site followed by the mKate2 coding sequence. When co-transfecting these two plasmids together with a plasmid expressing the corresponding recombinase, intermolecular site-specific recombination between the attB and the attP sites would result in the fusion of the two plasmids and therefore in the insertion of the promoterless mKate2 coding sequence in frame with the Hef1a promoter. This recombination event would therefore trigger the constitutive expression of mKate2 and interrupt the expression of EYFP as it becomes promoterless.

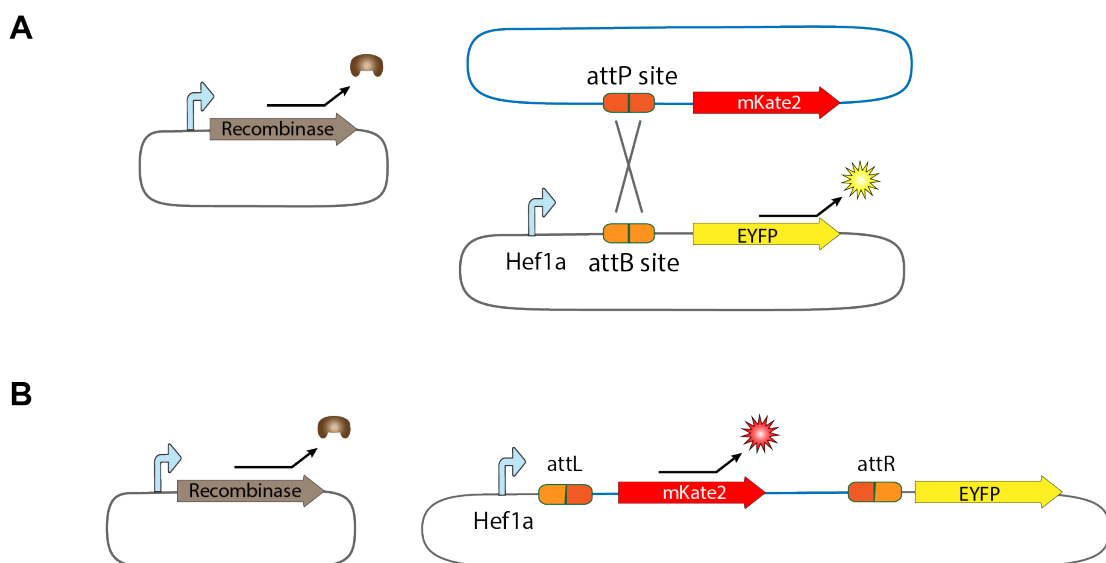


Figure 3-15 Intermolecular site-specific recombination mammalian reporter system. (A) The reporter system is composed of 2 plasmids: 1 plasmid harboring the attP site followed by the mKate2 coding sequence and 1 plasmid carrying the Hef1a promoter followed by the corresponding attB site and the EYFP coding sequence. I also created a third vector to constitutively expresses the recombinase. **(B)** After co-transfection of these 3 plasmids in mammalian cells, intermolecular site-specific recombination between the plasmids carrying attB and attP site should result in the constitutive expression of mKate2. Depending on the recombination efficiency, some plasmids may still express EYFP a few hours after transfection.

I co-transfected these two plasmids together with a third plasmid to express the recombinase and assessed the intermolecular recombination events based on the fluorescence output. If the recombinase mediates site-specific recombination between the separated attB and attP sites, the two plasmids fuse together and the mKate2 coding sequence is placed just after the Hef1a promoter while the EYFP coding sequence becomes promoterless.

As depicted on Figure 3-16, I demonstrated that the 4 new recombinases could mediate intermolecular recombination in mammalian cells.

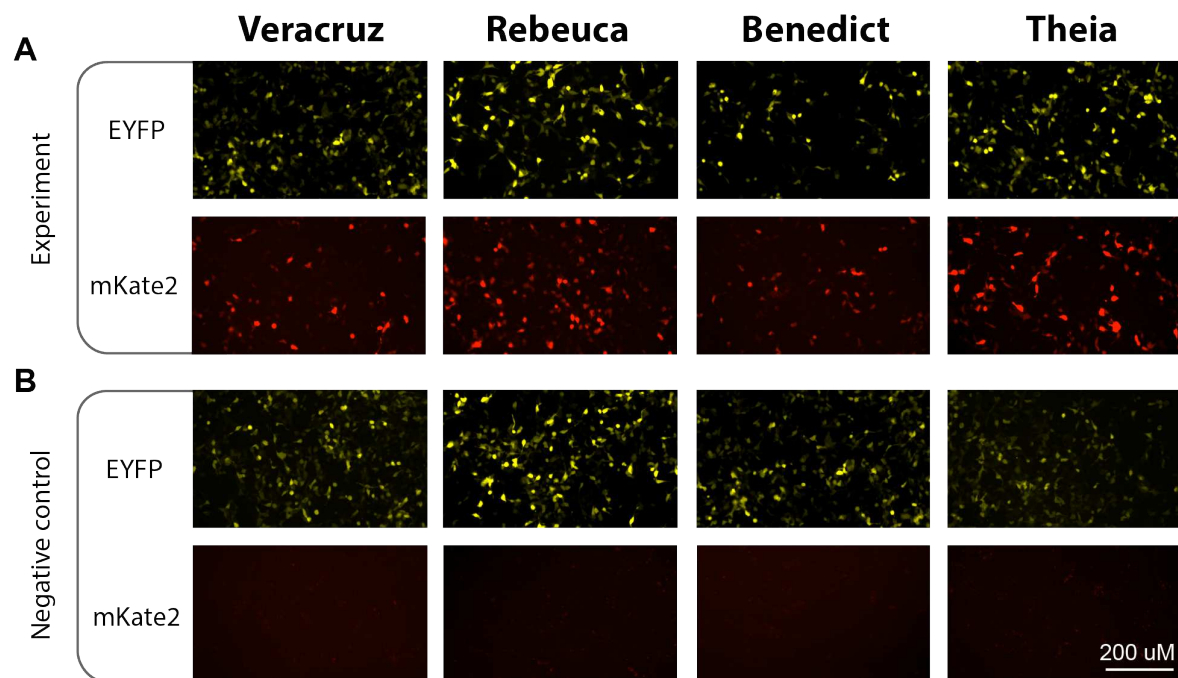


Figure 3-16 Fluorescent microscopy images of mammalian cells 36h after transfection of the intermolecular site-specific recombination reporter system. (A) Co-transfection of both the reporter system and the vector expressing the recombinase leads to efficient intermolecular recombination between the 2 plasmids of the reporter system as depicted by expression of mKate2. (B) Negative control of the same experiment: only the 2 plasmids of the reporter system were transfected without the vector expressing the recombinase.

3. 4. 2. Assessment of recombination crosstalk

To test whether these recombinases were orthogonal to each other, i.e. could only mediate site specific recombination between their own attB and attP sites identified in previous experiments, I co-transfected each recombinase expression vectors with the reporter systems corresponding to all recombinases. The recombinases that had different attB sites were completely orthogonal to each others (no mKate2 expression). However, the recombinases, which shared the same attB but a different attP site, could mediate site-specific recombination using both attP sites (significant mKate2 expression). The results are summarized in Figure 3-17.

		Recombinase					
		Peaches	BxB1	Benedict	Theia	Veracruz	Rebeuca
attB/attP	Peaches						
	BxB1						
	Benedict						
	Theia						
	Veracruz						
	Rebeuca						

Figure 3-17 Crosstalk of different large serine recombinases. Efficient site-specific recombination between attB/attP pair is represented by green rectangles. On the opposite, rectangles in red illustrate a non-functional recombinase for the corresponding attB/attP site pair.

3. 4. 3. Creation of a library of orthogonal sites

The nonpalindromic central dinucleotide at the center of *attP* and *attB* sites is the sole determinant of the prophage orientation ¹⁰⁴, and a single base pair substitution in the two sites inhibits recombination. However, when identical central dinucleotide mutations are in both *attP* and *attB* partners, recombination is restored, as demonstrated for BxB1 and PhiC31 in bacterial cells ^{104,105}. Moreover, for the directionality of the reaction to be retained, these two central dinucleotide have to be non-palindromic.

Based on these requirements, I therefore synthesized 6 different variants of the central dinucleotide sequence for each attP and attB partners that were non-palindromic and tested them against each other in transient transfections to see whether they were orthogonal in mammalian cells.

As predicted, only attP and attB that shared the exact same central dinucleotide sequence could recombine, and this for all 7 recombinases tested (i.e. 42 orthogonal sites) (Figure 3-18).

Based on these results, I have extended the landing pad concept developed in the first chapter of my PhD thesis. For this purpose, I have created a stable cell line with two landing pads, each one

carrying one BxB1 attB site orthogonal to the other. I then demonstrated that I could integrate a circuit in the landing pad of choice depending on the attP site carried by the integrative circuits (results not shown).

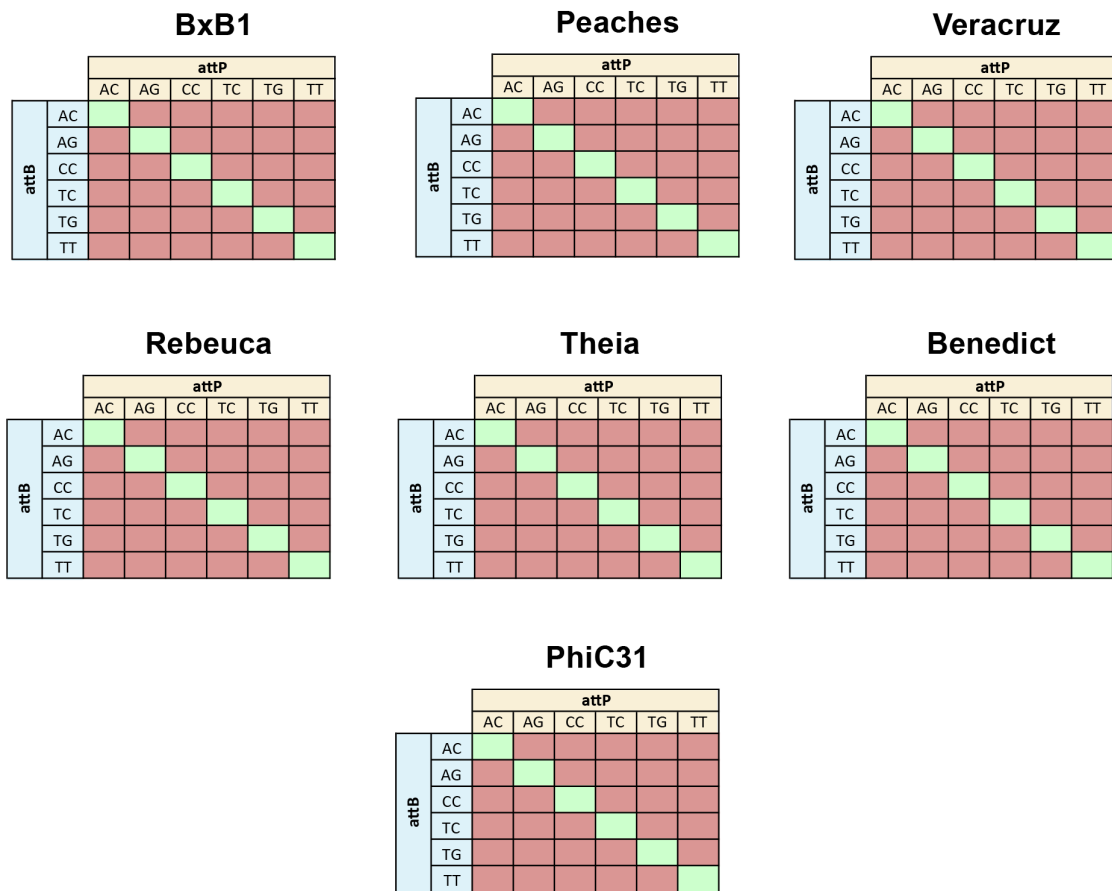


Figure 3-18 Library of orthogonal recombination sites for 7 large serine-recombinases.

3. 5. Conclusion and discussion

The recombinases that have been shown to work in eukaryotic cells are currently used to manipulate cells and animal's genomes (gene therapy for instance). They are however used empirically with little or no understanding of the factors or processes which may control their specificity and limit their efficiency. It would therefore be worthwhile to try to gain a deeper understanding of the extent to which these limitations can be overcome. This may help us to improve the efficiencies of site-specific recombinases but also to engineer their specificity. Such studies would benefit from the ability to compare the results obtained with different integrases and it seems therefore appropriate to itemize all of the serine integrases potentially useful in vertebrate cells.

Towards this goal, I have identified 26 new large serine recombinases from recently sequenced Mycobacteriophages genomes (2013) and engineered a dedicated plasmid rescue system to discover the specific attB/attP recombination sites recognized by these recombinases. While most of them shared a high degree of homology with the well-characterized BxB1, some had a significantly different amino-acid sequence. For the latter ones, it is interesting to note that the nucleotide sequence supposed to contain the attP site and preceding the recombinase coding sequence was also significantly different compared to the sequence at the same location from BxB1. This was another argument to suppose that these recombinases would use different attP and attB sites.

These assumptions were confirmed as I managed to discover 4 new attP/attB site pairs for Mycobacteriophages Theia, Benedict, Veracruz and Rebeuca used to stably integrate their own genome into *M. smegmatis* chromosome. On the one hand, Benedict and Theia recombinases used the same attB site within the 2-nitropropane dioxygenase CDS, and on the other hand, Rebeuca and Veracruz used the same attB site within the TmrH RNA methyltransferase CDS. While the attB specificity for these two pairs of recombinases is extremely high, the attP specificity is less stringent as Theia can use Benedict's attP and vice-versa, and Rebeuca can use Veracruz's attP and vice-versa. A similar case had already been studied with Peaches and BxZ2 mycobacteriophage recombinases.

The workflow I have set up to identify new serine recombinases and discover their att sites could now be automatically used for newly sequenced isolated mycobacteriophages in order to extend the toolbox of available recombinases.

A deeper analysis of these sites could even help us shorten the pace of att site identification by skipping the required experiments in the natural host *M. smegmatis*. Indeed, all these sites share common features such as 6 to 8 base pair homologies between attB and attP sites core sequence

and numerous inverted repeats in the flanking sequences. It could therefore become very handy to develop an algorithm to 1) itemize all possible large serine recombinases within sequenced mycobacteriophage genomes, 2) itemize all putative attP sites within 200bp flanking the start codon of these recombinases, 3) scan mycobacterium genomes to itemize all putative attB sites sharing an homologous core sequence flanked by numerous inverted repeats.

In order to extend not only the number of characterized site-specific recombination systems but also the usage of these systems, I validated the functioning of the 4 recombinases I discovered in heterologous hosts: E.coli and mammalian cells. I also created for each of these recombinases a library of orthogonal sites that could be used in parallel at the same time and therefore broaden the spectrum of applications.

While intramolecular site-specific recombination has proven useful for excision of constructs integrated in mammalian genomes, intermolecular site-specific recombination can be used to integrate complex circuits into mammalian chromosomes (cf Chapter 2). The ability to multiplex integrations by using orthogonal sites could help to integrate a variety of different circuits at different locations within a genome. This could be very helpful to prevent interference between circuits or attain higher levels of expression (with a lower coefficient of variation) by integrating the same construct in multiple copies. This could also allow engineering cell lines step by step, integrating one circuit after the other so as to incrementally increase the complexity.

The use of multiple serine recombinases with orthogonal sites could also lead to the development of new DNA assembly methods. For instance, one could create a system based on the Gateway principle, flanking each part with either an attB on its 5' end and an attP on its 3'. Having 6 orthogonal sites for one recombinase would allow assembling 5 parts per reaction, i.e. enough for a mammalian transcription unit. The assembled vectors could carry sites from another recombinase, and following the same principle, could be used to assemble 5 transcription units together. Such strategy could be used to assemble even larger constructs given that we now have 6 orthogonal recombinases. As site-specific recombination with serine recombinases is extremely efficient, fast and reliable, such an assembly method would be extremely convenient.

List of Supplementary Data (found in Annexes)

Supp Figure 3-1 Nucleotide sequences of the attP and attB sites used in this study

3. 6. Materials and methods

In-silico identification of large serine recombinases. Scripts were written and run with Matlab2013a.

Expression units and plasmids assembly. All expression units and plasmids were assembled with the Golden Gate framework and are listed in the Annexes. For all Golden Gate assembly reactions, we used: 0.4 μl of TypeIIS enzyme (either BsaI from NEB, or BpiI from Fermentas), 0.2 μl of T4 Ligase HC + 1 μl of T4 Ligase HC buffer (Promega), 1 μl of 10xBSA (NEB), 40 fmol for all vectors used in the reaction, ddH₂O up to a final total volume of 10 μl . The thermocycler program used for all assemblies included: 1 step of 15 min at 37°C; then 50 cycles of [2 min at 37°C followed by 5 min at 16°C]; 1 step of 15 min at 37°C, 1 step of 5 min at 50°C and 1 final step of 5 min at 80°C.

Bacterial cell cultures. Liquid cultures of *E.coli* MG1655 were grown in LB Medium (Difco) at 37°C. When appropriate, antibiotics were added as follows: spectinomycin (100 $\mu\text{g}/\text{mL}$), ampicillin (100 $\mu\text{g}/\text{mL}$) and kanamycin (25 $\mu\text{g}/\text{mL}$). For blue/white screening, we used X-gal at a final concentration of 40 $\mu\text{g}/\text{mL}$.

For liquid cultures, *M. smegmatis* was grown in Difco 7H9 liquid medium supplemented with 0.5% glycerol, 0.5% Tween 80 and 10% albumin–dextrose complex (ADC). Carbenicillin and cycloheximide were added to all cultures at concentrations of 50 mg ml^{-1} and 10 mg ml^{-1} respectively. When necessary, the following antibiotics were also added; kanamycin (8 mg ml^{-1}), Hygromycin (50 mg ml^{-1}) and tetracycline (0.5 mg ml^{-1}). For solid cultures, *M. smegmatis* was grown on Difco 7H10 agar supplemented with 0.5% glycerol and 10% ADC.

Mammalian cell culture and transfections. HEK293FT cell line was purchased from Invitrogen. HEK293FT cells were maintained in Dulbecco's modified Eagle medium (DMEM, Cellgro) supplemented with 10% FBS (PAA), 0.045 g/mL penicillin/streptomycin and non-essential amino acids (HyClone) at 37°C, 100% humidity and 5% CO₂. HEK293FT transfections were carried out in 24-well plates using Attractene reagent (Qiagen), 200000 cells and 200-300 ng total DNA per well (plasmid ratio 1 :1 :1). Media was changed 24 hours after transfection.

Microscope measurements and image processing. Fluorescence microscopy images of live cells were taken in glass-bottom dishes or 12-well plates using Zeiss Axiovert 200 microscope and Plan-Neofluar 10x/0.30 Ph1 objective. The imaging settings for the fluorophores were

S430/25x (excitation) and S470/30m (emission) filters for EYFP, and S565/25x (excitation) and S650/70m (emission) for mKate2. Data collection and processing were performed using AxioVision software (Zeiss).

PART 2

Engineering synthetic and orthogonal intercellular communication systems in mammalian cells

Contents

Introduction

Chapter 4

Hijacking plant metabolic pathways and bacterial nuclear receptors to create a de-novo small molecule based intercellular communication system

Chapter 5

Hijacking lentiviral capsids to create a multi-channel and modular protein-based intercellular

Part 2 Introduction

Multicellularity allows a living organism to overcome one of the major constraints imposed by diffusion: the size limitation. Single cells with increased size have indeed a decreased surface-to-volume ratio and have therefore difficulty in absorbing sufficient nutrients and transporting them throughout the cell. This confers multicellular organisms with the competitive advantages of an increase in size and enables an increase in complexity by allowing the differentiation of numerous cellular lineages within an organism.

The proper functioning of such multi-cellular organisms heavily relies on the ability of cells to communicate with each other. Many different kinds of molecules transmit information between the cells of multicellular organisms. Although all these molecules act as ligands that bind to receptors expressed by their target cells, there is considerable variation in the structure and function of the different types of molecules that serve as signal transmitters. Structurally, the signaling molecules used by plants and animals range in complexity from simple gases to proteins. Some of these molecules carry signals over long distances, whereas others act locally to convey information between neighboring cells. In addition, signaling molecules differ in their mode of action on their target cells. Some signaling molecules bind to receptors expressed on the target cell surface, whereas others are able to cross the plasma membrane to bind to intracellular receptors in the cytoplasm.

The development of these different kind of signaling molecule not only allowed cells to coordinate with one another to better survive in their respective environment, but also to organize the distribution and the execution of highly specific tasks and workload between different consortia of specialized cells. As such, intercellular communication increased the overall fitness of multicellular organisms.

Recently, the engineering of intercellular communication systems to coordinate population-level behavior has become of particular interest within the mammalian synthetic biology community; and this for two major reasons.

On the one hand, it can provide the tools to dissect important communication pathways broadly used across the living kingdom and finely understand the motifs and nuances of these

communications. Orthogonal systems have the major benefit of being partly buffered from contextual effects and therefore allow for precise study of specific components function without any cross-talk effect.

On the other hand, the use of cell-cell communication systems provides, from engineering prospective, the ability to design more complex genetic circuits and with new control dynamics for novel applications. The increasing complexity of gene circuits rises together with an increasing complexity of the encoded tasks to be performed by cells (Auslander et al, 2012a; Moon et al, 2012), which can hardly be performed by a single designer-cell. A proposed solution is therefore to distribute the tasks and metabolic workload among different specialized designer-cell populations, which will need to coordinate their activities to provide concerted actions. Such strategy has been successfully used to create systems to program pattern formation (Liu et al, 2011, Basu et al, 2005), to allow for increased precision and reliability in entire cell populations (Koseska et al, 2009) and to create interconnected multicellular assemblies close to those found in nature (Weber et al, 2007a, Bacchus et al, 2013). The possibility to spatially separate populations of communicating cells could also lead to new possibilities in various biomedical applications such as the design of biocompatible implants capable of communication and interfacing with our natural physiology (Duan and March, 2010).

While the communication modules used in these studies are promising, they rely on molecules with a widespread importance in existing mammalian cell-signaling pathways, which represent significant drawbacks. On the one hand, as the messaging molecules cross talk with endogenous cell-signaling pathways, it may results in undesired alteration of wild-type cells behavior, which can be problematic for any therapeutic application. On the other hand, any variation of metabolic states within an organism may result in fluctuations of signaling metabolite's concentration and therefore interfere with the synthetic cell-cell communication module's functioning of the engineered cell populations.

To overcome these limitations, I engineered and validated two new synthetic intercellular communication systems for mammalian cells that are orthogonal to endogenous mammalian cell-signaling systems. The first one uses the small and diffusible molecule Phloretin as a signaling molecule, while the second one uses virus like particles as carrier vectors to encapsulate and transfer the desired proteins from one cell to another.

Chapter 4 Hijacking plant metabolic pathways and bacterial nuclear receptors to create a de-novo small molecule based intercellular communication system

Contents

- 4.1 Introduction**
 - 4.2 Engineering of the receiver module**
 - 4.2.1 Phloretin responsive mammalian promoter (PRMP)
 - 4.2.1.1 Design*
 - 4.2.1.2 Influence of TtgR's operator sequence on the induction profile*
 - 4.2.1.3 Influence of transactivator vs. repressor ratio on induction profile*
 - 4.2.2 Creation and characterization of stable receiver cell lines
 - 4.2.2.1 Lentiviral mediated integration of the receiver module components*
 - 4.2.2.2 Characterization of induction profile of different receiver cell lines*
 - 4.3 Engineering of the sender module**
 - 4.3.1 Phloretin synthesis
 - 4.3.2 Creation of a stable sender cell line
 - 4.4 Validation of functional cell-to-cell communication**
 - 4.5 Conclusion and discussion**
 - 4.6 Materials and methods**
-

4. 1. Introduction

Hormones are a class of regulatory biochemicals that are produced by specialized cells in multicellular organisms. They serve as a major form of communication between different organs and tissue and are used in plants and animals to regulate a variety of physiological and behavioral activities, including growth and development, digestion and metabolism.

Whether it is to validate our current understanding of these systems, or to create new beneficial functions that would allow synthetic tissues to coordinate their actions, many believe it could be particularly interesting, though complex, to engineer synthetic hormone-like communication systems. Such advances could pave the way to next generation treatments for hormonal imbalance, which often leads to serious diseases requiring life-long treatment.

Towards these goals, a few synthetic systems have been designed but the signaling molecules they employ have a wide spread importance in endogenous mammalian cell signaling, resulting in significant cross talk with other natural gene networks. The most advanced piece of work, designed by Bacchus et al ¹⁸, demonstrated it was possible to create a synthetic mammalian two-way intercellular communication system with tunable dynamics. However, the molecules used as signaling molecules were L-tryptophan and acetaldehyde, which can be naturally found within the human body. While such strategy could be tuned to plug into the communication pathways of the human body, it could become challenging to obtain robust behavior of the systems given the significant natural fluctuations of endogenous signaling molecules concentrations.

To overcome such a limitation, I chose to design a new synthetic intercellular communication system based on an orthogonal signaling molecule. This molecule had to fulfill 4 major criteria. First, it had to be orthogonal to endogenous signaling systems (i.e. not naturally produced or encountered in the human body). Second, it had to be non-toxic in the concentration ranges required for the proper functioning of the synthetic communication system. Third, it had to be synthesizable by mammalian cells. And fourth, it had to be able to trigger the activation of a transgene.

After a thorough literature review, I identified the small and diffusible molecule phloretin as a good candidate. Phloretin is a natural antimicrobial mainly found in the root bark of apple trees, showed to inhibit the growth of many bacterial species ^{101,102}. In the specific case of

Pseudomonas putida, it has been demonstrated that phloretin induces the expression of the genes coding for the TtgABC efflux pump¹⁰³. The TtgR repressor protein, that naturally seats on its cognate operator (O_{TtgR}), changes conformation upon binding with phloretin and detach from the TtgABC promoter, enabling expression of the TtgABC genes¹⁰⁶. After showing that phloretin had a extended half-life in culture media and no negative influence on viability and growth of cells, Gitzinwer et al were able to transpose this system in mammalian cells to repress gene expression when adding phloretin to the culture media¹⁰⁷. Such a system is actually very close to the well characterized and widely used Tet-Off system. However, while production of Doxycycline (used to activate the Tet-Off system) had never been achieved in living cells, I found that three groups had already reported the efficient production of phloretin in genetically engineered *Escherichia coli* and *Saccharomyces cerevisiae* by building an artificial gene cluster containing the CHS and 4CL1 plant genes¹⁰⁸⁻¹¹⁰.

Based on all these results, I decided to engineer a synthetic mammalian communication system based on phloretin and inspired by endogenous hormone signaling pathways. I therefore designed a sender module and a receiver module that could be used either in separate cells to create a paracrine-like communication system (**Fig. 4-1A**) or in the same cells to create an autocrine-like communication system (**Fig. 4-1B**).

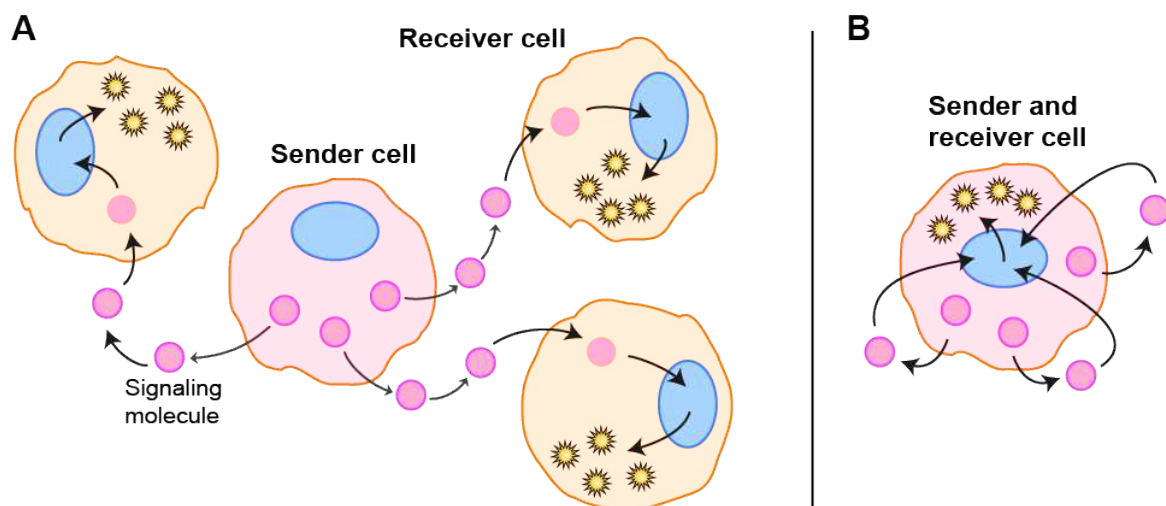


Figure 4-1. Overview of the synthetic hormone-like signaling system

(A) In the paracrine setup, sender cells are producing the signaling molecule, phloretin, which diffuses through their membrane and the media, to enter into the receiver cells. Phloretin then interacts with a synthetic receptor in the nucleus and triggers transgene expression. **(B)** In autocrine setup, both the sender and the receiver modules are integrated in the cells. The cell can therefore synthesize the signaling molecule but also respond to it in a dose-dependent manner.

In order to mimic the effect of hormones such as the thyroid hormone on its responsive tissues, I engineered a synthetic intracellular receptor based on *Pseudomonas putida* TtgR repressor, that would trigger transgene activation upon binding to phloretin in the receiver cell population (as opposed to Gitzinwer et al design). I characterized the dynamics and validated the proper functioning of this new and complex inducible system both in transient transfection and with stable cell lines. To achieve the synthesis of phloretin in the sender cell population, I assembled a *de-novo* metabolic pathway in the sender cell population by expressing the 4CL1 enzyme from the plant *Arabidopsis thaliana* and the CHS enzyme from the plant *Hypericum androsaemum*. To validate the paracrine-like function of the approach, I co-cultured both the sender and receiver cell populations and demonstrated the phloretin dose-dependent activation of a transgene in the receiver cell population. To our knowledge, this is the first study establishing a functional communication system in mammalian cells using a fully orthogonal signaling molecule.

4. 2. Engineering of the receiver module

4. 2. 1. Phloretin responsive mammalian promoter (PRMP)

4. 2. 1. 1. Design

In order to use phloretin as an orthogonal inducer of transgene activation within the receiver cell population, I have chosen to engineer a synthetic mammalian promoter mimicking the activation pattern of the natural *Pseudomonas putida*'s phloretin responsive promoter (P_{TtgR}).

In this case, gene expression is enabled by the release of the repressor protein TtgR from its cognate operator (O_{TtgR}), mediated by a conformational change of the repressor upon binding with phloretin.

To transpose the natural functional of the bacterial promoter into mammalian cells, I created the synthetic Phloretin Responsive Mammalian Promoter (PRMP) (**Fig. 4-2A**). This promoter is constituted of a repetition of 5 Gal4 binding domains, which are followed by the minimal human cytomegalovirus immediate early promoter (CMV). Such a basic architecture drives the constitutive expression of a downstream transgene when the Gal4VP16 transactivator would be expressed. I then decided to place on each side of the CMV promoter a TtgR cognate operator. In absence of phloretin, the expressed TtgR proteins should dimerize and attach to these sequences. Rationally, this would cause a steric hindrance for the proper assembly between the transcription machinery recruited by the Gal4-VP16 transactivator proteins and the CMV promoter; and should repress downstream transgene expression (**Fig. 4-2B**). Release of repression should be triggered by the phloretin-induced dissociation of the bound TtgR dimers,

allowing the proper interaction between Gal4-VP16 proteins, the transcription machinery and the CMV promoter (Fig. 4-2C).

As a whole, the functioning of the receiver module would therefore relies on the constitutive expression of both the Gal4-VP16 transactivator and the TtgR repressor; that would interact with this complex synthetic promoter in frame with a transgene coding sequence (Fig. 4-2D).

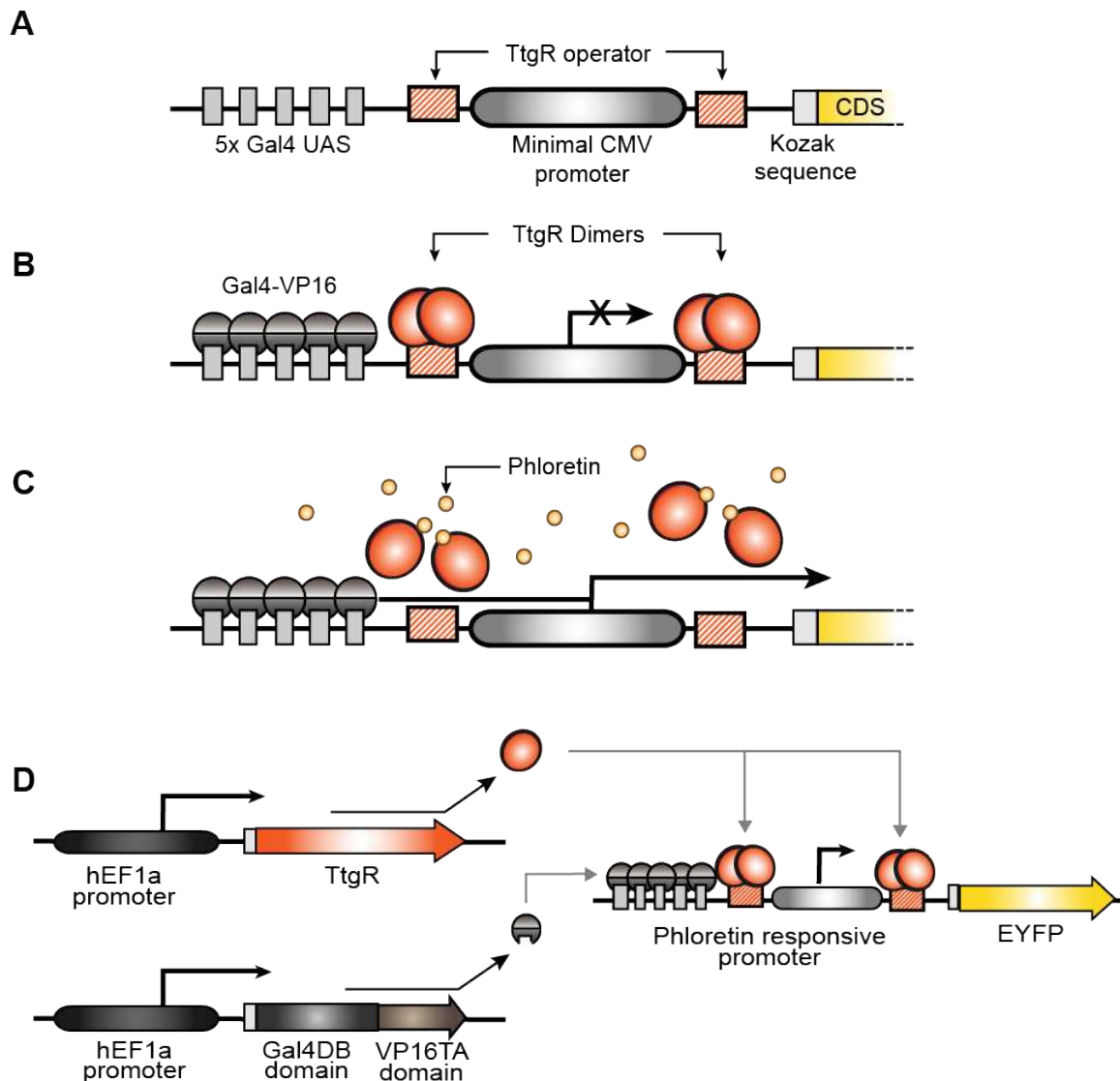


Figure 4-2. Phloretin Responsive Mammalian Promoter (PRMP)

(A) The PRMP is composed of Gal4 UAS operators, TtgR operators and a minimal CMV promoter. **(B)** In the absence of phloretin, TtgR dimers are bound to the TtgR operators and prevent transgene transcription. **(C)** In the presence of phloretin, the TtgR dimers dissociate from their cognate operators and enable transgene transcription. **(D)** The three elements that are required to create a complete receiver module.

4. 2. 1. 2. Influence of TtgR's operator sequence on the induction profile

Recent studies on the binding properties of TtgR dimers with their cognate operator have shown that its relative affinity could be increased 4 fold simply by changing 6 bp within the TtgR operator sequence ¹¹¹. These mutations (6bp) create longer inverted repeats (**Fig. 3**) that are likely to be linked with the reported increased affinity.



Figure 4-3. Comparison of O_{TtgR} and $O_{TtgR-Mut}$ nucleotide sequence

The 6 bp that have been mutated are in bold within the WT TtgR operator and in red within the Mutant TtgR operator. TtgR dimers have an increased affinity for the mutant TtgR operator due to longer inverted repeats (grey arrows).

To create synthetic promoters with different activation dynamics, I therefore synthesized two versions of the phloretin responsive mammalian promoter design: the PRMP with WT TtgR operators and the PRMP-Mut with the Mutant TtgR operators (**Fig. 4-4A,B**). I then transfected each of the corresponding reporter system together (PRMP/PRMP-Mut) with Gal4VP16 and TtgR expression vectors in a 1:1:1 ratio in Hek293FT cells. I then induced EYFP expression using different concentrations of phloretin and recorded fluorescence levels 48 hours later (**Fig. 4-4C**).

Induction of the receiver module with Phloretin resulted in activation of EYFP expression in a dose-dependent manner, though with different dynamics depending on the TtgR operator sequence used in the synthetic promoter. On the one hand, activation of PRMP reached its maximum with 10uM of phloretin (4 fold activation) and higher concentrations did not increase the transgene expression levels. On the other hand, activation of PRMP-Mut increased quasi-linearly together with phloretin up to 50uM of phloretin for which it reached a 5x induction level. As expected, the higher affinity of the TtgR dimers for the $O_{TtgR-Mut}$ of the PRMP-Mut compared to the O_{TtgR} of the PRMP resulted in a relatively lower basal activation of EYFP expression in absence of phloretin.

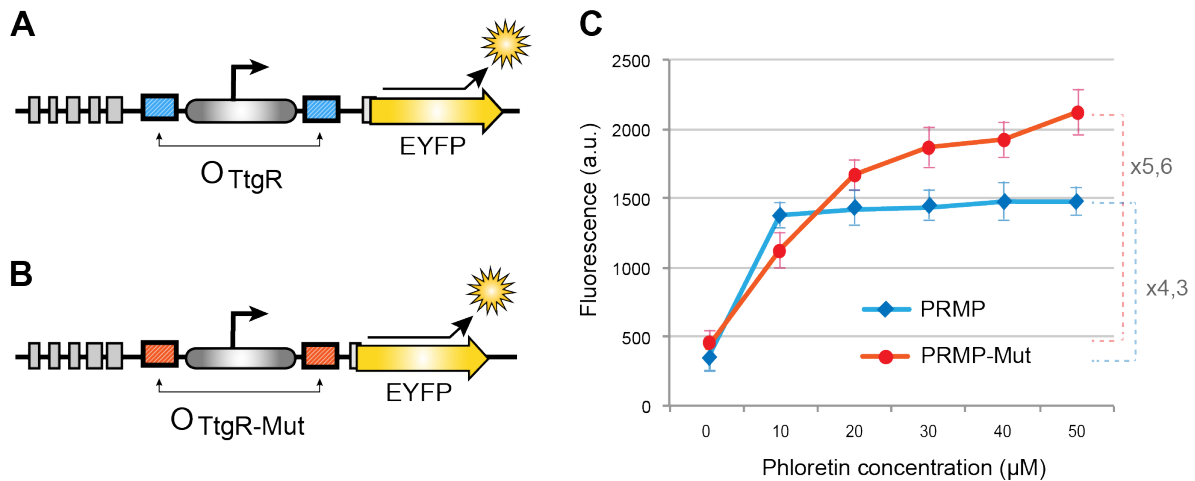


Figure 4-4. Dose dependent activation of Phloretin Responsive Mammalian Promoters in transient transfections

(A) Schematic of the PRMP reporter vector. (B) Schematic of the PRMP-Mut reporter vector. (C) Dose-dependent activation of the PRMP and PRMP-Mut measured by flow cytometry. Standard error was calculated based on two replicate experiments.

4. 2. 1. 3. Influence of transactivator vs. repressor ratio on induction profile

Given that the receiver module was relying on both the constitutive expression of a transactivator and a repressor, I wanted to test the influence of different ratios of these two proteins on the dynamics of the promoter.

I therefore transfected different ratios of Gal4VP16 transactivator expression vector vs. TtgR repressor expression vectors (2:1, 1:1 and 1:2) with a fixed amount of either PRMP or PRMP-Mut vectors and induced the unbinding of TtgR dimers from its operators using different concentrations of phloretin (Fig. 4-5). For both promoters, the ratio of activator vs. repressor influenced the promoter dynamics. When repressor relative concentration was higher than activator expression, the induction was less efficient but at the same time, the leakiness of the promoter was reduced.

The opposite was observed when activators were in excess compared to the repressors: while the non-induced module had a higher basal expression, the induction was more efficient.

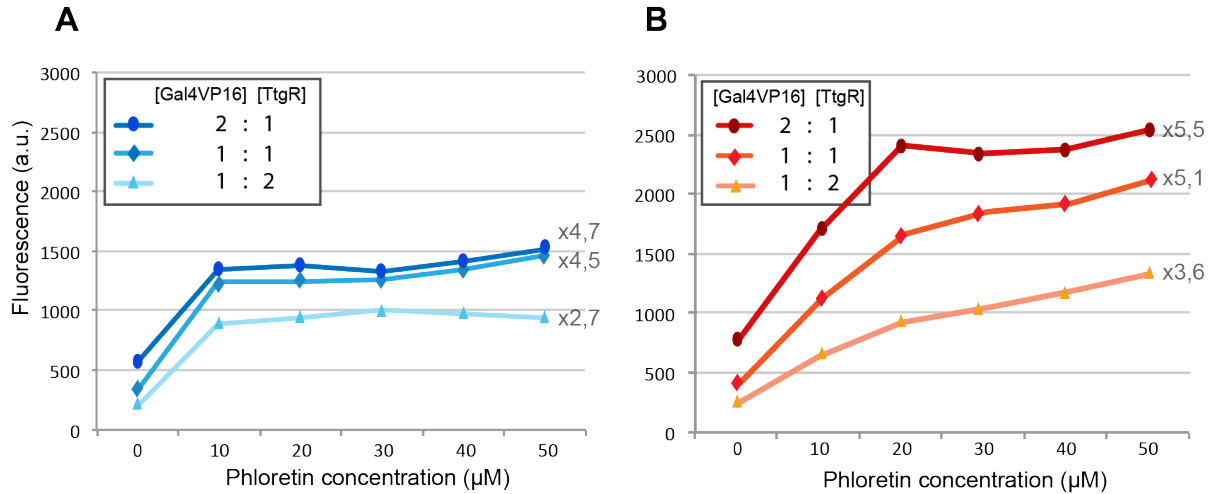


Figure 4-5. Influence of activator vs repressor ratio on the PRMP/PRMP-Mut activation dynamics

Boths PRMP **(A)** and PRMP-Mut **(B)** have a lower basal expression when TtgR concentration is higher than GalPV16 concentration. However, the opposite ratio allows a higher fold change when induced with phloretin.

4. 2. 2. Creation and characterization of stable receiver cell lines

4. 2. 2. 1. Lentiviral mediated integration of the receiver module components

To generate phloretin-responsive cell lines for robust induction of a transgene over time, I decided to use lentiviruses to stably integrate each component of the receiver module. Such a module had never been designed and therefore characterized stably integrated in the chromosome. Lentiviral strategy allows generating cell lines with many different integration profiles: different copy numbers and different integration loci. Such a strategy would therefore lead to the generation of a polyclonal cell line containing cells with a broad dynamic spectrum for the receiver. Moreover, I wanted to ensure a high level of induction of the system and this could only be achieved with multiple integrated copies of the transgene vector.

I therefore created a set of lentiviral vectors to stably integrate the 3 components of our receiver module:

- 1 lentiviral vector for the constitutive expression of the Gal4VP16 transactivator (**Fig. 4-6A**)
- 1 lentiviral vector for the constitutive expression of the TtgR repressor (**Fig. 4-6B**)
- 1 lentiviral vector for the PRMP reporter system (**Fig. 4-6C**)
- 1 lentiviral vector for the PRMP-Mut reporter system (**Fig. 4-6D**).

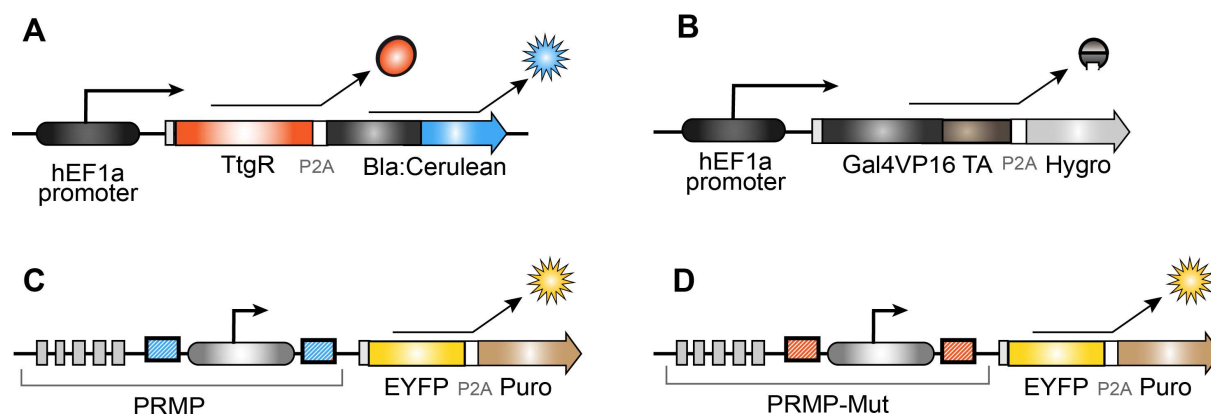


Figure 4-6. Set of lentiviral vectors to create PRMP and PRMP-Mut stable cell lines. (A) This vectors enable the constitutive expression of TtgR and a fluorescent protein coupled to a selection marker. (B) This vector enables the constitutive expression of the Gal4VP Transactivator coupled to a selection marker. (C, D) These vectors comprise either the PRMP or the PRMP-Mut.

The stable cell lines were generated in 4 successive steps (**Fig. 4-7**).

I first infected Hek293FT cell lines with both activator and reporter (PRMP or PRMP-Mut) vectors (**Fig. 4-7A**). This way, depending on the relative number of integrated copies of

Gal4VP16 transactivator vectors vs. PRMP or PRMP-Mut reporter vectors, I could select for integration a wide range of resulting EYFP levels. One week after transfection and 3 days after selection with puromycin, I decided to separate the EYFP positive cells in 2 different bins (HEK-PRMP a,b & HEK-PRMP-Mut a,b) (**Fig. 4-7B**), ranging from high to low EYFP levels, for which levels of Gal4-VP16TA were high enough to constitutively activate the PRMP promoter in the absence of the TtgR repressor.

I then infected these polyclonal populations with the repressor vector (**Fig. 4-7C**) and FACS sorted the cells 7 days later based on their EYFP levels. For each version of the promoter, I decided to isolate 3 bins resulting in 6 different polyclonal cell lines (**Fig. 4-7D**). I sorted in the first bin the cells, which had a high level of Cerulean, indicating a high level of TtgR. In these cells, the PRMP and PRMP-Mut were significantly repressed in the absence of phloretin. The second bin consisted of the cell with a low level of Cerulean, and for which EYFP expression was particularly un-repressed in the absence of phloretin. Finally, the last bin consisted of cells with fluorescence levels in between the cells from the other bins.

At the end of this 1 month long sorting experiment, I had a total of 12 polyclonal cell lines with different levels of basal activation and repression of EYFP expression (HEK-PRMP 1, 2, 3, 4, 5, 6 & HEK-PRMP-Mut 1, 2, 3, 4, 5, 6).

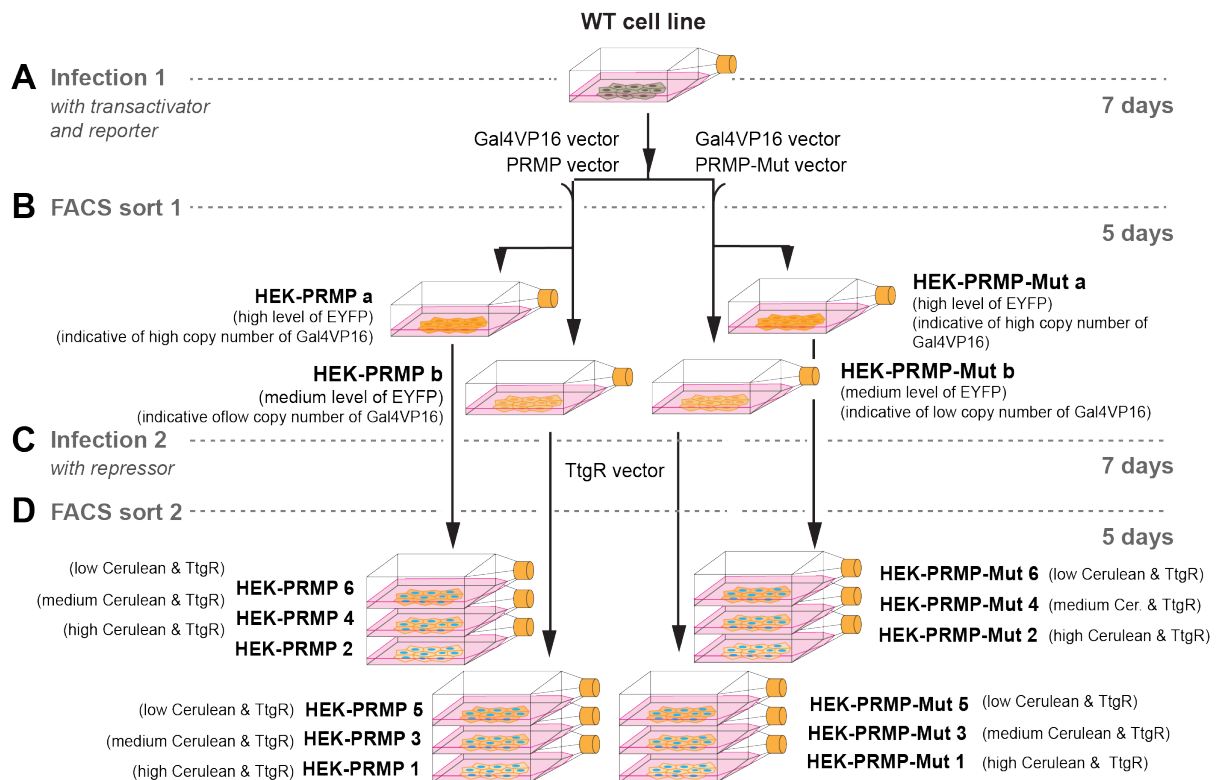


Figure 4-7. Workflow to create stable receiver cell lines.

4. 2. 2. 2. Characterization of induction profile of different receiver cell lines

I induced the activation of PRMP/PRMP-Mut with different concentrations of phloretin to characterize the induction dynamics of the different receiver cell lines, and measured EYFP levels 48h later by flow cytometry when cell density reached confluence. The dynamics profiles obtained with these stable cell lines confirmed the previous results obtained from transient transfection experiments: a sharp increase of transgene activation was measured with as little as 10uM of phloretin for cell-lines carrying the PRMP reporter system (2/3 of maximum recorded fold change for majority of cell lines). For all the PRMP cell lines, maximum expression levels were induced with 30uM of phloretin (**Fig. 4-8A**) while higher level of phloretin resulted in a relatively lower induction. Among all PRMP cell lines, Hek PRMP 6 was the one with the highest fold-change between non-induced and induced (x11.2) even though its basal expression level of EYFP without inducer was twice as high as other sorted cell lines (Hek PRMP 1,2).

Cell lines carrying the PRMP-Mut had a significantly more linear induction profile over the different ranges of Phloretin concentration, reaching maximum expression levels with 50uM of phloretin (**Fig. 4-8B**). Hek PRMP-Mut 6 has a particularly high fold-change between the non-induced and induced state (x10.9).

In both cases (PRMP and PRMP-Mut), cell lines with a high:low ratio of transactivator vs. repressor could attain higher level of fluorescence after induction with phloretin, resulting in higher fold-changes in transgene activation (cell lines 4-5-6 vs. 1-2-3).

While these results were observed on polyclonal populations, it is very likely that even higher fold changes could be attained with clonal cell lines. Given the tedious and time-consuming effort required to isolate, expand and test at least 4 or 5 clones for each of the 12 polyclonal populations, I preferred to first focus on the validation of the functioning of the intercellular communication and therefore proceeded with these polyclonal population for the co-culture experiments detailed later.

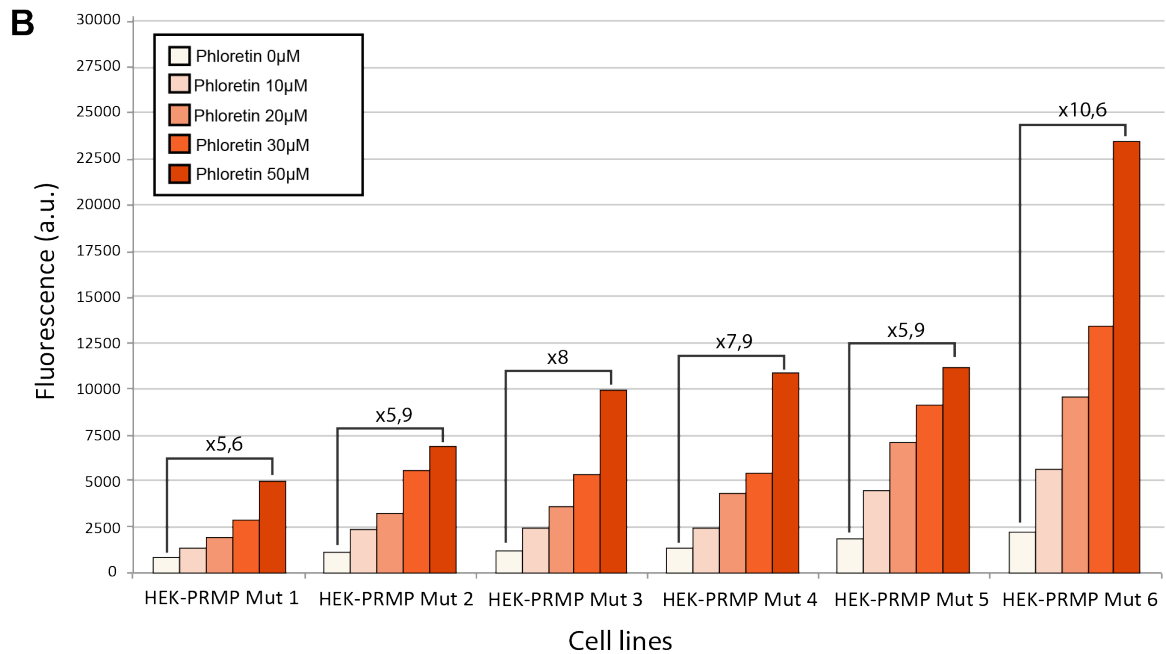
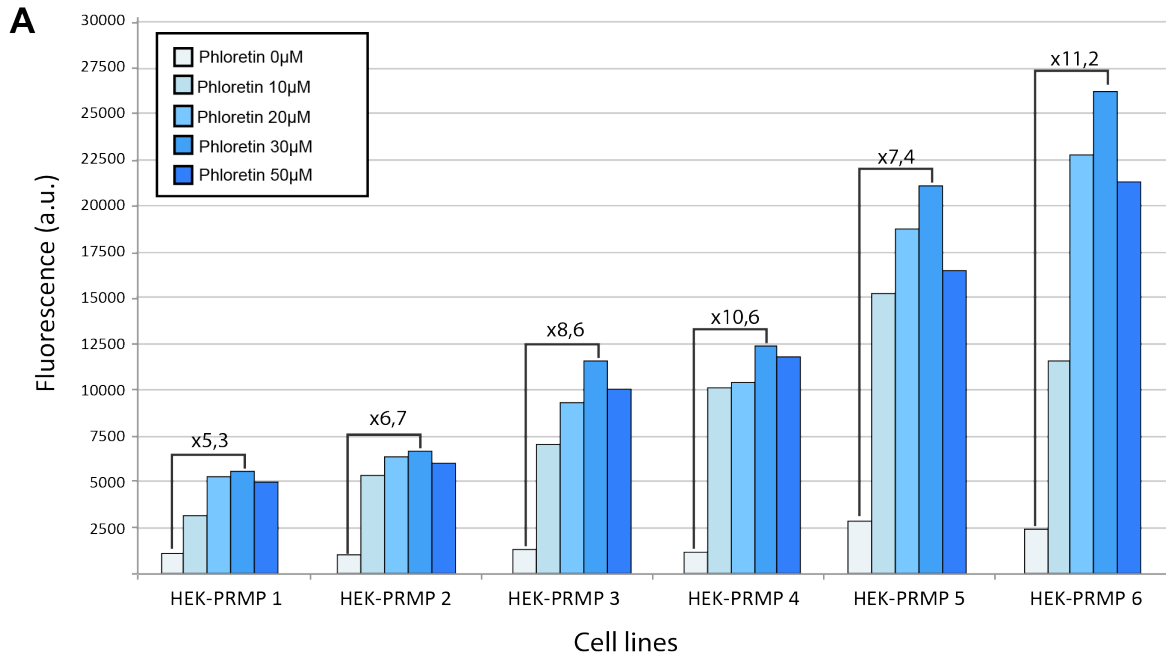


Figure 4-8. Induction profiles of the PRMP and PRMP-Mut cell lines

(A) PRMP polyclonal cell lines plateau-like induction profile. It is interesting to note.

(B) PRMP-Mut polyclonal cell lines quasi-linear induction profile. It is interesting to note that 50 μ M of phloretin allows for the highest induction rate for the PRMP-Mut cell-lines while it has a lower induction rate than a 30 μ M concentration for the PRMP cell lines.

4. 3. Engineering of the sender module

4. 3. 1. Phloretin synthesis

While I had engineered a working phloretin responsive promoter and generated phloretin inducible stable cell lines (receiver cell lines), it was still required to create a sender cell line that could synthesize phloretin as a communication molecule.

To achieve this, I engineered the first synthetic phenylpropanoid pathway for phloretin synthesis in mammalian cells (**Fig. 4-9**). Phloretin is a flavonoid which synthesis has been previously achieved in two enzymatic steps from the phloretic acid precursor (dihydro-p-coumaric acid) in *Escherichia coli* and *Saccharomyces cerevisiae*¹⁰⁹ for biotechnology purpose. First, the 4-coumarate:coenzyme A (CoA) ligase (4CL1) from the plant *Aradopsis thaliana* was used to convert dihydro-p-coumaric acid into dihydro-p-coumaroyl-CoA, which is then converted into phloretin by the chalcone synthase (CHS) from the plant *Hypericum androsaemum*.

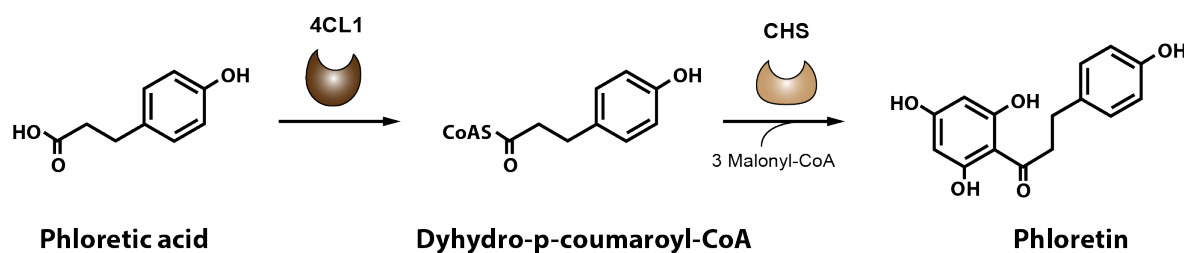


Figure 4-9. Synthetic phloretin pathway

4. 3. 2. Creation of a stable sender cell line

To achieve an optimal expression of these two enzymes in mammalian cells, I codon optimized their genetic sequences. I then created a polycistronic lentiviral vector expressing both enzymes together with a fluorescent marker (**Fig. 4-10**) so as to be able to FACS sort the cells based on the expression levels of the enzymes.

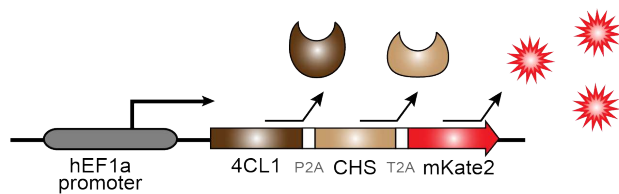


Figure 4-10. Phloretin biosynthesis polycistronic vector

The 2A tags I used have been shown to enable an efficient cleavage of the synthesized protein during translation, and should therefore result in a 1:1:1 ratio for the 3 proteins encoded by the single mRNA molecule.

This way, the intracellular concentration of 4CL1 and CHS enzymes should be directly correlated to the fluorescence level. One week after infection of HEK293FT cells with lentiviral particles containing the polycistronic construct, I FACS-sorted the cells with a rather high level of mKate2 fluorescence.

4. 4. Validation of functional cell-to-cell communication

To validate the proper functioning of our intercellular communication system, I co-cultured the stable sender cell lines together with the stable receiver cell lines in presence or in absence of the precursor.

Active metabolism by the sender cells of phloretic acid into dihydro-p-coumaroyl-CoA and then into phloretin would result in an increase of phloretin concentration in the media which would then trigger the activation of the reporter system within the receiver cells (**Fig. 4-11**).

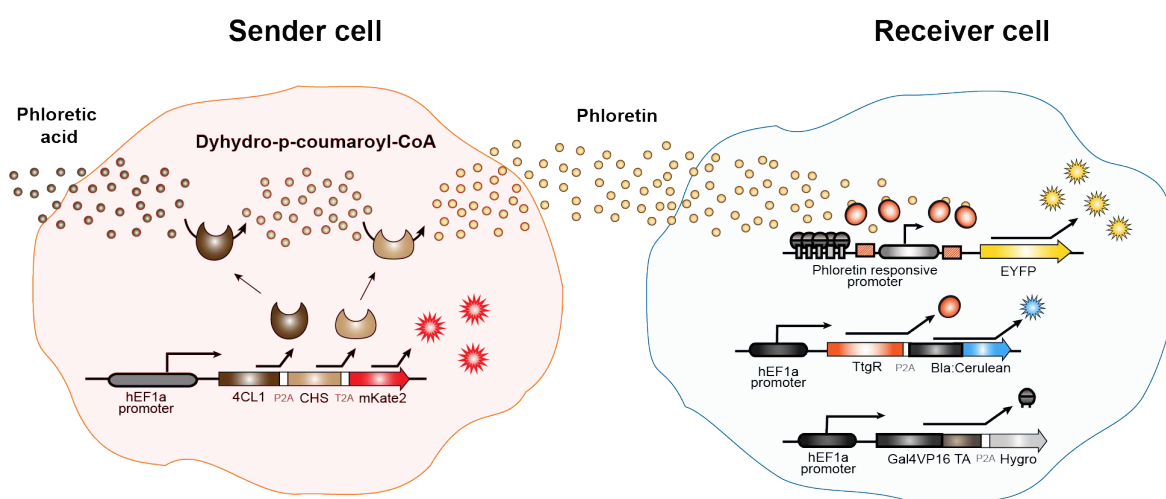


Figure 4-11. Overview of the complete phloretin cell-to-cell communication system.

To test the system, I decided to grow the cells in a 3:1 ratio (senders:receivers) and used 140uM of the precursor (phloretic acid) in order to ensure that a large amount of phloretin could be produced in a short period of time before the media had to be changed or cells split. To maximize the ratio of phloretin production vs. cellular growth, I seeded each well of 12 well plates with 50000 cells (25% confluence) which gave me about 72h before the wells would become confluent, which corresponds to the half-life time of phloretin in culture media (~70h, ref).

In the absence sender cells, EYFP levels did not increase in the receiver cells demonstrating that the precursor, even in excess, could not by itself activate the phloretin responsive promoters (**Fig. 4-12**).

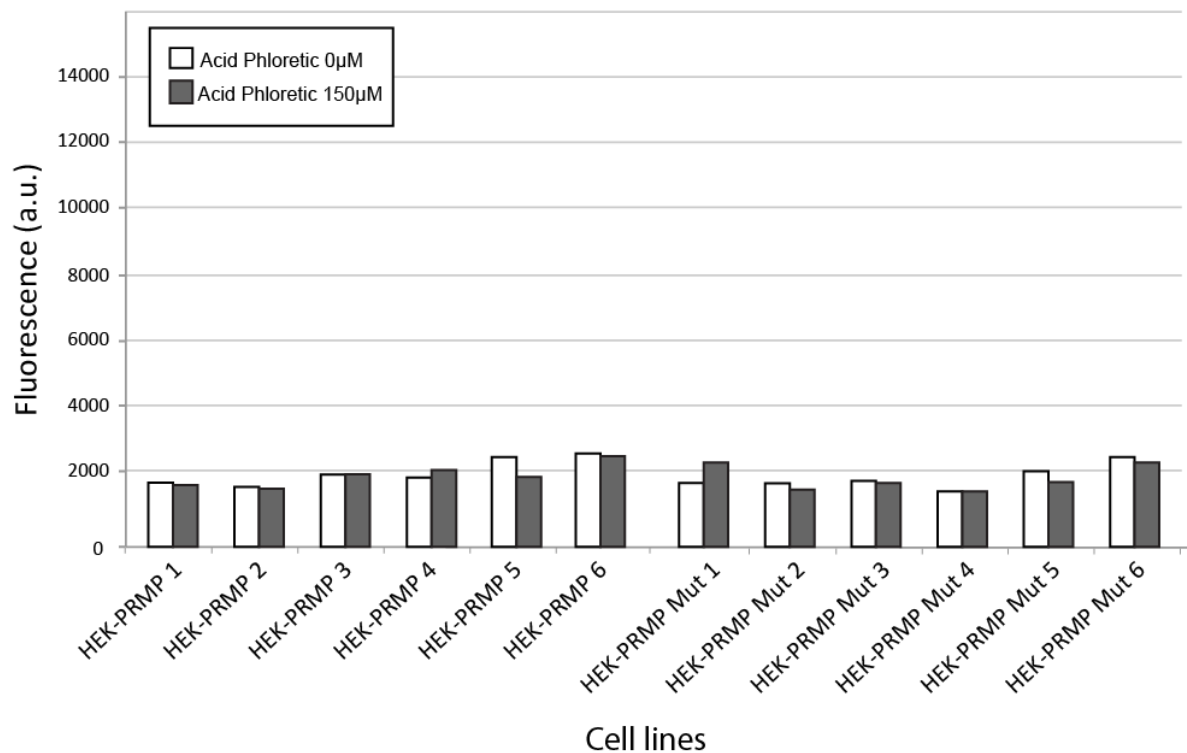


Figure 4-12. Activation profile of the reporter system without any phloretin

When sender cells were co-cultured with receiver cells, I could record a significant activation of the reporter system (**Fig. 4-13**) when I measured the EYFP levels in the receiver cells, which confirmed the synthesis of phloretin by the sender cells. By comparing the fold-change in transgene induction from this experiment to the calibration tests previously done, one could predict that the concentration of phloretin produced by the sender cells ranged between 20 and 30 µM.

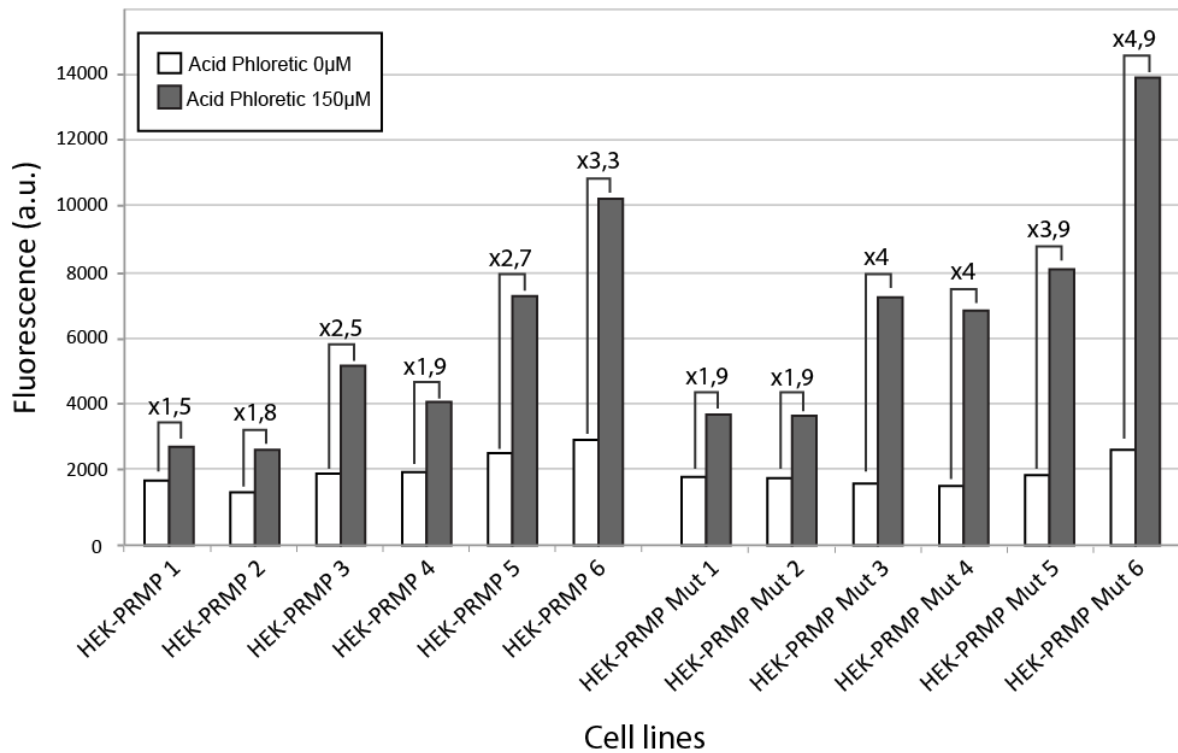


Figure 4-13. Activation profile of the reporter system when receiver cells are cultured with sender cells.

Altogether, these results demonstrate the proper functioning of the proposed intercellular-communication system. They are particularly encouraging given that the different modules have not yet been optimized, which means that better results could be attained. Notably, these results are based on polyclonal cell lines. It could be worth isolating a few clones from each of the 12 tested cell-lines and remake the co-culture experiments with each individual clone.

4. 5. Conclusion and discussion

In this study, I engineered the first synthetic intercellular communication system in mammalian cells that relies on a signaling molecule orthogonal to endogenous mammalian signaling systems and pathways. To achieve it, I first designed a new inducible system that uses phloretin, a small diffusible molecule, to trigger the transcription of a transgene in a dose-dependent manner, and then a biosynthetic pathway to produce this small molecule directly from mammalian cells.

As opposed to the PEACE repressor module created in 2012, the system I developed relies on the engineering of a synthetic phloretin responsive promoter that is turned off in the absence of phloretin and progressively turned on with an increasing concentration of phloretin in the media. The rationale behind the design was that in the absence of phloretin, binding of TtgR dimers to TtgR operators within the synthetic promoter sequence would prevent the transcription machinery to properly assemble. Dissociation of the TtgR dimers from their operator would then be triggered upon interaction with phloretin, which would therefore result in the transcription of the downstream gene thanks to the proper interaction between Gal4VP16 transactivators, transcription machinery and the CMV mini-promoter.

I demonstrated the functioning of this new inducible system, both when transiently transfected and stably integrated in mammalian cells. Fold-changes in induction ranged from 2x to more than 10x depending on the concentration of phloretin added to the media. Activation dynamics could be easily tuned by modifying the operators' sequence within the promoter.

Although these fold changes are not as high as what could be achieved with the extensively characterized and engineered TRE system, it could nonetheless become a convenient and useful complementary tool for many applications. Only a handful of reliable inducible systems have indeed been engineered up to date for use in mammalian cells. Moreover, most of the inducers currently used are not orthogonal to endogenous signaling pathways and can be only used for repression rather than for activation.

While I characterized some properties of the Phloretin Responsive Mammalian Promoter in this work, many other experiments could be done to explore a wider range of activation dynamics. One could for instance try to play both with the number of TtgR operators flanking the Mini-CMV and with the distance between these operators and the Mini-CMV or the Gal4 operators. This could significantly influence the leakage vs. activation ratio and therefore create a system which enables a higher fold-change in transgene expression level between the non-induced and the induced states. Such optimization could be particularly important for applications in which the

activated transgene would be a toxic gene. In this case, the PRMP leakage would have to be as little as possible when non-induced.

Another useful optimization of our proof of concept system would be the creation and integration of a single polycistronic vector to co-express the activator and the repressor in the receiver cells. In the current design, these components are expressed from different vectors. It is therefore rather hard to tune their expression levels given that they are stably integrated via lentiviruses. Co-expression from a single vector would allow for a better control of the ratio of these proteins. While the use of a P2A tag would result in a 1:1 ratio of activator vs. repressor protein concentration, the use of a different cleavage tag or different codon-optimized sequences could easily help attaining a different ratio of the expressed proteins and therefore more predictable activation dynamics.

One current limitation of this inducible system is that it relies on the competitive presence of both an activator and a repressor, which complexifies the tuning of its dynamics. A useful optimization would be to create a single polycistronic vector that.

In this study, the three components of the receiver module were stably integrated using lentiviruses, which prevented the control over locus insertion and over the number of integrated copies in the genome. This allowed me to create many different polyclonal stable cell lines with significantly different induction profiles. It could be interesting to isolate multiple clones from these different populations and characterize them separately in order to get even better fold-change in transgene activation for instance.

Once this receiver module had been validated, I had to create a sender module that could efficiently synthesize phloretin from an inert precursor molecule. Such a module would enable the transfer of information from a sender cell population to a receiver cell population containing the receiver module described above.

Towards this goal, I engineered and stably integrated a synthetic metabolic pathway using enzymes from two different plants that could lead to an efficient two-step synthesis of phloretin from phloretic acid, a small phenolic compound that is not produced by mammalian cells and orthogonal to endogenous mammalian signaling pathways.

I demonstrated the proper synthesis of phloretin from these receiver cells in a co-culture experiments in which I mixed these sender cells with the previously characterized receiver cells and supplemented the growth media with the precursor molecule.

In this experiment, I observed the activation of the reporter system in the receiver cells, corresponding to a concentration of approximately 30uM of phloretin.

To further understand and characterize the dynamics of this proof-of-concept intercellular communication system, a couple of experiments could be done.

First, one would have to quantitatively evaluate the exact production rate of phloretin by the sender cells. This could be achieved by measuring by HPLC the amount of phloretin produced over time depending on the concentration of precursor added to the media. Such experiment would help to determine the required ratio of sender vs. receiver cells to use to attain the desired activation of the reporter system in the receiver cells in a co-culture experiment.

To summarize, I have validated the *in-vitro* paracrine-like functioning of a new intercellular communication system in mammalian cells by engineering separate sender and receiver modules.

Besides deeper characterization and optimization of the modules themselves and of the complete system as a whole, there are a few developments that could really demonstrate the impact of such a technology.

A direct application of this cell-to-cell communication system based on the production of a diffusible small molecule would be the induction of a genetic program to create engineered organs. One could imagine to spatially arrange sender and receiver cells in a way such that upon induction with the precursor, sender cells trigger the activation of a differentiation gene in the receiver cells to create specific motifs depending on their distance from the sender cells and therefore responding to the gradient of phloretin diffusing through the tissue.

Another application I envision could be the creation of a synthetic tissue that could autonomously control its size, transforming the current paracrine-like system into an autocrine-like system. To achieve it, one could integrate both the sender and receiver system in the same cell population and connect the reporter system to a growth control mechanism. Such a population density-sensor would behave similarly to bacterial quorum-sensing: at first, when the size of the tissue is rather small, the synthesized phloretin would rapidly diffuse in the media and would therefore be immediately diluted in the overall growth culture. In this situation, the activation of the reporter system would be prevented. As the cells are dividing, they start to occupy more and more space, resulting in an increase of phloretin concentration in the media. At one point, when this concentration would exceed the activation threshold, the reporter system would be activated. If the induced transgene is directly linked to cellular growth arrest or apoptosis, one could imagine that its expression would either lead to the arrest of the tissue growth or to the fractional killing of the cell population, therefore regulating the size of the tissue.

For now, the functioning of this system relies on the supplementation with the precursor phloretic acid, which is not produced by mammalian cells. However, it has been demonstrated that phloretic acid could be synthesized from the essential amino acid L-tyrosine by another plant enzyme. If this system were to be used for therapeutic applications, one could envision integrating the corresponding gene into the sender module so that it could be directly be fed by an intake of L-tyrosine from the patient diet.

List of Supplementary Data (found in Annexes)

Supp Text 4-1	Nucleotide sequences of Phloretin promoters
Supp Text 4-2	Nucleotide sequences of phloretin production enzymes
Supp Text 4-3	Nucleotide sequence of TtgR repressor

4. 6. Materials and methods

Promoter and gene synthesis. PRMP, PRMP-Mut 4CL1 and CHS sequences were commercially synthesized. CHS and 4CL1 were codon-optimized to attain high expression levels in mammalian cells. The sequences are detailed in the Annexes.

Bacterial cell cultures. Liquid cultures of *E.coli* MG1655 were grown in LB Medium (Difco) at 37°C. When appropriate, antibiotics were added as follows: spectinomycin (100 µg/mL), ampicillin (100 µg/mL) and kanamycin (25 µg/mL). For blue/white screening, we used X-gal at a final concentration of 40 µg/mL.

Cell cultures, induction and transfections. HEK293FT cell lines were purchased from Invitrogen and maintained in Dulbecco's modified Eagle medium (DMEM, Cellgro) supplemented with 10% FBS (PAA), 0.045 g/mL penicillin/streptomycin and non-essential amino acids (HyClone) at 37 °C, 100% humidity and 5% CO₂. Phloretin and Phloretic Acid were purchased from Sigma (P7912 and H52406), dissolved in pure Ethanol, aliquoted and conserved at -20°C. HEK293FT transfections were carried out in 24-well plates using Attractene reagent (Qiagen), 200000 cells and 200-300 ng total DNA per well.

Lentiviral particle production and infection. We used Gateway ⁷⁷ cloning to construct the integration vectors. Our lentiviral Gateway destination vectors contain pFUGW^{18,25} (Addgene plasmid 14883) backbone and Gateway cassette (comprising chloramphenicol resistance and ccdB genes flanked by attR4 and attR2 recombination sites) followed by blasticidin or puromycin resistance markers expressed constitutively. LR reaction of the destination vectors with entry vectors carrying human elongation factor 1 alpha (hEF1a) promoter and either TtgR-P2A-Blasticidin-Cerulean or Gal4VP16-P2A-Hygromycin or 4CL1-P2A-CHS-T2A-mKate2 was used to create the following expression vectors: pLV-Hef1a_TtgR-P2A-Blasticidin-Cerulean, pLV-Hef1a_Gal4VP16-P2A-Hygromycin and pLV-Hef1a_4CL1-P2A-CHS-T2A-mKate2.

LR reaction of the destination vectors with entry vectors carrying the PRMP or PRMP-Mut and EYFP-P2A-Puromycin was used to create the following expression vectors: pLV_PRMP_EYFP-P2A-Puro and pLV_PRMP-Mut_EYFP-P2A-Puro.

For production of lentiviral particles $\sim 2 \times 10^6$ HEK293FT cells (Invitrogen) in 3 mL of DMEM complete media were plated into gelatin-coated 60 mm dishes (Corning Incorporated). Three hours later the $\sim 80\%$ confluent cells were co-transfected with 0.5 mg of the pLV-hEF1a-mKate2-P2A-Puromycin expression vector, 1.1 mg packaging plasmid pCMV-dR8.2 (Addgene plasmid 8455) and 0.55 mg envelope plasmid pCMV-VSV-G (Addgene plasmid 8454)^{78,101,102} using Attractene reagent (Qiagen) and following manufacturer's protocol. Media containing viral particles produced from transfected HEK293FT cells were harvested ~ 48 h post-transfection and filtered through a 0.45 mm syringe filter. 1.5 mL of the filtrate was added to $\sim 20\%$ confluent HEK293FT cells in 12-well plate seeded immediately before infection.

Flow cytometry and data analysis. Cells were analyzed with LSRFortessa flow cytometer, equipped with 405, 488 and 561 nm lasers (BD Biosciences). We collected 30000–50000 events, using a forward scatter threshold of 5000. Fluorescence data were acquired with the following cytometer settings: 488 nm laser and 530/30 nm bandpass filter for EYFP, 561-nm laser and 610/20 nm filter for mKate2, 405 laser and 525/50 filter for AmCyan and 405 nm laser, 450/50 filter for Pacific Blue. Data analysis was performed with FACSDiva software (BD Biosciences) and FlowJo (<http://www.flowjo.com/>). For histogram analysis, flow cytometry data in .FCS format were exported into text format using FCS Extract 1.02 software (E.F. Glynn, Stowers Institute for Medical Research) and analyzed in Microsoft Excel. For cell sorting, cells were collected directly into an 8-well micro-slide (Ibidi) by a FACS Aria cell sorter.

Microscope measurements and image processing. Fluorescence microscopy images of live cells were taken in glass-bottom dishes or 12-well plates using Zeiss Axiovert 200 microscope and Plan-Neofluar 10x/0.30 Ph1 objective. The imaging settings for the fluorophores were S430/25x (excitation) and S470/30m (emission) filters for AmCyan, and S565/25x (excitation) and S650/70m (emission) for mKate2. Data collection and processing were performed using AxioVision software (Zeiss).

Chapter 5 Hijacking lentiviruses to create a multi-channel protein-based intercellular communication system

Contents

- 5.1 Introduction**
 - 5.2 Engineering of a protein-based intercellular communication**
 - 5.2.1 Design
 - 5.2.2 Validation of sender and receiver modules proper functioning
 - 5.3 From single-channel to multi-channel intercellular communication**
 - 5.3.1 Design
 - 5.3.2 Validation of the AND gate functioning
 - 5.4 Conclusion and discussion**
 - 5.5 Materials and methods**
-

5. 1. Introduction

Viruses have evolved to efficiently transfer their genetic material from one cell to another in order to reproduce by hijacking their host machinery. For a safe and efficient transfer, the viral genome (either DNA or RNA) is packaged in a protective shell of core proteins (**Figure 5-1**). An envelope, composed of lipids and proteins, surrounds the capsid to help penetrate host cells.

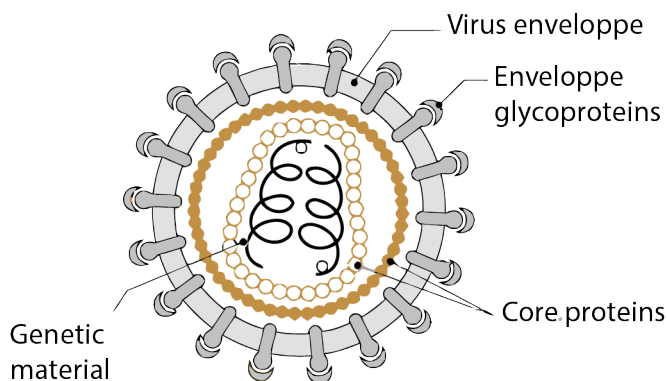


Figure 5-1. Schematic of a virion.

Viral particles (known as virions) consist of two or three parts: i) the genetic material made from either DNA or RNA; ii) a protein coat that protects these genes; and in some cases iii) an envelope of lipids coated with glycoproteins that surrounds the capsid.

In the last decade, scientists have managed to hijack these viral particles (virions) for mainly two applications: vaccines and more recently protein transfer. By eliminating the packaging signal from the viral genomes, they demonstrated that they could create Virus Like Particles that resemble virions but do not contain any genetic material, and therefore cannot replicate. This way, non replicative VLPs could be used as vaccines, as they contain repetitive high density displays of viral surface proteins which present conformational viral epitopes that can therefore elicit strong T cell and B cell immune responses.

Two recent studies demonstrated that VLPs could also be used to deliver macromolecules of interest into the cytosol and nucleus of living cells. More specifically, Voelke et al and Kaczmarczyk et al ¹¹² engineered VLPs derived from retroviruses to deliver the Flp and Cre recombinases to mammalian cells. To achieve that, Voelke et al intercalated the recombinase gene between viral genes while Kaczmarczyk et al fused it directly to the C-ter of the *gag* gene, deleting all other genes from the viral genome.

The *gag* gene encodes the major structural proteins of the virion (**Figure 5-2**): matrix (MA), p12, capsid (CA) and nucleocapsid (NC); which are transcribed and translated as a single chain polyprotein. It is followed by the *pol* gene and the *env* gene. The *pol* gene encodes the viral protease (PR – which cleaves the single chain polyproteins during maturation of the virion), the reverse-transcriptase (RT – which reverse-transcript the RNA genome of the virus into DNA

after infection) and the integrase (IN – which stably integrate the viral DNA into the infected cell chromosome). Finally, the mature product of the env gene is the viral spike protein, which has two main parts: the surface protein (SU) and the transmembranar protein (TM). The SU protein domain determines the tropism of the virus because it is responsible for the receptor-binding function of the virus. The SU domain therefore determines the specificity of the virus for a single receptor molecule.

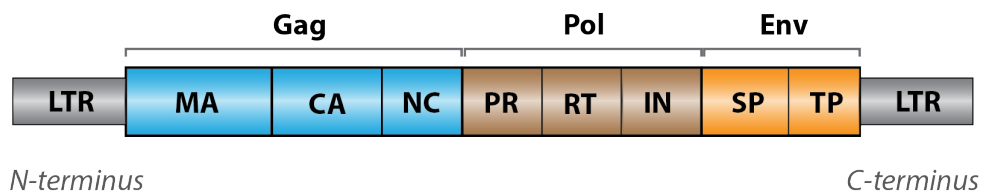


Figure 5-2. Simplified genome of retroviruses.

Retroviral genomes are composed of 3 genes. First, the *gag* gene encodes for the N-terminal matrix protein (MA), the capsid protein (CA), the p12 protein and the nucleocapsid protein (NC). The *pol* gene codes for the viral protease (PR), the reverse transcriptase (RT) and the integrase (IN). Most of the retrovirus also have the *env* gene which codes for the envelope proteins. Long Terminal Repeats (LTR) flank the genome and are used in the process of integration.

For VLP formation, for which it had been previously shown that Gag polyprotein is the only retroviral protein required, thousands of Gag polyproteins accumulate at the cytoplasmic side of the cell membrane through myristoylation of the N-terminal matrix protein. By fusing a protein of interest to the C-terminus of the *gag* gene of the Rous Sarcoma Virus (RSV), Kaczmarczyk et al. demonstrated that the fusion protein not only retained its packaging ability, but that it could be efficiently released into the cytosol of the infected cells.

VLPs then become a useful mechanism to transfer highly concentrated proteins from one cell to another, protected by the viral particle during the transfer. I therefore decided to extend this work in order to create a modular and orthogonal intercellular communication mechanism, harnessing the potential of VLPs to transfer proteins of interest from a sender cell population to trigger an output in a distant receiver cell population (**Fig. 5-3**).

I first engineered an inducible sender system that could be triggered to package, into VLPs, either the Cre recombinase or the B3 recombinase fused to the Rous Sarcome Virus *gag* polyprotein from the Rous Sarcoma Virus (same as the one used by Kaczmarczyk et al). I then created reporter cell lines to validate the efficient intracellular delivery of these fusion proteins. I demonstrated that different proteins of interests, here recombinases, could be actively transferred from sender cells to receiver cells after being encapsulated into engineered VLPs

(Fig. 5-3A). Finally, I created a reporter cell line to demonstrate that this system could pave the way for complex intercellular computing thanks to its modularity and the possibility to send many different and orthogonal messages in between population of cells (Fig. 5-3B).

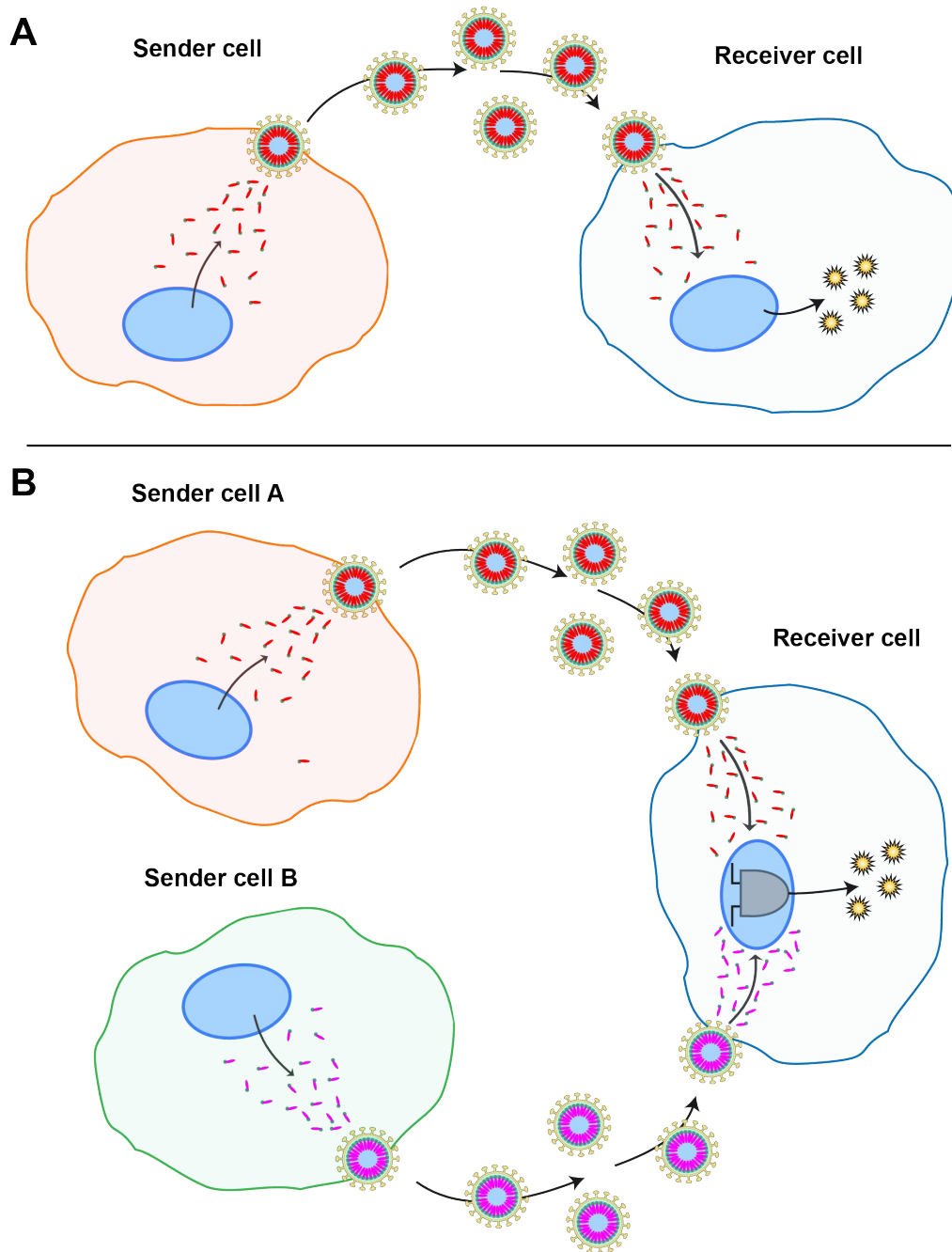


Figure 5-3. Schematic of engineered VLP based intercellular communication.

(A) Gag-fusion proteins are concentrated into budding engineered VLPs which then diffuse in the media. When the VLPs fuse with the receiver cell, the fusion proteins are delivered intracellularly and can then activate a receiver circuit. (B) Such a system can be multiplexed by sending different and orthogonal signals to a receiver cell that can then integrate both signals.

5. 2. Engineering of a protein-based intercellular communication

5. 2. 1. Design

To create a reporter system that could easily reflect the efficient delivery of a protein of interest into the cytosol of receiver cells, I chose to build a excision based system that could be trigger the constitutive expression of a fluorescent protein upon delivery of the protein of interest. In such design, the protein of interest would therefore be a site-specific recombinase. Based on Kaczmarczyk et al study, I first assembled the synthetic genes to express in the sender cells two different *gag* proteins fused on their C-terminal end to Cre or B3 recombinases (**Fig. 5-4**).

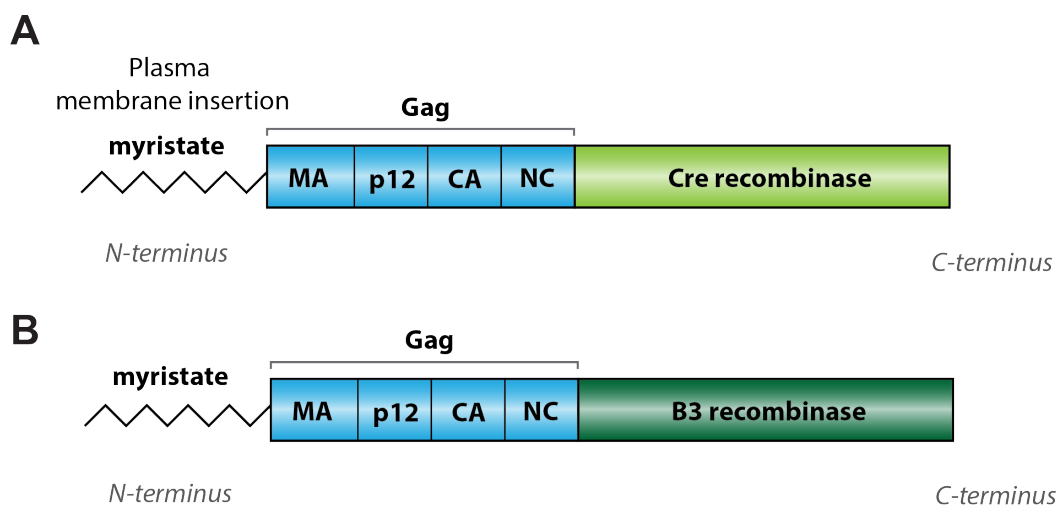


Figure 5-4. Schematic of RSV Gag-fusion proteins.

A recombinase is fused to the C-terminus of the Gag protein. Engineered retroviral particles assemble by associating a few thousand Gag-precursor polyproteins, which accumulate at the cytoplasmic side of the cell membrane through myristoylation of the N-terminal matrix (MA) protein.

I also assembled a vector to express the VSV-G envelope protein (**Fig. 5-5**) so as to pseudotype the lentiviral vector¹¹³, conveying the ability to transduce a broad range of mammalian cells. I placed this gene under the control of the TRE-tight (TRET) inducible promoter so as to be able to regulate the assembly of functional VLPs in the sender cells. To complete the sender module, I finally assembled a vector to constitutively express the rtTA3 protein required for TRE-tight activation in presence of doxycycline.

In order to produce as many VLPs as possible from the sender cell population, I chose to stably integrate these genes with lentiviruses (multiple integrations) rather than with the previously developed Landing pad method (single integration).

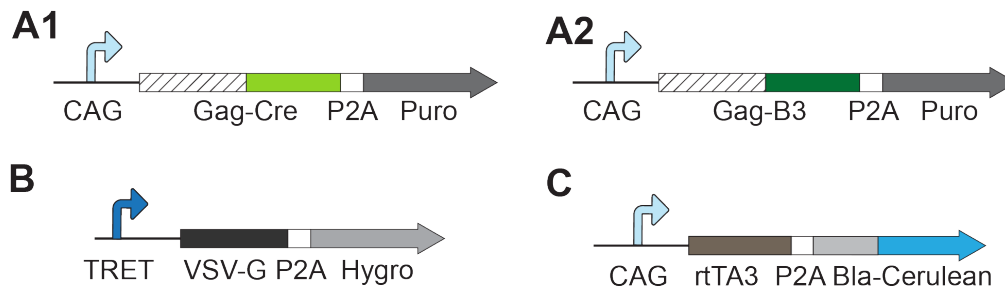


Figure 5-5. Vectors used for the sender module.

(A, B) These vectors allow the constitutive expression of the Gag-recombinase fusion proteins. (C) This vector will allow the inducible expression of the VSV-G envelope protein to enable the fusion of the VLP with mammalian cells. (D) This vector is required to activate the expression of the VSV-G protein which is under the control of the TRE-tight promoter. All these vectors also allow the expression of a resistance gene to enable the selection of stable integrations to create stable sender cell lines.

For the receiver system, I created two different synthetic circuits: one responsive to Cre recombinase, the other to B3 recombinase (**Fig. 5-6**). These circuits were stably integrated in a single copy the landing pad of Hek293FT chassis cell-line. They consist of non-fluorescent gene placed after a constitutive CAG promoter, and flanked by either lox or att B3 recombination sites. When recombination happens between these two sites oriented in the same direction, the gene is excised and places a promoterless mKate2 gene in frame with the CAG promoter, resulting in permanent expression of the fluorescent protein, even after dilution of the recombinase.

I decided to integrate these receiver circuits in a single copy using the Landing Pad platform described in the second chapter of the thesis to assess the sensitivity of the receiver module and to have a homogeneous response over the whole receiver population.

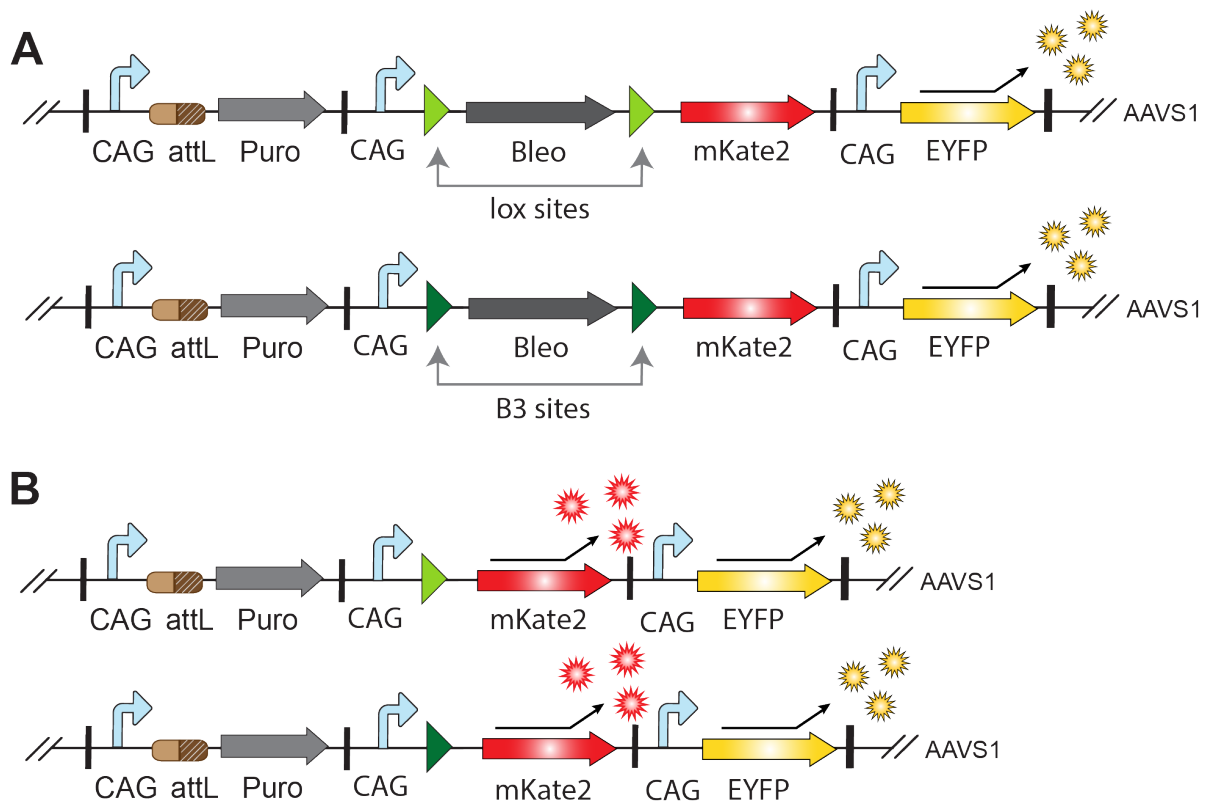


Figure 5-6. Details of the receiver module.

(A) These circuits integrated in the landing pad result in constitutive EYFP expression before excision of the Bleo gene by site-specific recombinase. (B) After excision of the Bleo genes by either Cre or B3 recombinase, mKate2 fluorescent protein is constitutively expressed.

5. 2. 2. Validation of sender and receiver modules proper functioning

To demonstrate that the activation of the sender module (both with Cre and B3 recombinase) could result in the transfer of Cre or B3 recombinases from budding VLPs to the receiver cells, we co-transfected Hek293 FT cells with the 3 sender systems vectors: pLV_CAG_RSV-GAG-Cre_P2A_Puro or pLV_CAG_RSV-Gag-B3_P2A_Puro, pLV_CAG_rtTA3_P2A_Bla-Cerulean and pLV_TREt_VSV-G_P2A_Puro. For each RSV-Gag fusion protein system, the transfected cells were split in two batches. One batch served as a negative control for which no doxycycline was added. The culture media of the other batch was supplemented with 2ug/mL of doxycycline. Sender cells were grown for 48h without changing the media. The supernatant of each batch was then used to replace the media of receiver cells cultures. Finally, fresh media was added to the receiver cells 24h later.

To detect whether or not excision events happened in the receiver cells circuits, I measured fluorescence levels 72h after the initial media replacement. Such delay should be

enough to allow the fusion of VLPs with the cell membranes, the delivery of their contents within the cell cytosol, the site-specific recombination events within the circuits, and the reach of high enough fluorescence levels from the expression of a single copy of the mKate2 gene.

When VSV-G expression was not induced in the sender cells, supplementing the media of the receiver cells by the supernatant from the sender cells did not result in any mKate2 expression within the sender cell population (**Fig. 5-7**). However, when VSV-G expression was induced in the sender cells, I could record that approximately 30 to 40% of the receiver cells had switched on the expression of mKate2 gene, for both Cre and B3 systems.

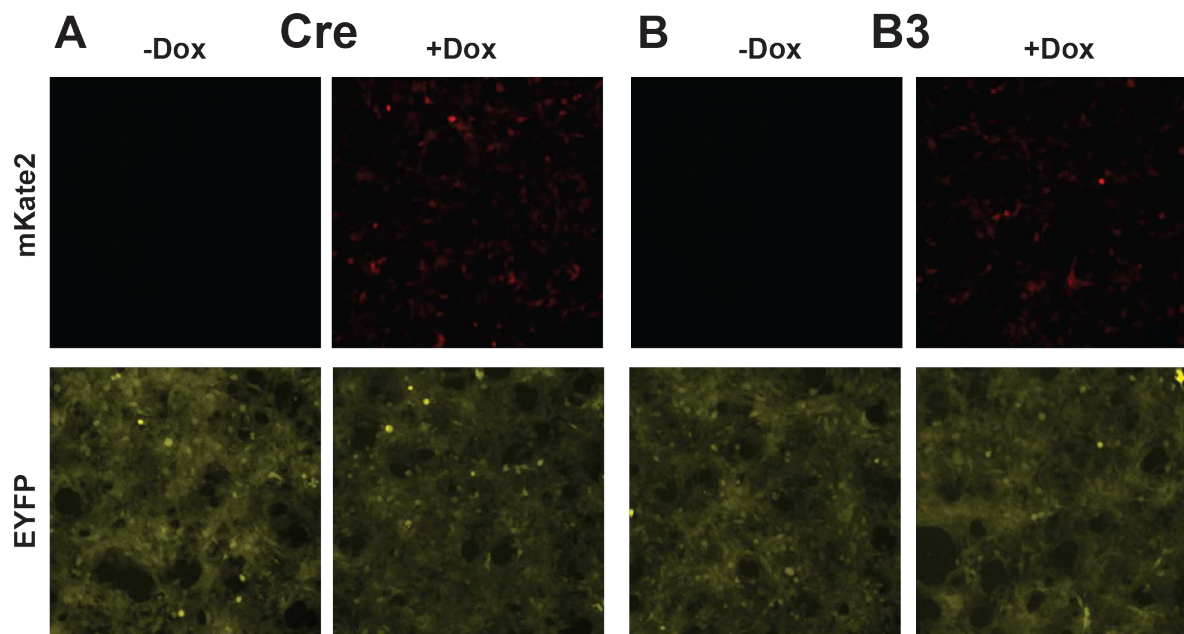


Figure 5-7. Fluorescence microscopy pictures of receiver cells 72h after sender cells supernatant addition.

(A) These receiver cells contain the Cre responsive circuit. mKate2 expression when dox has been added to the sender cells demonstrates the proper functioning of both the sender module and the receiver module. (B) The proper functioning of both sender and receiver modules for the B3 communication system is also confirmed.

5. 3. From single-channel to multichannel intercellular communication

5. 3. 1. Design

Now that the ability to transfer of different proteins from a population of senders cells to a population of receiver cells was confirmed, I wanted to demonstrate that these two

communication channels could be combined in order to perform intercellular computing based on different communication signals.

Based on the orthogonality of Cre and B3 recombinases, I created and stably integrated a simple AND gate circuit (**Fig. 5-8**) in the receiver cells. Such a circuit would result in the expression of a fluorescent protein only and only if both recombinases are delivered within the receiver cells. It consists of a constitutive promoter followed by two non-fluorescent genes, respectively flanked by lox sites and att B3 sites (oriented in the same direction). A promoterless EYFP fluorescent gene is located after these two genes.

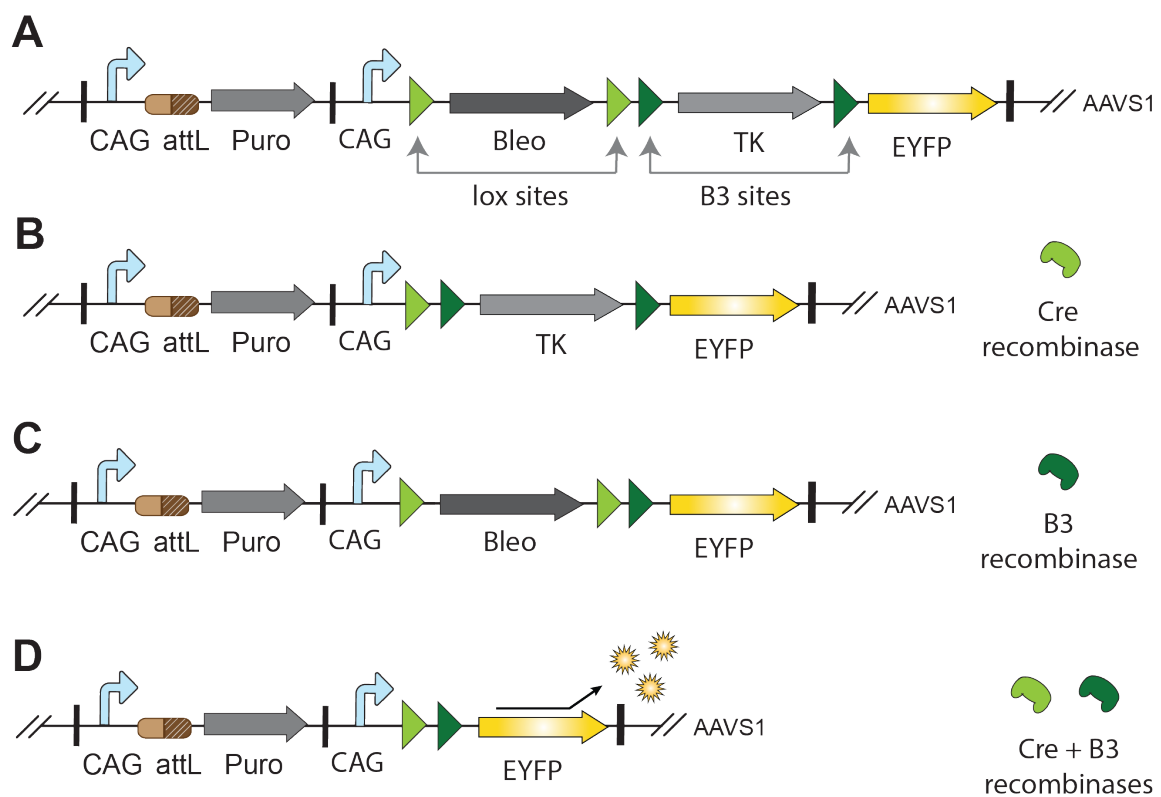


Figure 8. Design and functioning of the AND gate circuit in the receiver cells.

(A) In the initial state, the AND gate is composed of the constitutive CAG promoter followed by two non fluorescent genes flanked by recombination sites and followed by the EYFP gene. In this state, EYFP is not expressed. (B) Excision of the Bleo gene by Cre recombinase only places the TK gene in frame with the CAG promoter but EYFP is still not expressed. (C) Excision of the TK by B3 recombinase only puts the EYFP closer to the EYFP promoter but its expression is still blocked by the Bleo gene. (D) Bleo and TK are both excised only when B3 and Cre are present, which result in expression of the EYFP gene.

5. 3. 2. Validation of the AND gate functioning

Once I assembled the AND gate circuit, I stably integrated it in the Hek293FT chassis cell line. To test whether it could support multichannel communication based on different recombinases as signaling molecules, I co-transfected the receiver cell lines with either the pLV_RSV-Gag-Cre vector, the pLV_RSV-Gag-B3 vector or both vectors. I then measured the EYFP fluorescence 72h after transfection for the different conditions (**Fig. 5-9**). When only Cre or B3 fusion proteins were expressed, no EYFP expression could be observed, confirming the orthogonality of these recombinases between their respective second sites. The positive EYFP output was observed only when both Cre and B3 fusion proteins were expressed.

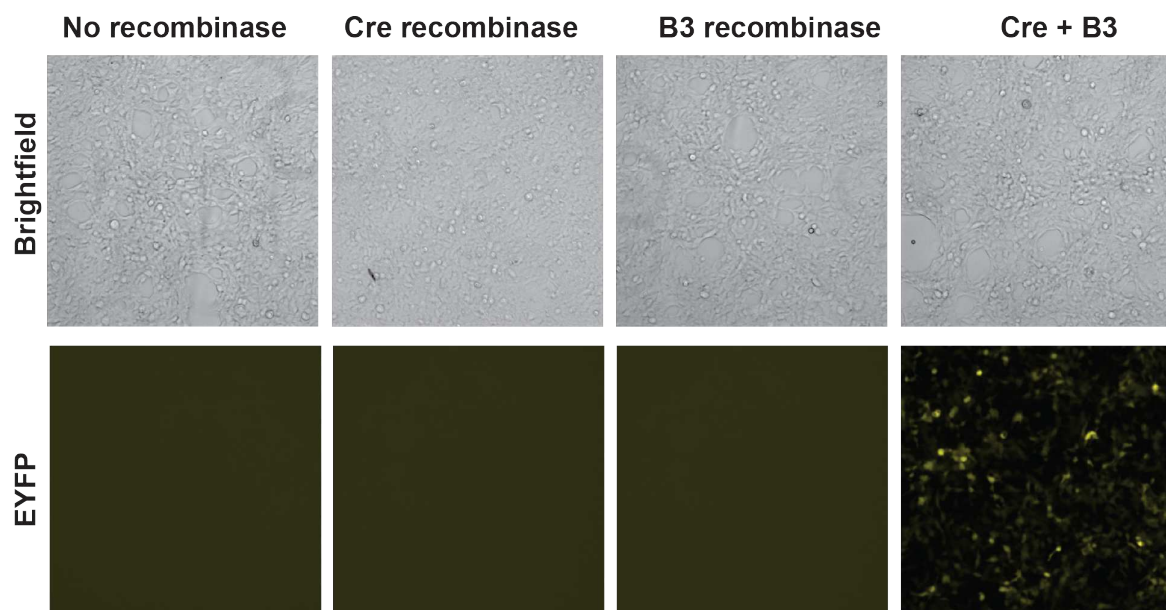


Figure 5-9. Fluorescence microscopy pictures of AND gate testing in the receiver cells. EYFP was expressed only when pLV-RSV-Gag-B3 and pLV-RSV-Gag-Cre were co-transfected in the receiver cells.

Now that the proper functioning of the AND gate from the receiver cell line was confirmed, the next step would be to repeat this experiment by transfecting all the vectors required to express VLPs into a two different sender cell lines this time (one for Cre and the other for B3), collect the supernatant and add it to the culture media of the receiver line.

5. 4. Conclusion and discussion

In this study, I extended the proof of concept that VLPs could be engineered to create a protein-based intercellular communication system and potentially be used to achieve intercellular computation. Based on previous work by Kaczmarczyk et al., I created an inducible system to enable VLPs expression using a Rous Sarcoma Virus gag protein fused to either Cre or B3 recombinase. I demonstrated that they could be produced from a sender cell population after transfections of the vectors followed by induction of their expression, and that they could deliver the gag-Cre or gag-B3 proteins into a population of receiver cells that carried a responsive genetic circuit integrated in a single copy in their chromosomes. These fusion proteins were able to genomically rearrange the circuit to turn ON the production of a fluorescent protein within the receiver cell population.

To move this proof of concept to the next level, it would be critical to create stable sender cell lines with the gag-fusion and VSV-G vectors stably integrated in their genome. I believe that it would be necessary to isolate different clonal cell populations carrying different ratios of gag-fusion vs. VSV-G vectors to find the ideal ratio for a high production of engineered VLPs. However, it would also be important to have the ability to control VSV-G expression levels with doxycycline so as not to reach to high levels in order not to kill the sender cells. Indeed, pseudo-typing protein VSV-G can induce the fusion of cell membranes.

Once a good stable sender cell line would have been isolated, we could further characterize the system by growing sender and receiver cells in the same media but spatially separated so as to study the dynamics of the overall system. This could for instance help us characterize the diffusion dynamics of the VLPs and the amount of sender cells required to activate the receiver cell population. Given that a single VLP is supposed to carry thousands of gag-fusion proteins and that one sender cell is also supposed to produce a significant number of VLPs, one could imagine that only a very little amount of sender cells would be necessary to activate a large number of receiver cells over time.

Building on the modularity of this protein based intercellular communication system, I wanted to create the first synthetic and intercellular computation system using different and orthogonal signaling molecules. To this end, I designed and stably integrated an AND gate circuit in a receiver cell population, which would be activated when both B3 and Cre recombinases are delivered to the receiver cells. Unfortunately, I did not have the time to generate a new batch of Gag-B3 and Gag-Cre VLPs to be applied on the clonal receiver population. Even though I could not achieve my goal, I am fairly confident that the system as a whole would work since each of

the modules has been previously validated (Gag-fusion transfer to a receiver cell population & functioning of the AND gate when both recombinases are present in the receiver cells).

To go beyond this proof of concept, it would be extremely interesting to extend the capabilities of the system by broadening the repertoire of proteins used as signaling molecules so as to create more complex logic circuits. While I have only tested the system with two recombinases, one could create Gag-fusion proteins with the handful of orthogonal recombinases that are currently characterized. One could also try to build a system that would use the transactivator rtTA3 or the KRAB silencing protein-domain as a signaling molecule to transiently activate or repress the expression of one or multiple transgenes in a population of receiver cells. With such tools in hand, a wide range of synthetic intercellular network topologies could then be designed and characterized: oscillators, feedback loops, feed-forward motif, etc. A sender population could for instance secrete VLPs that would release the rtTA3 transactivator in a receiver population, which will result in the secretion of VLPs from this receiver population carrying a Gag-B3 recombinase. These VLPs could then induce the constitutive expression of a specific transgene in a second receiver cell population. By combining the mMoClo system, the Landing Pad integration platform and these engineered VLPs, it could become convenient to create such cell populations and study complex networks.

On a longer-term, one could envision to use the system described in this chapter to integrate this long-range intercellular protein-based intercellular communication within a human body. This system may represent a possibility to allow distant synthetic tissues to communicate with each other. Advances in viral pseudotyping have demonstrated that viral capsid could be precisely targeted to specific cells depending on their cellular receptors. This way, one could envision creating a VLP-producing tissue that would enable the delivery of proteins specifically to another tissue of choice.

List of Supplementary Data (found in Annexes)

Supp Text 5-1 Nucleotide sequences of VLP proteins

5. 5. Materials and methods

Golden Gate reactions. For all Golden Gate assembly reactions we used: 0.4 μ l of TypeIIS enzyme (either BsaI from NEB, or BpiI from Fermentas), 0.2 μ l of T4 Ligase HC + 1 μ l of T4 Ligase HC buffer (Promega), 1 μ l of 10xBSA (NEB), 40 fmol for all vectors used in the reaction, ddH₂O up to a final total volume of 10 μ l. The thermocycler program used for all assemblies included: 1 step of 15 min at 37°C; then 50 cycles of [2 min at 37°C followed by 5 min at 16°C]; 1 step of 15 min at 37°C, 1 step of 5 min at 50°C and 1 final step of 5 min at 80°C.

Expression vectors assembly. We used Gateway[™] cloning to construct the expression vectors. Our lentiviral Gateway destination vectors contain pFUGW²⁵ (Addgene plasmid 14883) backbone and Gateway cassette (comprising chloramphenicol resistance and ccdB genes flanked by attR4 and attR2 recombination sites) followed by blasticidin or puromycin resistance markers expressed constitutively. LR reaction of the destination vectors with entry vectors carrying CAG promoter and either rtTA3-P2A-Blasticidin-Cerulean or RSV-Gag-Cre_P2A_Puro or RSV-Gag-B3_P2A_Puro was used to create the following expression vectors: pLV-CAG-rtTA3-P2A-Blasticidin-Cerulean, pLV-CAG-RSV-Gag-Cre_P2A_Puro and pLV-CAG-RSV-Gag-B3_P2A_Puro.

LR reaction of the destination vectors with entry vectors carrying the TRE-tight promoter and VSV_G-P2A-Hygromycin was used to create the following expression vector: pLV_TRE-tight_VSV_G-P2A-Hygromycin.

Bacterial cell cultures. Liquid cultures of *E.coli* MG1655 were grown in LB Medium (Difco) at 37°C. When appropriate, antibiotics were added as follows: spectinomycin (100 μ g/mL), ampicillin (100 μ g/mL) and kanamycin (25 μ g/mL). For blue/white screening, we used X-gal at a final concentration of 40 μ g/mL.

Cell cultures, induction and transfections. HEK293FT cell lines were purchased from Invitrogen and maintained in Dulbecco's modified Eagle medium (DMEM, Cellgro) supplemented with 10% FBS (PAA), 0.045 g/mL penicillin/streptomycin and non-essential amino acids (HyClone) at 37 °C, 100% humidity and 5% CO₂. For induction of the TRE promoter, 2 μ g/mL Doxycycline was used. HEK293FT transfections were carried out in 24-well plates using Attractene reagent (Qiagen), 200000 cells and 200-300 ng total DNA per well. Media containing engineered VLPs produced from transfected HEK293FT cells were harvested ~ 48 h post-transfection and filtered through a 0.45 μ m syringe filter. The whole filtrate was added to ~

20% confluent HEK293FT cells in 12-well plate seeded immediately before infection. After 70 h, fresh media was added to the cells.

BxB1-mediated integration of circuits. To integrate circuits into the landing pad of HEK293FT chassis cell-line, we co-transfected 150 ng of the appropriate multi-gene integration vector with 150 ng of BxB1 integrase expression vector using attractene (Qiagen) in 24-well format. 72 hours post transfection cells were transferred to 6-well plates and culture medium supplemented with 1 μ g/mL Puromycin (InvivoGen). Selection was maintained for 5 days. Cells were typically assayed 14 days post transfection.

Microscope measurements and image processing. Fluorescence microscopy images of live cells were taken in glass-bottom dishes or 12-well plates using Zeiss Axiovert 200 microscope and Plan-Neofluar 10x/0.30 Ph1 objective. The imaging settings for the fluorophores were S430/25x (excitation) and S470/30m (emission) filters for EYFP, and S565/25x (excitation) and S650/70m (emission) for mKate2. Data collection and processing were performed using AxioVision software (Zeiss).

Chapter 6 General conclusion

Contents

6.1 Summary

6.2 Perspectives

6. 1. Summary

In this PhD thesis, I proposed and validated a number of concepts and approaches to move mammalian synthetic forward. More specifically, I focused on the development of new platforms to streamline the assembly, characterization and long term study of large genetic circuits in mammalian cells; and the engineering of synthetic orthogonal intercellular communication systems.

To implement therapeutically relevant complex functions in mammalian cells, a fine-tuned, precise and robust behavior of the synthetic circuits is required over time. There are however major limitations that prevent such achievement: whether it is from a technical point of view or from context-dependency effects inherent to mammalian biology. To address these problems, I developed a framework for modular and combinatorial assembly of functional (multi)gene expression vectors and their efficient and specific targeted integration into a well-defined chromosomal context in mammalian cells. I demonstrated the potential of this framework by assembling and integrating different functional mammalian regulatory networks including the largest gene circuit built and chromosomally integrated to date (6 transcription units, 27kb) encoding an inducible memory device. Using a library of 18 different circuits as a proof of concept, I also demonstrated that the framework enables one-pot/single-flask chromosomal integration and parallel screening of circuit libraries. As a whole, this rapid and powerful prototyping platform provides an efficient and reliable strategy for correct and sustained execution of complex genetic programs in mammalian cells and could be well suited for comparative studies of genetic regulatory elements, genes and multi-gene circuits as well as facile development of libraries of isogenic engineered cell lines.

One of the bottlenecks of this method based on site-specific recombinases is that we have to preintegrate recombination sites in mammalian chromosomes in order to target large circuits integration to the locus of choice. It would therefore become much more convenient and powerful to engineer these recombinases to recognize endogenous sites in order to bypass the preintegration step. Our current ability to engineer these enzymes is however still extremely limited and I therefore decided to focus on the development of a platform to identify new large serine recombinases and extend the toolbox currently available. Based on an automated phage genome analysis, the script I created can efficiently identify putative recombinases that have specific amino acid motifs corresponding to large serine recombinases. After analyzing more than 400 mycobacteriophages genomes, I isolated 24 new putative recombinases, different from the ones previously characterized. I then created a framework that allowed me to identify their respective recombination sites within the mycobacterium genomes and then to test each recombinases activity in *E. coli* and mammalian cells. To further extend the use of site-specific

recombinases in mammalian cells, I also demonstrated that libraries of orthogonal sites could be created for large serine recombinases. As a whole, these contributions may enable to streamline the discovery of new recombinases and new recombination sites from mycobacteriophage genomes, and therefore extend the toolbox currently available for site-specific recombination based technologies.

Another current limitation for the development of complex functions in mammalian cells is the lack of synthetic intercellular communication systems that are orthogonal to endogenous signaling systems. I therefore spent quite some time designing two novel communication strategies with completely different messaging dynamics. The first one that I engineered is based on the production of a small metabolite, phloretin, from a sender cell population, which can activate transgene expression in the receiver cell population in a dose-dependent manner. On the one hand, I engineered the biosynthesis of phloretin by co-expressing 3 exogenous plant genes to create a synthetic enzymatic pathway. On the other hand, I created a synthetic sensor module based on a synthetic promoter carrying operator sequences recognized by the TtgR repressor, which binding is phloretin-dependent. To validate the design of the new communication system, I first demonstrated that commercially bought phloretin could induce significant transgene activation in receiver cell lines (10x) and I then performed co-culture experiments with both sender and receiver cell lines. It was particularly amazing to see that without fine-tuning of the system, i.e. working only with polyclonal cell lines created with lentiviruses, co-culture experiments resulted in 5x transgene activation in the receiver cell populations. While these results are satisfactory enough for a proof of concept of this completely new intercellular communication system, there is room for many improvements that I did not have the time to do: explore the phloretin production dynamics, isolate monoclonal cell lines, change the number of operator sequences within the synthetic promoter etc.... I also had in mind the use of this system to create a quorum sensing like mechanism, by integrating both the sender and the receiver modules into the same cell population. This would however require a very fine-tuning in order to enable the bistability of the system.

The second intercellular communication system I developed relies on a completely different signaling mechanism: protein transfer via virus like particles. To create this system, I extended previous results showing the efficient packaging of proteins into VLPS and their delivery into the cytosol of receiver cells. I first created inducible cell lines that could secrete VLPs containing two different functional recombinases (Cre and B3), and then the corresponding stable receiver cell lines. I demonstrated that induction of VLP production led to the secretion of VLPs that could efficiently deliver the recombinases into the cytosol of the stable receiver cell lines, and result in the activation of a transgene following excision events triggered

by the transferred recombinases. I then proposed to use this approach to create a synthetic intercellular computation system that could use different proteins as inputs to be integrated by a receiver cell line thanks to an AND gate circuit based on two separate excision modules. The proof of concept being done, one could envision pushing such a strategy forward by create more complex intercellular computation systems using different proteins as signals such as transcription activators or repressors, engineered nucleases, kinases, etc....

6. 2. Perspectives

"The way you deal with mammalian cells now is by luck,"

Pamela Silver, system biologist at Harvard Medical School in Boston.

When I started my PhD, there were only very few labs working on mammalian synthetic biology. While complex circuits had already been built in simpler unicellular organisms, it seemed feasible to transpose these systems to mammalian cells. However, given the non-replicability of bacterial plasmids in mammalian cells, circuits could only allow transient expression of synthetic genes and therefore prevented long-term implementation of new functionalities.

It became clear to us that we would need to create technologies and platforms to enable stable integration and characterization of large circuits. When I started to develop an efficient integration method, there were actually no studies demonstrating stable integration of functional genetic circuits carrying more than two independent transcription units. While robust and long-term transgene expression was possible thanks to lentivirus technology (which was severely limiting the size of the insert), such approach prevented the creation of independent transcription units as no polyA signals could be placed within the insert. In parallel, new genome engineering technology started to emerge and allow for targeted integration of DNA fragments, though with limits on insert size. However, it allowed unveiling the significant complexity of transgene expression once integrated in mammalian cells. Context-dependency effects appeared to be numerous and we therefore had to take them into consideration when building our circuits. Based on some previous studies about the role and influence of insulating sequences on the limitation of these effects, we proposed a specific architecture for synthetic mammalian transcription units and circuits. Although we have validated the approach by assembling, integrating and showing long term functioning of a large synthetic circuit, there is still a long way until we could accurately predict the expression profile of the different integrated

transgenes. We have for instance seen that transgene expression can be significantly altered when shuffling the order of the transcription units or when using a specific combination of promoters in the different units, which demonstrates that there are still some context-effects that have yet not been understood and harnessed. I therefore believe that the modular assembly and integration platform I developed may be of great help to test and try all possible combinations of circuits in order to elucidate what are the mechanisms affecting transgene expression. To my mind, such studies will be necessary to move integrated synthetic circuits to the clinic and therefore have them used for real therapeutic applications that will rely on finely tuned transgene expression levels.

A potential solution could however be attained by physically separating the different modules of such circuits on the chromosome. It might therefore be particularly useful to create cell lines with multiple landing pads located in different places throughout the genome. These landing pads should however be orthogonal to each other and could for instance carry orthogonal recombination sites. For this purpose, one could use the large serine recombinases that I have discovered and characterized. Such a strategy would also help understanding the specificity of each chromosomal loci and their effect on transgene expression.

To my mind, we are still in a period where the potential of mammalian synthetic biology resides in other applications than engineered-cells based therapies, such as antibody production for instance. In this case, while using a single landing pad can enable long term production of antibody proteins, the overall expression level will never reach the levels attained by lentivirus-based integrations, which enable numerous copies to be integrated at the same time. Using multiple landing pad however could be a very interesting strategy given that we could potentially create a chassis cell-line carrying up to 30 orthogonal sites (5 recombinases x 6 orthogonal sites for each). The number of integrated transgene copies and therefore the overall expression level would then be as good as with lentiviruses, but integration locus would be known and expression could be sustained for a much longer period of time, reducing the overall cost of antibody production in an industrial setup.

Even though the landing pad platform and therefore the use of recombinases can be extremely useful for high-throughput screening of circuits' libraries or for therapeutic purposes, lentiviruses mediated integrations are still useful to rapidly validate the proof of concept of applications necessitating high-levels of transgene expression. For my intercellular communication co-culture experiments for instance, the outputs had to be measured before the media was depleted of nutrients and before the cells reached confluence. Such constraints therefore translate into the need for both a rapid and an efficient production of the signaling

molecules. I therefore used lentivirus to create the sender cell lines, which allowed me to rapidly validate the concept of the two synthetic intercellular communication systems I designed.

It is currently rather difficult to design, built and test synthetic biological systems with precise spatio-temporal properties and characteristics. Both time constraints (for instance half-life of the signaling molecule) and spatial constraints (diffusion constant) need to be satisfied. Last but not least, one has to take into account the structure of the system itself which is constantly evolving (growth of the cells).

Therefore, pushing the two communication systems beyond the proof of concept would not require a more robust characterization but it would also mostly require the design of new culture setups. Such setups would notably require enabling both growth and microscopic observation of single layer cell tissues over extended period of times, without the need to detach (or passage) the cells and to change the whole media culture. Indeed, replacing the media would result in a drastic reduction of the signaling molecule concentration. Also, a significant density of cells is required to produce a large amount of signaling molecules, which therefore rapidly reach confluence and become unstable (more fragile and more prone to epigenetic changes, which often perturbs the expression of the transgenes).

Once a setup that would prevent these problems is found, both intercellular communication systems could be integrated for interesting projects such as pattern formation with applications in tissue engineering. If we could demonstrate that VLPs and phloretin have different diffusion constants, one could then envision creating a synthetic Turing pattern. For the required morphogens, phloretin could act as an activator molecule to express a specific characteristic while virus like particles could be filled up with repressor proteins to act as inhibitors, which have the ability to suppress the activator's expression. One could even imagine using a third morphogen by filling up the virus like particles with a different transactivator protein in order to create more complex patterns.

To my mind, we are still years away from therapies using engineered cells carrying synthetic circuits to repair damaged or non-functional organs or to create de-novo tissues. I tried to advance the field by designing and validating new platforms to fasten both the development of new synthetic systems and the understanding of the underlying mechanisms of their functioning. Eager to apply engineering principles to solve real-life problems, I decided to focus on the engineering of simpler "organisms" to cure diseases. Right after my PhD, I will be working on the use of engineered phages to manipulate the composition of human microbiome, which is known to be heavily involved in the health status of individuals.

Chapter 7 Bibliography

1. Chan, L. Y. *et al.* Refactoring bacteriophage T7. *Mol Syst Biol* **1**, 2005.0018 (2005).
2. Gardner, T. S. *et al.* Construction of a genetic toggle switch in *Escherichia coli*. *Nature* **403**, 339–342 (2000).
3. Danino, T., Mondragón-Palomino, O., Tsimring, L. & Hasty, J. A synchronized quorum of genetic clocks. *Nature* **463**, 326–330 (2010).
4. Moon, T. S. *et al.* Genetic programs constructed from layered logic gates in single cells. *Nature* **491**, 249–253 (2012).
5. Purnick, P. E. M. & Weiss, R. The second wave of synthetic biology: from modules to systems. *Nature Reviews Molecular Cell Biology* **10**, 410–422 (2009).
6. Wang, B., Kitney, R. I., Joly, N. & Buck, M. Engineering modular and orthogonal genetic logic gates for robust digital-like synthetic biology. *Nat Commun* **2**, 508 (2011).
7. Friedland, A. E. *et al.* Synthetic Gene Networks That Count. *Science* **324**, 1199–1202 (2009).
8. Elowitz, M. B. & Leibler, S. A synthetic oscillatory network of transcriptional regulators. *Nature* **403**, 335–338 (2000).
9. Siuti, P., Yazbek, J. & Lu, T. K. Synthetic circuits integrating logic and memory in living cells. *Nature Biotechnology* **31**, 448–452 (2013).
10. You, L. *et al.* Programmed population control by cell–cell communication and regulated killing. *Nature* **428**, 868–871 (2004).
11. Chen, M.-T. & Weiss, R. Artificial cell-cell communication in yeast *Saccharomyces cerevisiae* using signaling elements from *Arabidopsis thaliana*. *Nature Biotechnology* **23**, 1551–1555 (2005).
12. Prindle, A. *et al.* A sensing array of radically coupled genetic 'biopixels'. *Nature* **481**, 39–44 (2012).
13. Kramer, B. P. *et al.* An engineered epigenetic transgene switch in mammalian cells. *Nature Biotechnology* **22**, 867–870 (2004).
14. Ausländer, S., Ausländer, D., Müller, M., Wieland, M. & Fussenegger, M. Programmable single-cell mammalian biocomputers. *Nature* (2012). doi:10.1038/nature11149
15. Smith, M. C. M. *et al.* A synthetic mammalian gene circuit reveals antituberculosis compounds. *Proceedings of the National Academy of Sciences* **105**, 9994–9998 (2008).
16. Xie, Z. *et al.* Multi-input RNAi-based logic circuit for identification of specific cancer cells. *Science* **333**, 1307–1311 (2011).
17. Kemmer, C. *et al.* Self-sufficient control of urate homeostasis in mice by a synthetic circuit. *Mol Ther* **28**, 355–360 (2010).
18. Bacchus, W. *et al.* Synthetic two-way communication between mammalian cells. *Nature Biotechnology* **30**, 991–996 (2012).
19. Ryan, R. P. & Dow, J. M. Diffusible signals and interspecies communication in bacteria. *Microbiology* **154**, 1845–1858 (2008).
20. Peterson, D. O., Beifuss, K. K. & Morley, K. L. Context-dependent gene expression: cis-acting negative effects of specific procaryotic plasmid sequences on eucaryotic genes. Human gene expression. (1999).
21. Nilsson, L. M. *et al.* Mouse genetics suggests cell-context dependency for Myc-regulated metabolic enzymes during tumorigenesis. *PLoS Genet.* **8**, e1002573 (2012).
22. Chen, C.-M., Krohn, J., Bhattacharya, S. & Davies, B. A Comparison of Exogenous Promoter Activity at the ROSA26 Locus Using a PhiC31 Integrase Mediated Cassette Exchange Approach in Mouse ES Cells. *PLoS ONE* **6**, e23376 (2011).

24. Qin, J. Y. *et al.* Systematic Comparison of Constitutive Promoters and the Doxycycline-Inducible Promoter. *PLoS ONE* **5**, e10611 (2010).
25. Lois, C. *et al.* Germline transmission and tissue-specific expression of transgenes delivered by lentiviral vectors. *Science* **295**, 868–872 (2002).
26. Knowling, S. & Morris, K. V. in *Progress in Molecular Biology and Translational Science* **102**, 1–10 (Elsevier, 2011).
27. Eyquem, J., Poirot, L., Galetto, R., Scharenberg, A. M. & Smith, J. Characterization of three loci for homologous gene targeting and transgene expression. *Biotechnol. Bioeng.* **110**, 2225–2235 (2013).
28. Schneider, R. & Grosschedl, R. Dynamics and interplay of nuclear architecture, genome organization, and gene expression. *genesdev.cshlp.org*
29. Recillas-Targa, F. *et al.* Position-effect protection and enhancer blocking by the chicken beta -globin insulator are separable activities. **99**, 6883–6888 (2002).
30. Lombardo, A. *et al.* Site-specific integration and tailoring of cassette design for sustainable gene transfer. *Nat Meth* **8**, 861–869 (2011).
31. Sharma, N. *et al.* The impact of cHS4 insulators on DNA transposon vector mobilization and silencing in retinal pigment epithelium cells. *PLoS ONE* **7**, e48421 (2012).
32. Jordan, J. D., Landau, E. M. & Iyengar, R. Signaling networks: the origins of cellular multitasking. *Cell* **103**, 193 (2000).
33. Bögre, L. *Cell Signalling Mechanisms in Plants*. *els.net* (John Wiley & Sons, Ltd, 2001). doi:10.1002/9780470015902.a0020134
34. Lanning, N. J. & Carter-Su, C. Recent advances in growth hormone signaling. *Rev Endocr Metab Disord* **7**, 225–235 (2007).
35. Andersson, E. R., Sandberg, R. & Lendahl, U. Notch signaling: simplicity in design, versatility in function. *dev.biologists.org*
36. Bacchus, W. & Fussenegger, M. Engineering of synthetic intercellular communication systems. *Metabolic Engineering* **16**, 33–41 (2013).
37. Payne, S. & You, L. Engineered Cell-Cell Communication and Its Applications. *Adv. Biochem. Eng. Biotechnol.* (2013). doi:10.1007/10_2013_249
38. Rusk, N. Synthetic biology: Let's start talking. *Nat Meth* **9**, 1052–1052 (2012).
39. Matsuda, M., Koga, M., Nishida, E. & Ebisuya, M. Synthetic Signal Propagation Through Direct Cell-Cell Interaction. *Science Signaling* **5**, ra31–ra31 (2012).
40. Weber, W. *et al.* Engineered *Streptomyces* quorum-sensing components enable inducible siRNA-mediated translation control in mammalian cells and adjustable transcription control in mice. *J. Gene Med.* **7**, 518–525 (2005).
41. Wang, W.-D., Chen, Z.-T., Kang, B.-G. & Li, R. Construction of an artificial intercellular communication network using the nitric oxide signaling elements in mammalian cells. *Experimental Cell Research* **314**, 699–706 (2008).
42. Weber, W., Schuetz, M., Déneraud, N. & Fussenegger, M. A synthetic metabolite-based mammalian inter-cell signaling system. *Mol. BioSyst.* **5**, 757 (2009).
43. Weber, W. *et al.* Synthetic ecosystems based on airborne inter- and intrakingdom communication. *Proceedings of the National Academy of Sciences* **104**, 10435–10440 (2007).
44. Bleris, L. *et al.* Synthetic incoherent feedforward circuits show adaptation to the amount of their genetic template. *Mol Syst Biol* **7**, 519 (2011).
45. Mukherji, S. *et al.* MicroRNAs can generate thresholds in target gene expression. *Nat Genet* **43**, 854–859 (2011).
46. Tigges, M. *et al.* A tunable synthetic mammalian oscillator. *Nature* **457**, 309–312 (2009).
47. Siciliano, V. *et al.* Construction and modelling of an inducible positive feedback loop stably integrated in a mammalian cell-line. *PLoS Comp Biol* **7**, e1002074 (2011).
48. Lanitis, E. *et al.* Chimeric antigen receptor T Cells with dissociated signaling domains exhibit focused antitumor activity with reduced potential for toxicity in vivo. *Cancer Immunol. Res.* **1**, 43–53 (2013).
49. Grushkin, D. The new drug circuit. *Nat Med* **18**, 1452–1454 (2012).

50. Ruder, W., Lu T. & Collins, J. Synthetic Biology Moving into the Clinic. *Science* **333**, 1248–1252 (2011).
51. Folcher, M. & Fussenegger, M. Synthetic biology advancing clinical applications. *Current Opinion in Chemical Biology* **16**, 345–354 (2012).
52. Bacchus, W., Aubel, D. & Fussenegger, M. Biomedically relevant circuit-design strategies in mammalian synthetic biology. *Mol Syst Biol* **9**, (2013).
53. Ye, H., Aubel, D. & Fussenegger, M. Synthetic mammalian gene circuits for biomedical applications. *Current Opinion in Chemical Biology* **17**, 910–917 (2013).
54. Ye, H. *et al.* A Synthetic Optogenetic Transcription Device Enhances Blood-Glucose Homeostasis in Mice. *Science* **332**, 1565–1568 (2011).
55. Moehle, E. A. *et al.* Targeted gene addition into a specified location in the human genome using designed zinc finger nucleases. *Proc. Natl. Acad. Sci. U.S.A.* **104**, 3055–3060 (2007).
56. Carlson, D. F., Fahrenkrug, S. C. & Hackett, P. B. Targeting DNA With Fingers and TALENs. *Mol Ther Nucleic Acids* **1**, e3 (2012).
57. Reyon, D. *et al.* FLASH assembly of TALENs for high-throughput genome editing. *Nature Biotechnology* **30**, 460–465 (2012).
58. Perez-Pinera, P. *et al.* Gene targeting to the ROSA26 locus directed by engineered zinc finger nucleases. *Nucleic Acids Res.* **40**, 3741–3752 (2012).
59. Hendel, A. *et al.* Quantifying Genome-Editing Outcomes at Endogenous Loci with SMRT Sequencing. *Cell* **7**, 293–305 (2014).
60. Zhu, F. *et al.* DICE, an efficient system for iterative genomic editing in human pluripotent stem cells. *Nucleic Acids Res.* **42**, e34 (2014).
61. Olivares, E. C. *et al.* Phage R4 integrase mediates site-specific integration in human cells. *Gene* **278**, 167–176 (2001).
62. Zhou, H., Liu, Z.-G., Sun, Z.-W., Huang, Y. & Yu, W.-Y. Generation of stable cell lines by site-specific integration of transgenes into engineered Chinese hamster ovary strains using an FLP-FRT system. *Journal of Biotechnology* **147**, 122–129 (2010).
63. Bushman, F. *et al.* Efficient integration of transgenes into a defined locus in human embryonic stem cells. *Nucleic Acids Res.* **38**, e96 (2010).
64. Bai, H. *et al.* Single-molecule analysis reveals the molecular bearing mechanism of DNA strand exchange by a serine recombinase. *Proceedings of the National Academy of Sciences* **108**, 7419–7424 (2011).
65. Hockemeyer, D. *et al.* Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. *Nature Biotechnology* **27**, 851–857 (2009).
66. Sadelain, M., Papapetrou, E. P. & Bushman, F. D. Safe harbours for the integration of new DNA in the human genome. *Nature Reviews Cancer* **12**, 51–58 (2012).
67. Weber, E., Engler, C., Gruetzner, R., Werner, S. & Marillonnet, S. A modular cloning system for standardized assembly of multigene constructs. *PLoS ONE* **6**, e16765 (2011).
68. Engler, C. *et al.* Golden gate shuffling: a one-pot DNA shuffling method based on type II restriction enzymes. *PLoS ONE* **4**, e5553 (2009).
69. Torella, J. P. *et al.* Rapid construction of insulated genetic circuits via synthetic sequence-guided isothermal assembly. *Nucleic Acids Res.* **42**, 681–689 (2014).
70. Guye, P., Li, Y., Wroblewska, L., Duportet, X. & Weiss, R. Rapid, modular and reliable construction of complex mammalian gene circuits. *Nucleic Acids Res.* **41**, e156 (2013).
71. Schmid-Burgk, J. L. *et al.* Rapid hierarchical assembly of medium-size DNA cassettes. *Nucleic Acids Res.* **40**, e92 (2012).
72. Shiyuza, H. *et al.* Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in *Escherichia coli* using an F-factor-based vector. *Nucleic Acids Res.* **89**, 8794–8797 (1992).
73. Chevalier, C., Henry, I., Nicolas, JF. CpG content affects gene silencing in mice: evidence from novel transgenes. *Genome Biol.* **4**, R53 (2003).
74. Chen, Z. *et al.* Improved Production and Purification of Minicircle DNA Vector Free of Plasmid Bacterial Sequences and Capable of Persistent Transgene Expression In Vivo.

- Hum. Gene Ther.* **16**, 126–131 (2005).
75. Anastassiadis, K. *et al.* Dre recombinase, like Cre, is a highly efficient site-specific recombinase in E. coli, mammalian cells and mice. *Nucleic Acids Res.* **41**, e156 (2013).
 76. Chen, Y. *et al.* Genetic control of mammalian T-cell proliferation with synthetic RNA regulatory systems. *Proceedings of the National Academy of Sciences* **107**, 8531–8536 (2010).
 77. Hartley, J. L., Temple, G. F. & Brasch, M. A. DNA cloning using in vitro site-specific recombination. *Genome Research* **10**, 1788–1795 (2000).
 78. Stewart, S. A. *et al.* Lentivirus-delivered stable gene silencing by RNAi in primary cells. *RNA* **9**, 493–501 (2003).
 79. Malchin, N. *et al.* Optimization of coliphage HK022 Integrase activity in human cells. *Gene* **437**, 9–13 (2009).
 80. Krishnakumar, R. *et al.* Simultaneous non-contiguous deletions using large synthetic DNA and site-specific recombinases. *Nucleic Acids Res.* (2014). doi:10.1093/nar/gku509
 81. Park, J. T. & Leach, S. D. TAILOR: transgene activation and inactivation using lox and rox in zebrafish. *PLoS ONE* **8**, e85218–e85218 (2012).
 82. Long, D. *et al.* In vivo site-specific integration of transgene in silkworm via PhiC31 integrase-mediated cassette exchange. *Insect Biochem. Mol. Biol.* **43**, 997–1008 (2013).
 83. Turan, S. *et al.* Expanding Flp-RMCE options: the potential of Recombinase Mediated Twin-Site Targeting (RMSTT). *Gene* (2014). doi:10.1016/j.gene.2014.06.002
 84. Osterwalder, M. *et al.* Dual RMCE for efficient re-engineering of mouse mutant alleles. *Nat Meth* **7**, 893–895 (2010).
 85. Jain, P. *et al.* Specialized transduction designed for precise high-throughput unmarked deletions in Mycobacterium tuberculosis. *MBio* **5**, e01245–14 (2014).
 86. Wang, Y. *et al.* Targeted DNA recombination in vivo using an adenovirus carrying the cre recombinase gene. *Proc. Natl. Acad. Sci. U.S.A.* **93**, 3932–3936 (1996).
 87. Grindley, N. D. F., Whiteson, K. L. & Rice, P. A. Mechanisms of Site-Specific Recombination*. *Annu. Rev. Biochem.* **75**, 567–605 (2006).
 88. G Pan, K. L. P. D. S. & G Pan, K. L. P. D. S. Mechanism of cleavage and ligation by FLP recombinase: classification of mutations in FLP protein by in vitro complementation analysis. *Molecular and Cellular Biology* **13**, 3167 (1993).
 89. Esposito, D., Esposito, D., Scocca, J. J. & Scocca, J. J. The integrase family of tyrosine recombinases: evolution of a conserved active site domain. *Nucleic Acids Res.* **25**, 3605–3614 (1997).
 90. Mumm, J. P., Landy, A. & Gelles, J. Viewing single lambda site-specific recombination events from start to finish. *EMBO J.* **25**, 4586–4595 (2006).
 91. Zuwen Zhang, B. L. & Zuwen Zhang, B. L. Cre recombinase-mediated inversion using lox66 and lox71: method to introduce conditional point mutations into the CREB-binding protein. *Nucleic Acids Res.* **30**, e90 (2002).
 92. Gaj, T. *et al.* Structure-guided reprogramming of serine recombinase DNA sequence specificity. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 498–503 (2011).
 93. Hubbard, E. J. A. & Hubbard, E. J. A. FLP/FRT and Cre/lox recombination technology in C. elegans. *Methods* (2014). doi:10.1016/j.ymeth.2014.05.007
 94. Turan, S. *et al.* Recombinase-mediated cassette exchange (RMCE) - a rapidly-expanding toolbox for targeted genomic modifications. *Gene* **515**, 1–27 (2013).
 95. Smith, M. & Thorpe, H. M. Diversity in the serine recombinases. *Mol Microbiol* **44**, 299–307 (2002).
 96. Xu, Z. *et al.* Accuracy and efficiency define Bxb1 integrase as the best of fifteen candidate serine recombinases for the integration of DNA into the human genome. *BMC Biotechnol* **13**, 87 (2013).
 97. Thyagarajan, B. *et al.* Site-specific genomic integration in mammalian cells mediated by phage phiC31 integrase. *Molecular and Cellular Biology* **21**, 3926–3934 (2001).
 98. Kim, A. I. *et al.* Mycobacteriophage Bxb1 integrates into the Mycobacterium smegmatis groEL1 gene. *Mol Microbiol* **50**, 463–473 (2003).

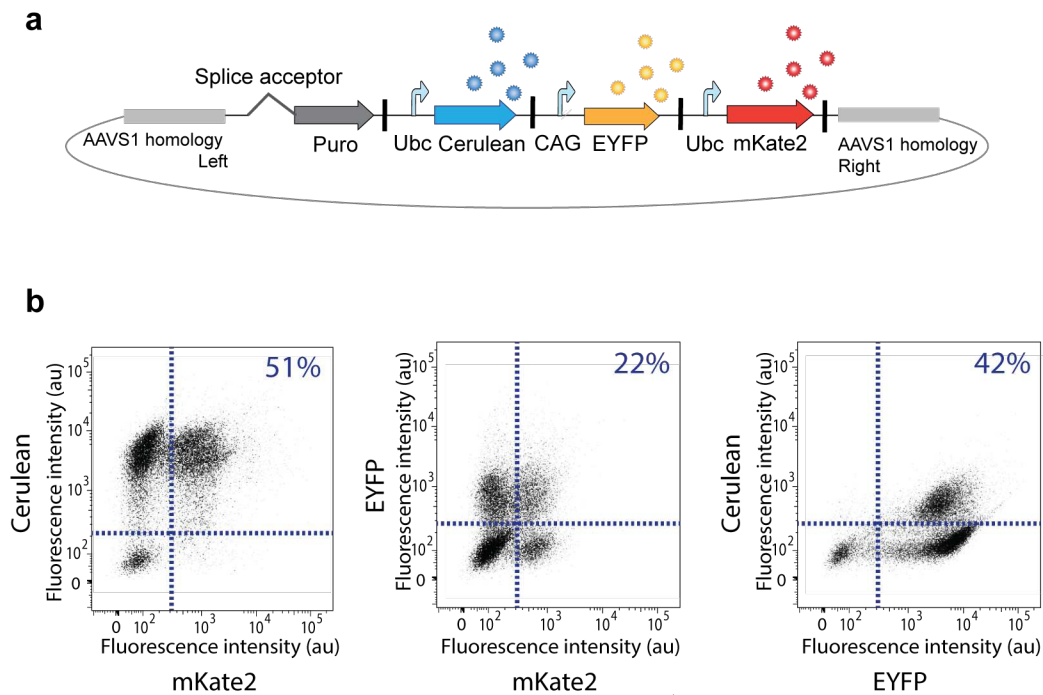
99. Singh, S., Rockenbach, K., Dedrick, R. M., VanDemark, A. P. & Hatfull, G. F. Cross-talk between Diverse Serine Integrases. *Journal of Molecular Biology* **426**, 318–331 (2014).
100. Hatfull, G. F. *et al.* Complete Genome Sequences of 138 Mycobacteriophages. *Journal of Virology* **86**, 2382–2384 (2012).
101. Park, A., Jeong, H.-H., Lee, J. & Lee, C.-S. The inhibitory effect of phloretin on the formation of Escherichia coli O157:H7 biofilm in a microfluidic system. *BioChip J* **6**, 299–305 (2012).
102. Barreca, D. *et al.* Biochemical and antimicrobial activity of phloretin and its glycosylated derivatives present in apple and kumquat. *Food Chemistry* **160**, 292–297 (2014).
103. Teran, W. *et al.* Antibiotic-Dependent Induction of Pseudomonas putida DOT-T1E TtgABC Efflux Pump Is Mediated by the Drug Binding Repressor TtgR. *Antimicrobial Agents and Chemotherapy* **47**, 3067–3072 (2003).
104. Rutherford, K., Yuan, P., Perry, K., Sharp, R. & Van Duyne, G. D. Attachment site recognition and regulation of directionality by the serine integrases. *Nucleic Acids Res.* **41**, 8341–8356 (2013).
105. Singh, S., Ghosh, P. & Hatfull, G. F. Attachment site selection and identity in Bxb1 serine integrase-mediated site-specific recombination. *PLoS Genet.* **9**, e1003490 (2013).
106. Terán, W. *et al.* Effector-repressor interactions, binding of a single effector molecule to the operator-bound TtgR homodimer mediates derepression. *Journal of Biological Chemistry* **281**, 7102–7109 (2006).
107. Gitzinger, M. *et al.* Controlling transgene expression in subcutaneous implants using a skin lotion containing the apple metabolite phloretin. *Proceedings of the National Academy of Sciences* **106**, 10638–10643 (2009).
108. Watts, K. T., Lee, P. C. & Schmidt-Dannert, C. Exploring Recombinant Flavonoid Biosynthesis in Metabolically Engineered Escherichia coli. *ChemBioChem* **5**, 500–507 (2004).
109. Jiang, H. *et al.* Metabolic Engineering of the Phenylpropanoid Pathway in Saccharomyces cerevisiae. *Applied and Environmental Microbiology* **71**, 2962–2969 (2005).
110. Hwang, Il, E., Kaneko, M., Ohnishi, Y. & Horinouchi, S. Production of plant-specific flavanones by Escherichia coli containing an artificial gene cluster. *Applied and Environmental Microbiology* **69**, 2699–2706 (2003).
111. Krell, T. *et al.* Optimization of the palindromic order of the TtgR operator enhances binding cooperativity. *Journal of Molecular Biology* **369**, 1188–1199 (2007).
112. Kaczmarczyk, S. J., Sitaraman, K., Young, H. A., Hughes, S. H. & Chatterjee, D. K. Protein delivery using engineered virus-like particles. *Proc. Natl. Acad. Sci. U.S.A.* **108**, 16998–17003 (2011).
113. Farley, D., *et al.* Factors that influence VSV-G pseudotyping and transduction efficiency of lentiviral vectors-in vitro and in vivo implications. *J. Gene Med.* **9**, 345–356 (2007).

Chapter 8 Annexes

Annexes for Chapter 2 - A platform for rapid prototyping of synthetic genetic networks in mammalian cells

- Supp Figure 2-1.** Non-homogeneity of transgene expression after zinc-finger nuclease mediated integration of a 4-gene circuit.
- Supp Figure 2-2.** Landing pad chassis cell line construction: Southern blot genomic analysis of landing pad integration with zinc-finger nuclease in wild-type cells.
- Supp Figure 2-3.** Landing pad chassis cell line construction: PCR confirmations of mono-allelic landing pad integration.
- Supp Figure 2-4.** Details of the mMoClo backbones used in the assembly workflow depicted in Figure 2.
- Supp Figure 2-5.** Effect of nucleotide sequence optimization and nuclear localization signal on the integration efficiency of BxB1 recombinase.
- Supp Figure 2-6.** Influence of vector ratios (BxB1 expression vector vs. integration vector) on the integration efficiency.
- Supp Figure 2-7.** Microscopy images of different model chassis cell lines after integration of a simple circuit.
- Supp Figure 2-8.** Comparison of circuit integration efficiency in three different chassis cell lines.
- Supp Figure 2-9.** Flow cytometry analysis of chassis cell line fluorescence before and after integration of circuits and selection with puromycin.
- Supp Figure 2-10.** Specificity of integration into the landing pad: Southern blot genomic analysis assessing the specificity of BxB1 mediated integration.
- Supp Figure 2-11.** Sequence alignment of insert flanking sequences of the 42 isolated clones.
- Supp Figure 2-12.** Circuit performance in HeLa and CHO chassis cell-lines.
- Supp Figure 2-13.** Early genomic rearrangement of the 7-gene payload following transfection.
- Supp Figure 2-14.** Microscopy images of isolated colonies representing the 18 different phenotypes of the circuit library.
- Supp Figure 2-15.** Annotated maps of vectors used in this study.
- Supp Table 2-1.** Comparison of integration efficiencies of different recombinases.
- Supp Text 2-1.** DNA sequences of the probes used for the Southern Blots.
- Supp Text 2-2.** DNA sequences of the primers used for genomic PCRs.
- Supp Text 2-3.** DNA sequences of the genetic components from our mammalian parts library used in this study.

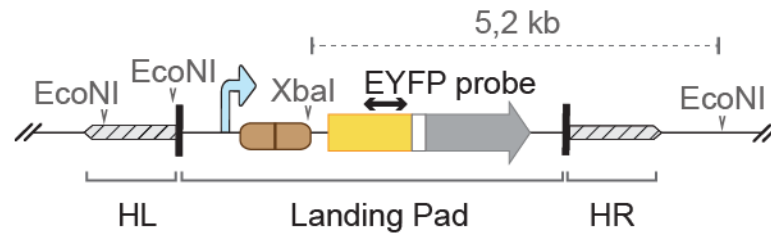
Supplementary Figure 2-1. Non-homogeneity of transgene expression after zinc-finger nuclease mediated integration of a 4-gene circuit.



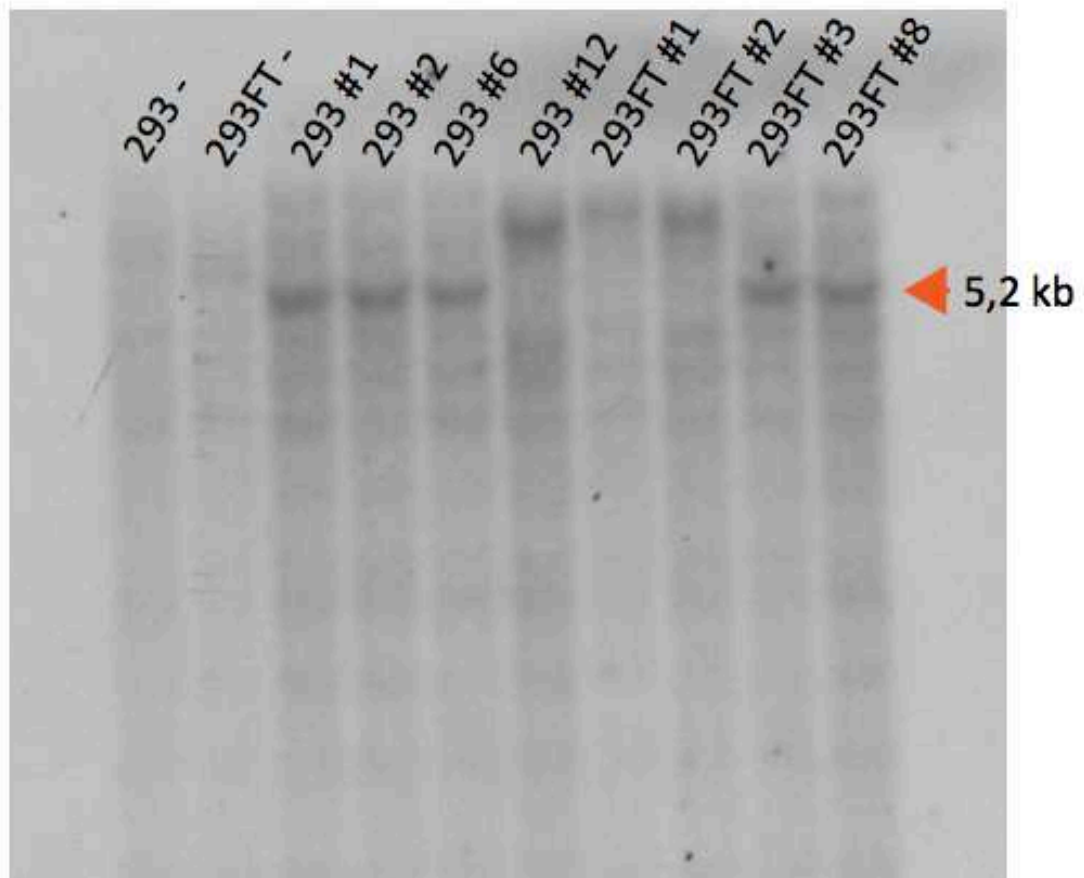
(a) Schematic representation of the 4-gene circuit used for integration; the circuit comprises (1) a promoterless puromycin resistance gene activated only after integration in the AAVS1 locus¹, (2) Ubc promoter constitutively expressing a Cerulean reporter, (3) CAG promoter constitutively expressing an EYFP reporter and (4) UbC promoter constitutively expressing an mKate2 reporter. AAVS1 homology regions flank the circuit to target integrations into the AAVS1 locus. **(b)** Representative two-dimensional fluorescence density plots demonstrating significant heterogeneity in reporter expression levels within the polyclonal population of cells after integration of the circuit and selection with Puromycin. Selection was started 3 days post transfection and cells were assayed 14 days post transfection.

Supplementary Figure 2-2. Landing pad chassis cell line construction: Southern blot genomic analysis of landing pad integration with zinc-finger nuclease in wild-type cells.

A



B

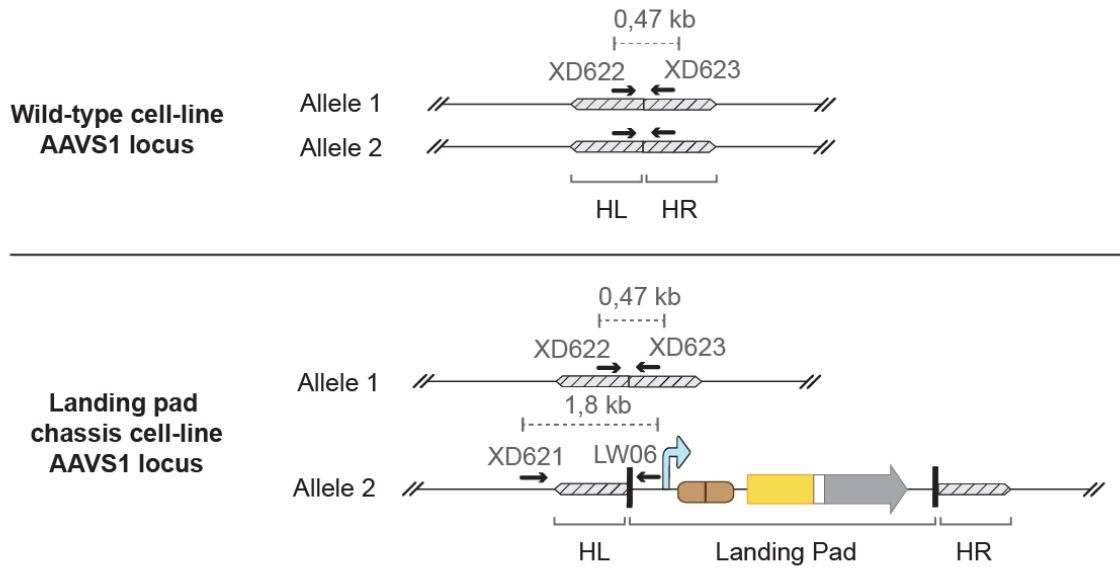


(A) Schematic of the Southern probe and the restriction enzyme cut sites. The landing pad is inserted between the left homology and right homology sequences (HL & HR). The 300 bp probe used for the Southern Blots corresponds to an internal sequence of EYFP. Digestion of chromosomal DNA with EcoNI and XbaI results in a 5.2kb band containing the EYFP gene for the correct clones.

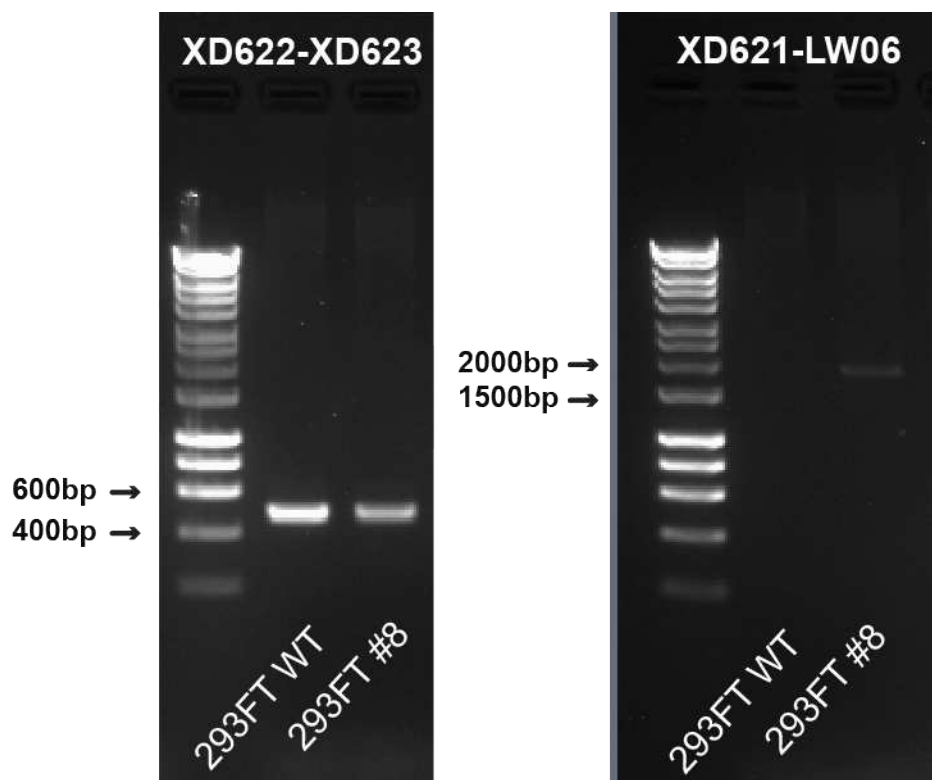
(B) Wild-type cell lines of HEK293FT and HEK293 with no landing pad were used as negative controls (first two lanes on the left). Integration in the correct locus was confirmed for 3 out of 4 clones for HEK293 cell lines, and for 2 out of 4 clones for HEK293FT. Clone HEK293FT#8 (denoted in the text as HEK293FT-LP) was chosen as the monoclonal chassis cell line for all downstream circuit integrations.

Supplementary Figure 2-3. Landing pad chassis cell line validation: PCR confirmations of mono-allelic landing pad integration.

A



B

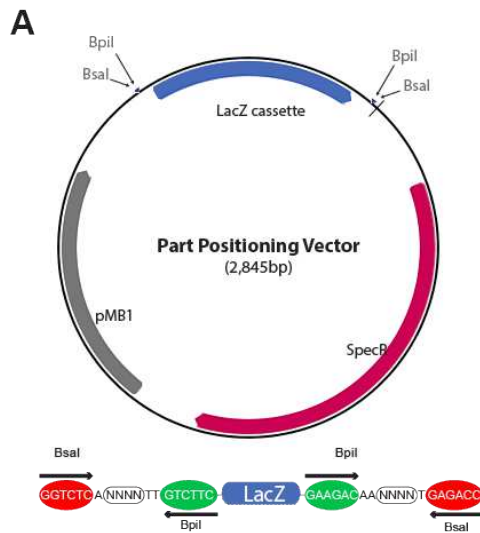


(A) Schematic of the PCRs and primers used. Primers XD622 and XD623 allow amplification of a 0.47kb sequence that overlaps the ZFN cleavage site within the AAVS1 locus. XD622 and XD623 are respectively located within the Left Homology arm (HL) and the Right Homology arm (HR) used for integration of the landing pad. The primers XD621 and LW06 allow amplification of a 1.8kb sequence when the landing pad is specifically inserted in the AAVS1 locus in the right orientation.

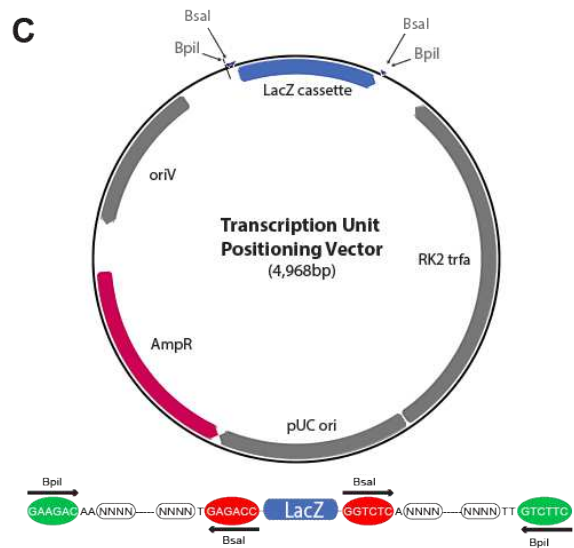
(B) PCR gels on WT and Chassis cell lines. Control PCRs on the wild type cell line with the two primer pairs resulted in a single band at 0.47kb and no band at 1.8kb. PCRs on the chassis cell line resulted in both a band at 0.47kb and a band at 1.8kb, confirming the presence of one non-modified allele and one allele carrying the landing pad.

Supplementary Figure 2-4. Details of the mMoClo backbones used in the assembly workflow depicted in Figure 2.

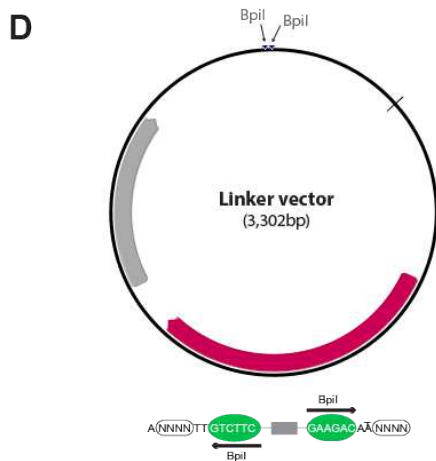
The mMoClo library includes a library of empty backbones that are used to assemble (multi)gene circuits. **(A)** The Part Positioning Vectors are used to create a library of single genetic parts (required later to build functional transcription units): Insulator, Promoter, 5'UTR, Gene coding sequence, 3'UTR and polyA signal. **(B)** To clone a genetic part into the spectinomycin resistant Part Positioning Vector of choice, the part has to be flanked by a BpiI restriction site on each end, together with 4bp overhangs matching the ones from the desired Part Positioning Vector. The BpiI mediated Golden Gate reaction between the genetic part and the Part Positioning Vector replaces the lacZ cassette with the desired genetic part into the Part Positioning Vector and thus allows a rapid and easy visual blue/white screening of the transformed bacteria. The library Part vector created this way does not contain BpiI restriction sites anymore. **(C)** To create a functional transcription unit, the BsaI mediated Golden Gate reaction requires the presence of 6 plasmids: 1 Insulator plasmid, 1 Promoter plasmid, 1 5'UTR plasmid, 1 Gene plasmid, 1 3'UTR plasmid, 1 polyA plasmid (all from the library of parts) and 1 ampicillin resistant Transcription Unit Positioning Vector. Assembly of the transcription unit into the latter one replaces the lacZ cassette and thus allows a rapid and easy visual blue/white screening of the transformed bacteria. The choice of the pTU (Transcription Unit positioning Vector) is directly linked to the position of transcription unit in the final circuit. The transcription unit vector created this way does not contain BsaI restriction sites anymore. **(D)** To create an integration vector (containing a BxB1 recombination site) that is used to assemble a complete circuit, an integration cassette is assembled first into the kanamycin resistant Destination vector, using a mock promoter part together with the BxB1 attB site as the 5'UTR part and a fluorescence/resistance marker of choice for the Gene part. Assembly of this promoterless transcription unit replaces the lacZ gene from the Destination Vector, allowing for a red/green selection of the transformed bacteria. Before this step, the destination vector expresses both lacZ gene and the *crt* red pigment operon, resulting in green pigmentation of the bacteria carrying an intact Destination Vector. The integration vector assembled this way can then be repeatedly used to assemble multi-gene circuits with a BpiI based Golden Gate reaction by combining a set of assembled transcription unit vectors (following each other) together with the linker vector corresponding to the position of the last transcription unit vector. The assembled circuit replaces the *crt* red pigment producing operon and thus allows for easy read/white screening to pick the correct clones.



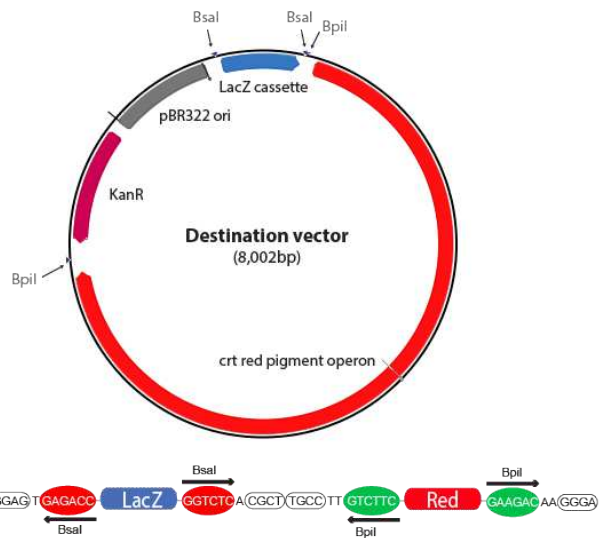
Basic Parts	5' Overhang <i>BspI/BsaI</i>	3' Overhang <i>BspI/BsaI</i>
p_Insulator	GGAG	TACT
p_Promoter	TACT	AATG
p_5'UTR	AATG	AGGT
p_Gene	AGGT	GCTT
p_3'UTR	GCTT	CAAC
p_polyA	CAAC	CGCT



Position Vectors	5' Overhang <i>BspI</i>	3' Overhang <i>BspI</i>
pTU-1	TGCC	GCAA
pTU-2	GCAA	ACTA
pTU-3	ACTA	TTAC
pTU-4	TTAC	CAGA
pTU-5	CAGA	TGTG
pTU-6	TGTG	GAGC
pTU-7	GAGC	AACG
pTU-8	AACG	CTTC
pTU-9	CTTC	AGAC



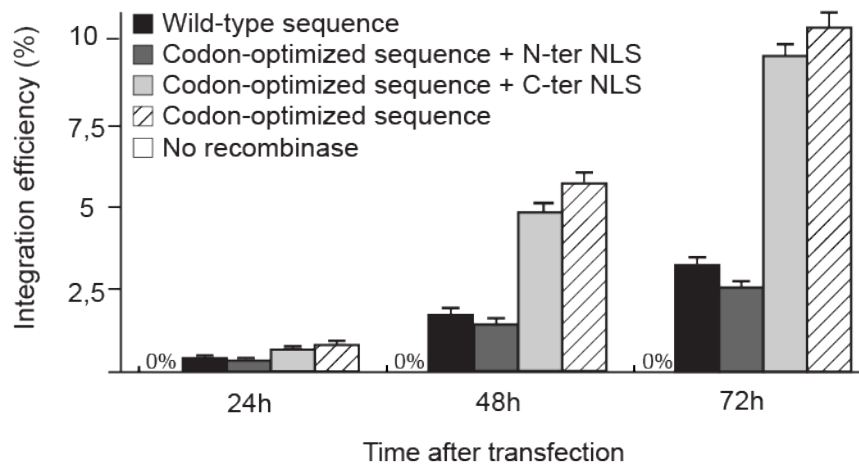
Linkers	5' Overhang <i>BspI</i>	3' Overhang <i>BspI</i>
pLink-1	GCAA	GGGA
pLink-2	ACTA	GGGA
pLink-3	TTAC	GGGA
pLink-4	CAGA	GGGA
pLink-5	TGTG	GGGA
pLink-6	GAGC	GGGA
pLink-7	AACG	GGGA
pLink-8	CTTC	GGGA
pLink-9	AGAC	GGGA



Integration vector	5' Overhang <i>BsaI</i>	3' Overhang <i>BsaI</i>
Integration cassette	GGAG	CGCT

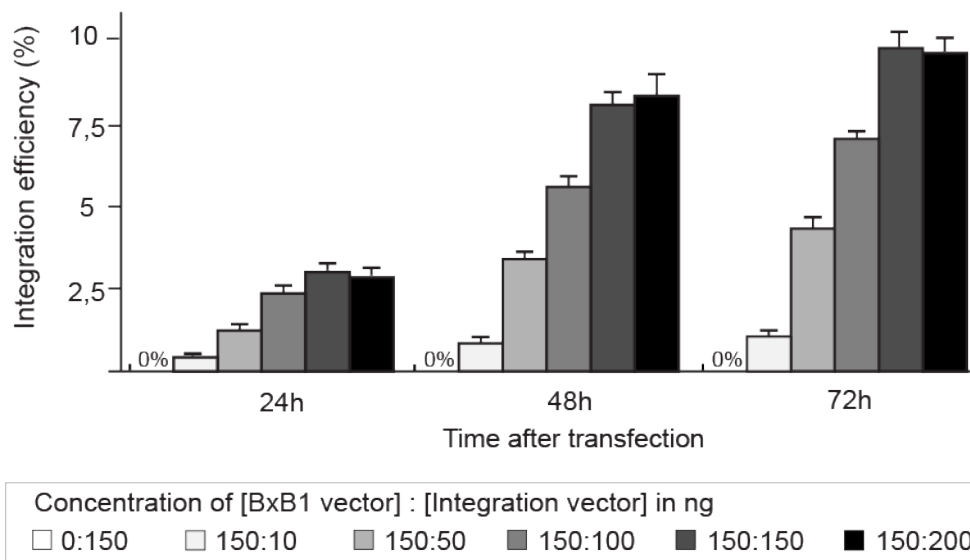
Integration vector	5' Overhang <i>BspI</i>	3' Overhang <i>BspI</i>
Circuit	TGCC	GGGA

Supplementary Figure 2-5. Effect of nucleotide sequence optimization and nuclear localization signal on the integration efficiency of BxB1 recombinase.



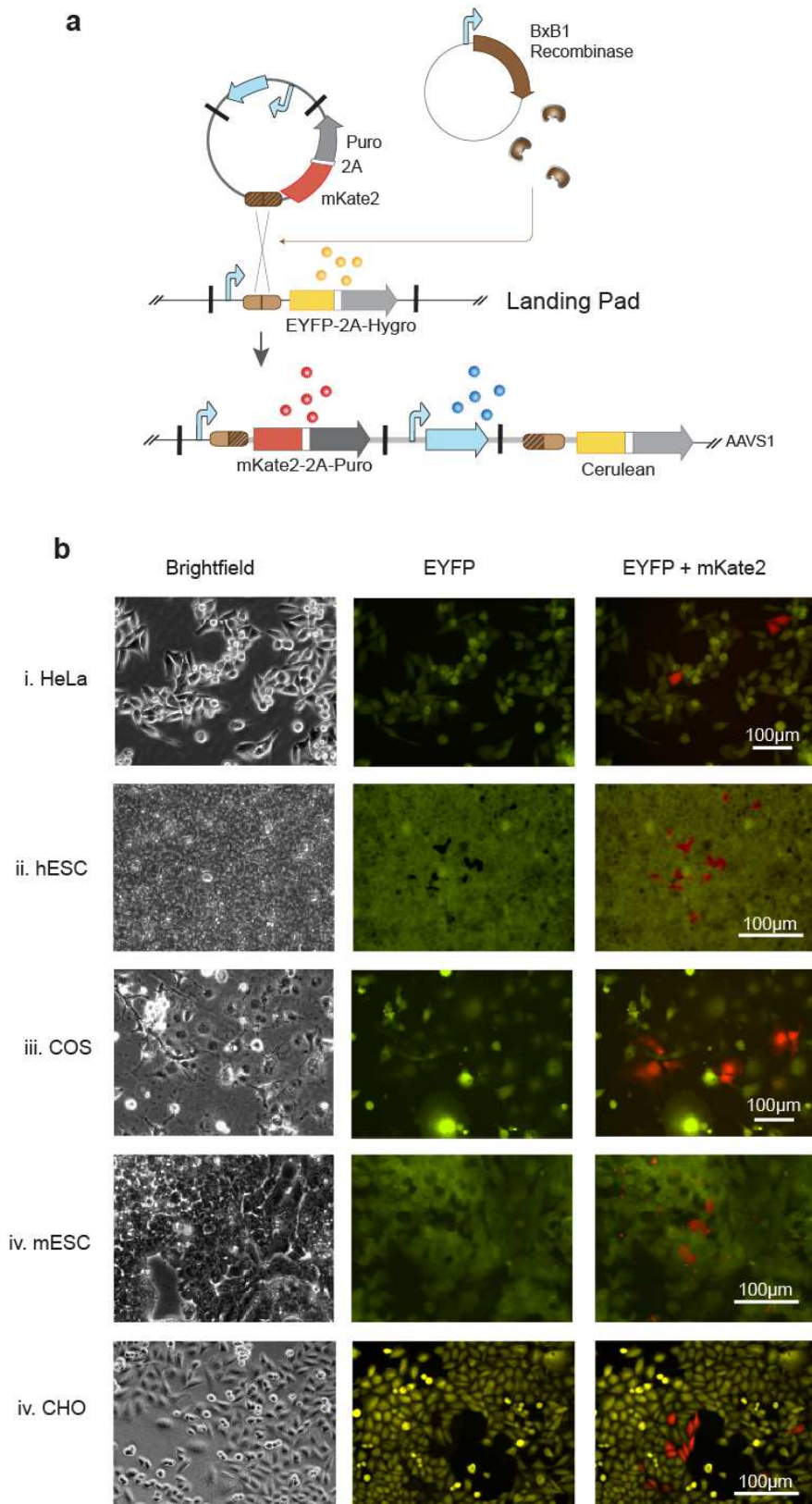
A 12kb integrative plasmid comprising a promoterless Puromycin-2A-mKate2 followed by constitutively expressed Cerulean was co-transfected with BxB1 expressing plasmid into the HEK293FT-LP chassis cell line. Integration efficiency was measured as percentage of mKate2 positive cells within the transfected cell population. The number of cells expressing mKate2 increased with time and reached a maximum 3 days after transfection (we monitored the expression of mKate2 up to 10 days post-transfection, data not shown). Nuclear Localization Signal (NLS) fusion to BxB1 did not improve the integration efficiency. When NLS was added at the N-terminus of the recombinase, significant negative impact on integration efficiency was observed. Our hypothesis is that in this case the NLS may sterically interfere with the N-terminal domain involved in the catalytic cleavage and ligation of DNA. Error bars represent standard deviation from 3 independently repeated experiments.

Supplementary Figure 2-6. Influence of vector ratios (BxB1 expression vector vs integration vector) on integration efficiency.



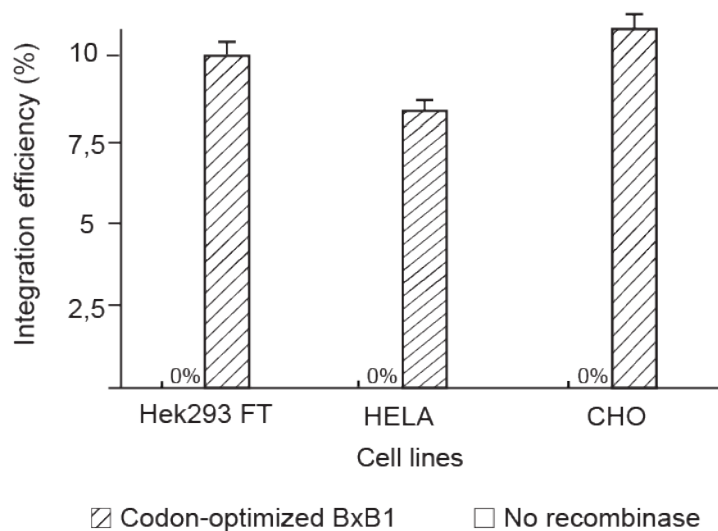
We co-transfected, in different ratios, a 12kb integrative plasmid comprising a promoterless Puromycin-2A-mKate2 construct followed by constitutively expressed Cerulean together with a codon-optimized BxB1 expressing plasmid into the HEK293FT-LP chassis cell line. The transfection mix was supplemented with mock plasmid DNA to always transfect the cells with a total of 400ng of DNA. Integration efficiency was measured as percentage of mKate2 positive cells within the transfected cell population. Error bars represent the standard deviation from 3 independently repeated experiments.

Supplementary Figure 2-7. Microscopy images of different model chassis cell lines after integration of a simple circuit.



(A) The integrative plasmid comprising a promoterless mKate-2A-puro gene cassette and a constitutively expressed Cerulean fluorescent reporter was co-transfected with the BxB1 recombinase expression plasmid in various monoclonal chassis cell lines. Integration events in the landing pad can be monitored in the transfected cells by both expression of mKate2 and reduction in expression of EYFP. **(B)** Fluorescent microscopy images of monoclonal chassis cell lines four days after co-transfection with BxB1 expression plasmid and the integrative plasmid: (i) HeLa cells, (ii) human Embryonic Stem Cells, (iii) COS cells, (iv) mouse Embryonic Stem Cells and (v) CHO cells.

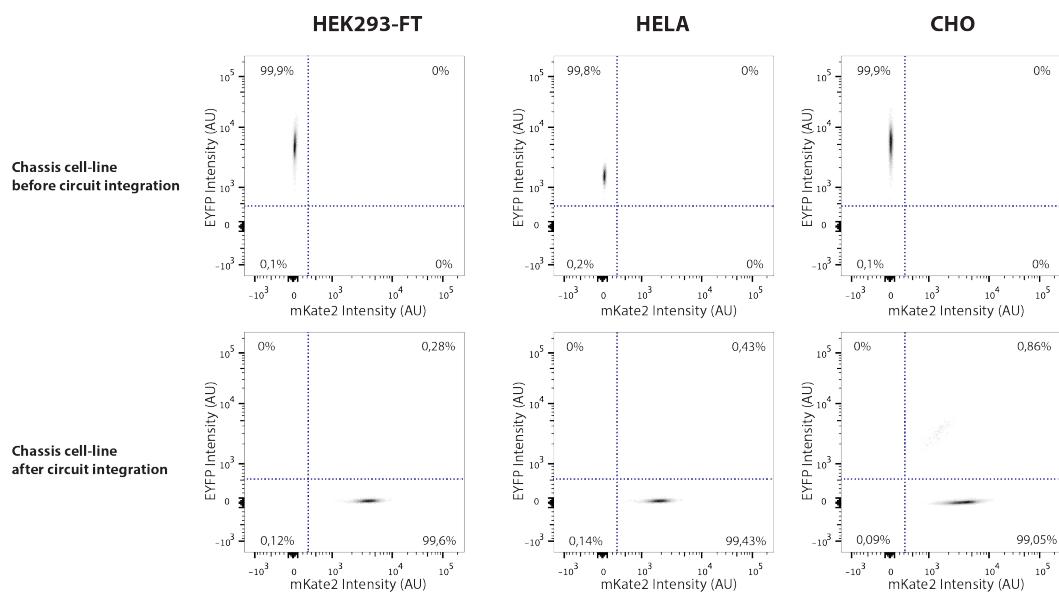
Supplementary Figure 2-8. Comparison of circuit integration efficiency in three different chassis cell lines.



We co-transfected a 12kb integrative plasmid comprising a promoterless Puromycin-2A-mKate2 followed by constitutively expressed Cerulean together with a codon-optimized BxB1 expressing plasmid into our HEK293FT, HeLa and CHO chassis cell-lines. Integration efficiency was measured as percentage of mKate2 positive cells within the transfected cell population 3 days after transfection.

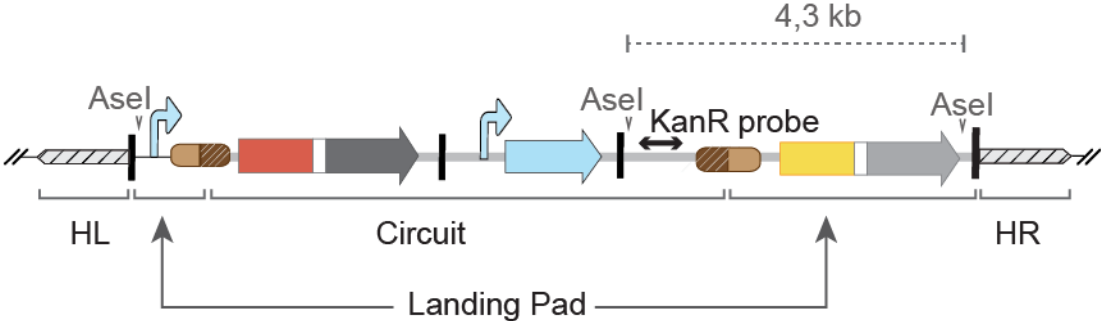
Supplementary Figure 2-9. Flow cytometry analysis of chassis cell line fluorescence before and after integration of circuits and selection with puromycin.

FACS comparative analysis of EYFP and mKate2 fluorescence levels in different chassis cell-lines before integration of the integration vector carrying the promoterless mKate2-2A-puro cassette, and 14 days after integration and selection with puromycin. After selection, more than 99% of cells that survive selection are mKate2 positive and EYFP negative, demonstrating the specificity of the integration events into the landing pad.

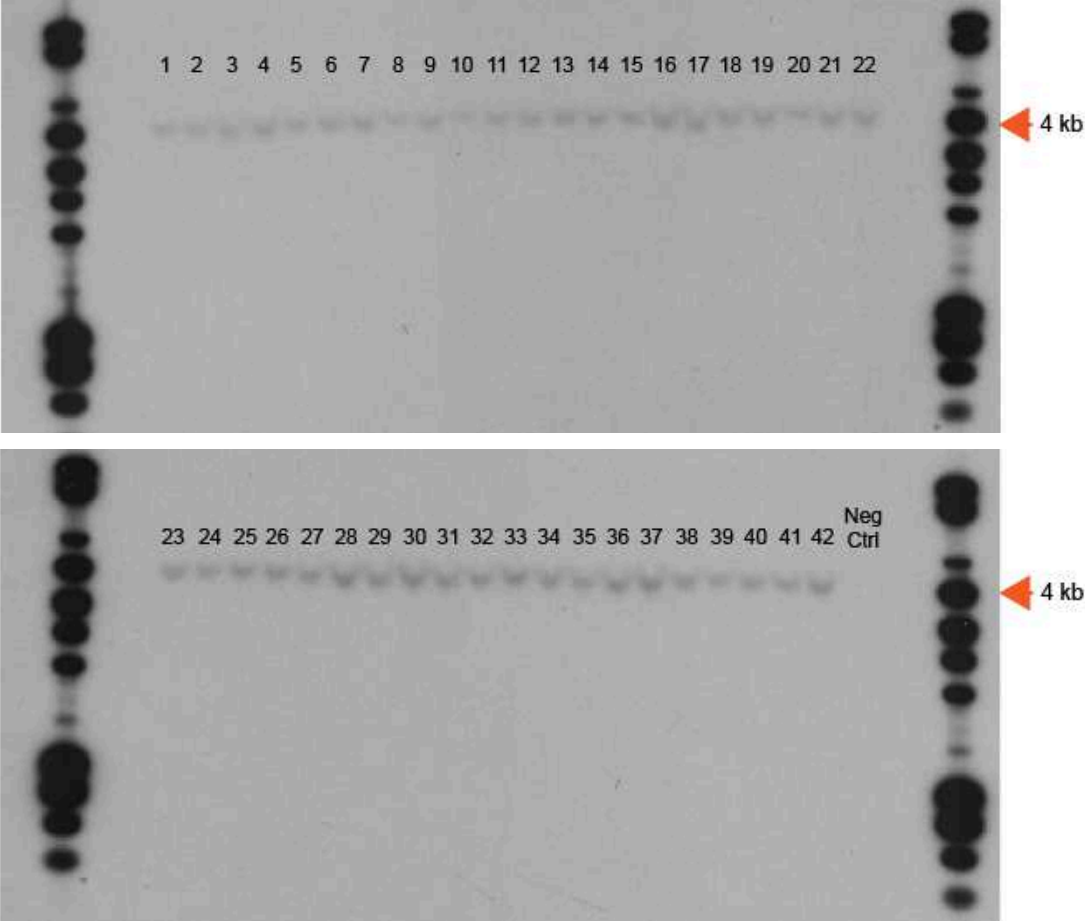


Supplementary Figure 2-10. Specificity of integration into the landing pad: Southern Blot genomic analysis assessing the specificity of BxB1 mediated integration.

A



B



(A) Schematic of the Southern probe and the restriction enzyme cut sites. The 300 bp probe used for the Southern Blots corresponds to an internal sequence of the KanR gene from the integrated plasmid. Digestion of the chromosomal DNA with AseI results in a 4.3kb band containing the KanR gene if the circuit has been correctly integrated into the landing pad.

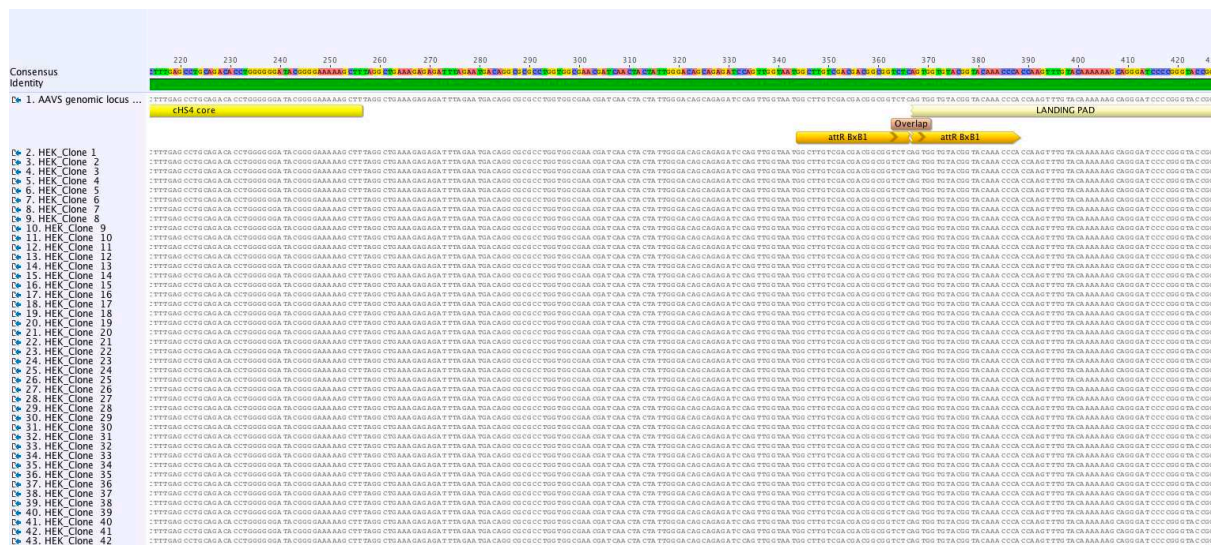
(B) Targeted integration of the integrative plasmid in the landing pad was confirmed for 42 randomly sorted clones that were expanded after selection with puromycin. For all the 42 clones, a single band of the expected size is revealed on the SB. Genomic DNA of the HEK293FT#8 monoclonal chassis cell line before payload integration is used as a negative control (last lane on the right).

Supplementary Figure 2-11. Sequence alignment of insert flanking sequences of the 42 isolated clones.

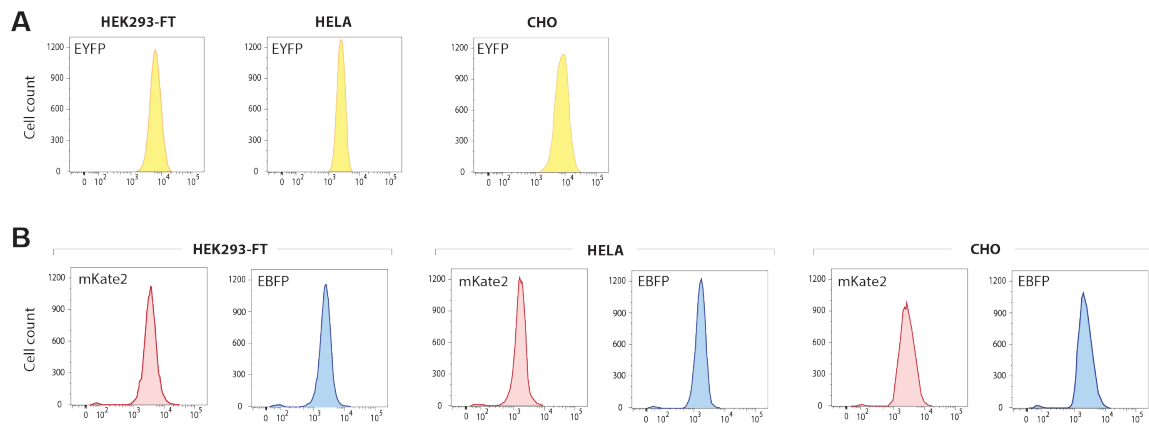
Sequences close to the attL junction



Sequences close to the attR junction

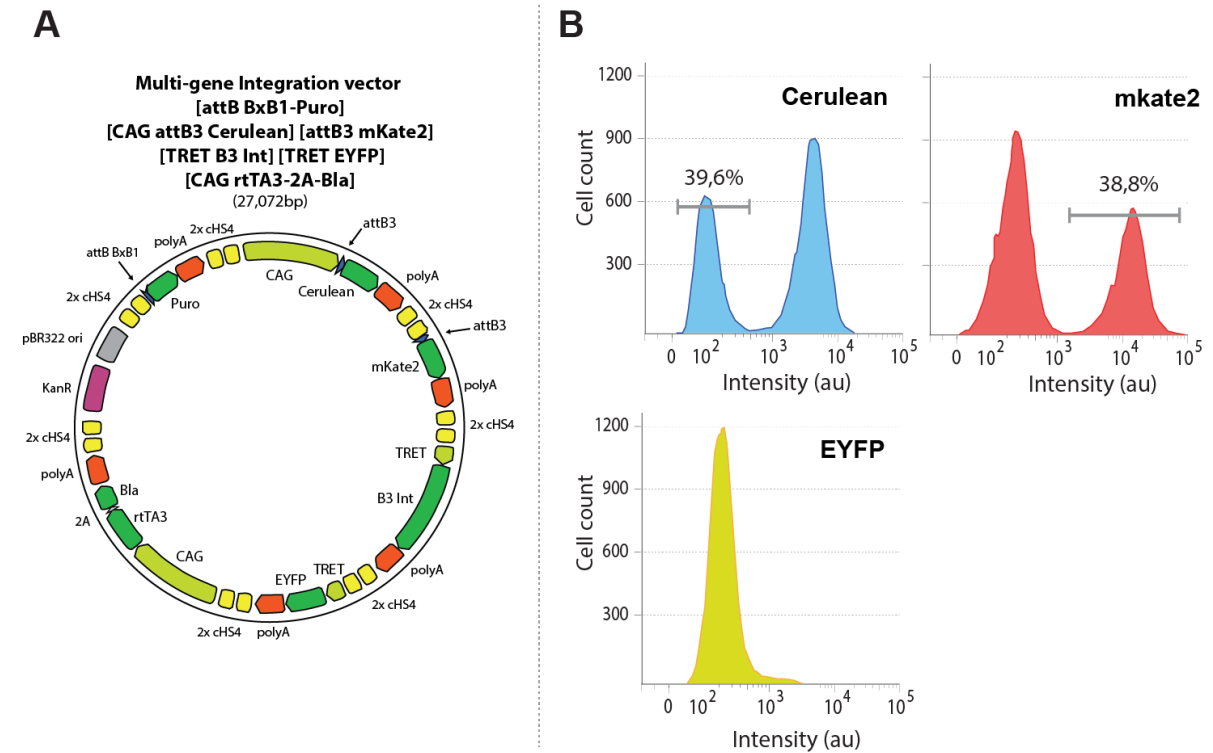


Supplementary Figure 2-12. Circuit performance in Hek293FT, HeLa and CHO chassis cell-lines.



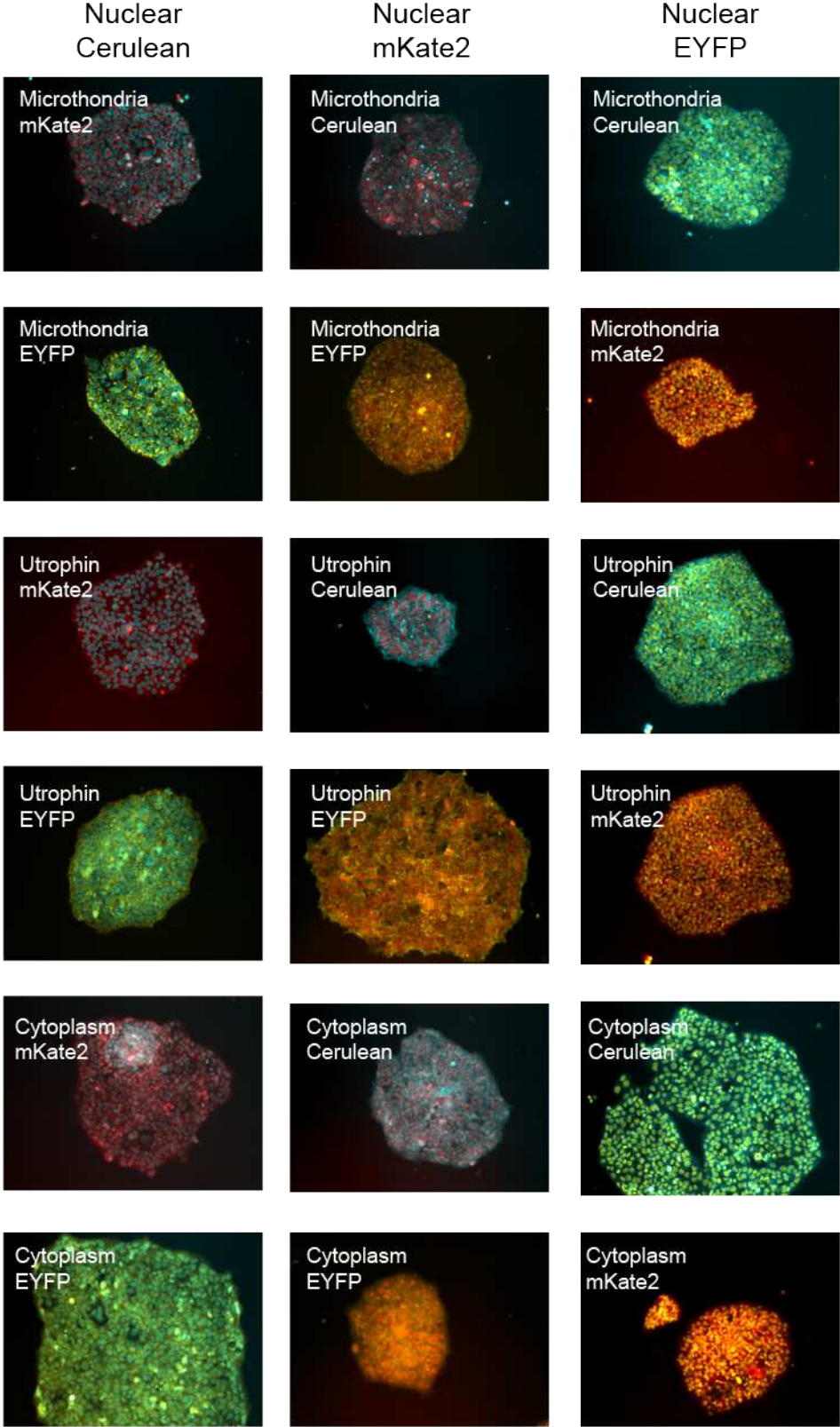
(A) Homogeneity of transgene (EYFP) expression in different landing-pad chassis cell lines. **(B)** FACS histograms of cells selected with puromycin after BxB1 mediated integration in different landing-pad cell lines.

Supplementary Figure 2-13. Early genomic rearrangement of the 7-gene payload following transfection.

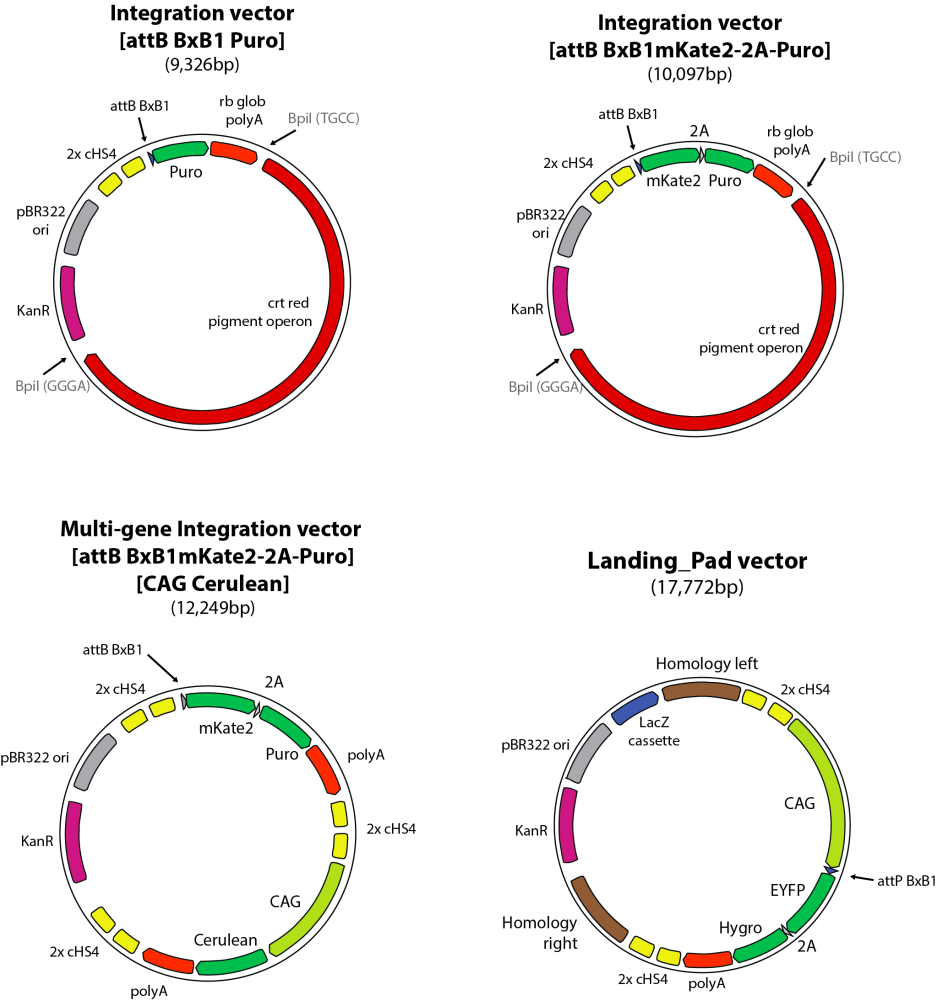


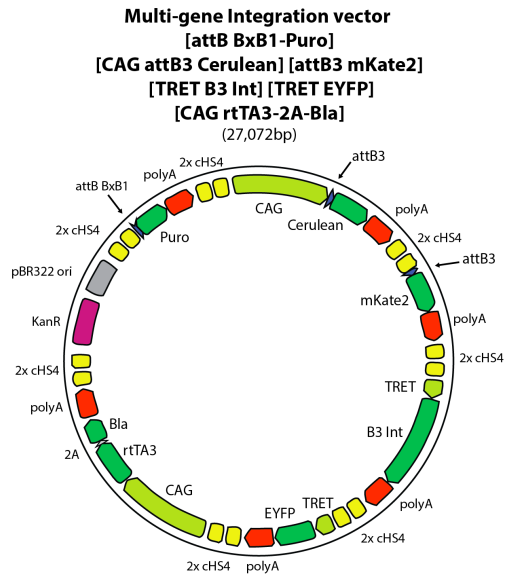
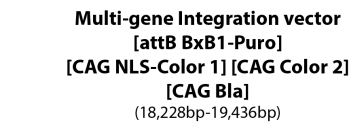
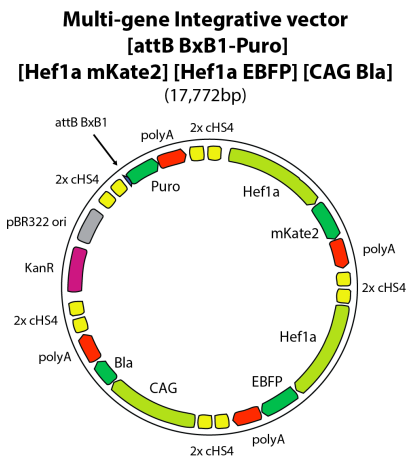
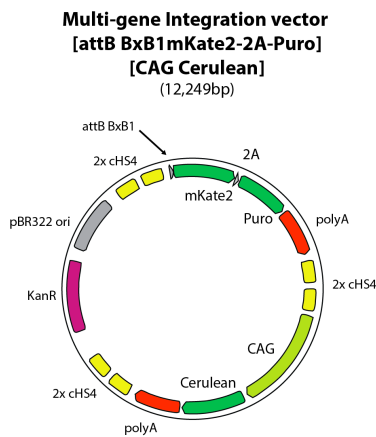
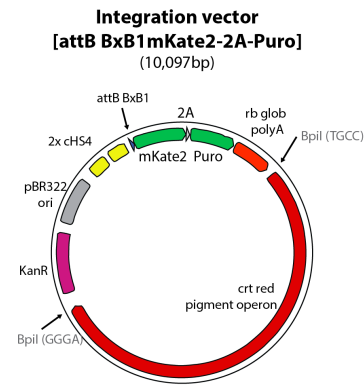
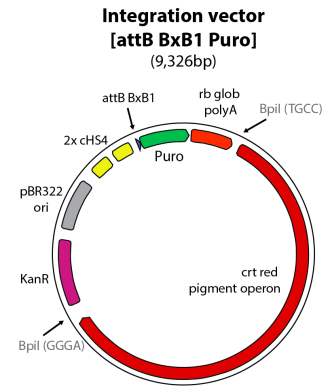
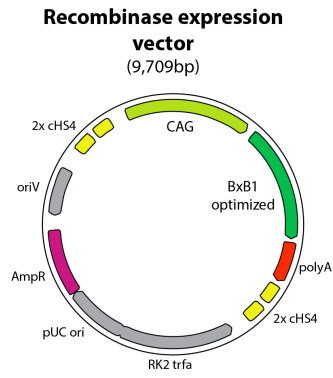
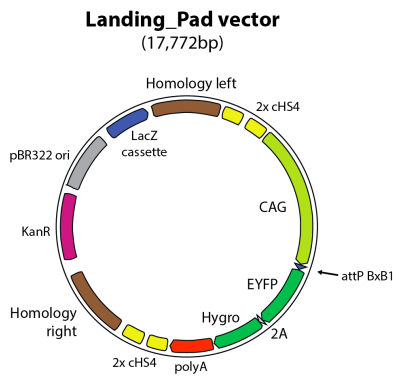
(A) Annotated map of the fully assembled circuit. **(B)** Representative FACS analysis histograms of resistant polyclonal population after integration of the 7-gene payload and selection with Puromycin for 7 days. About 40% of the cells have already excised the Cerulean transcription unit and switched to the final mKate2 expressing state. Given that essentially no cells are in the intermediate state, we hypothesized the switch occurred soon after transfection (complete degradation of the cerulean fluorescent protein takes up to a week). Leakiness of the TRE promoter from the transfected plasmids in the presence of high levels of rtTA3 (constitutively expressed) was sufficient to trigger expression of B3 integrase at low levels. Even in absence of dox, this resulted in moderate excision of the Cerulean expression cassette in the pool of plasmids. The experiment in Figure 3 is therefore performed with cells harboring a non-rearranged payload (red negative sorted cells).

Supplementary Figure 2-14. Microscopy images of isolated colonies corresponding to the 18 different phenotypes of the circuit library.



Supplementary Figure 15. Annotated maps of vectors used in this study.





Supplementary Table 2-1. Comparison of integration efficiencies of site-specific recombinases.

Recombinase	Integration efficiency	Cell type	Reference
BxB1	$\sim 1 \times 10^{-1}$	HEK293	This study
PhiC31	$\sim 2 \times 10^{-3}$	HEK293	2
R4	$\sim 1 \times 10^{-4}$	HEK293	3
Cre	$\sim 1 \times 10^{-5}$	HEK293	4
Flp	$\sim 1 \times 10^{-6}$	HEK293	5

REFERENCES

1. Hockemeyer, D. *et al. Nat Biotechnol* **27**, 851–857 (2009).
2. Thyagarajan, B. *et al. Molecular and Cellular Biology* **21**, 3926–3934 (2001).
3. Olivares, E. C. *et al. Gene* **278**, 167–176 (2001).
4. Fukushige, S., Fukushige, S., Sauer, B. & Sauer, B. *Proc Natl Acad Sci USA* **89**, 7905–7909 (1992).
5. O’Gorman, S., Fox, D. T. & Wahl, G. M. *Science* **251**, 1351–1355 (1991).

Supplementary Text 2-1. DNA sequences of the probes used for the Southern Blots.

YFP (Landing Pad integration test) :

caaagaattcgaccgtggttctgtctggtcaaccaccggtctcagtggtgtacggtacaaaccaccaagttgtacaaaaagcagggatc
cccgggtaccggtgccaccatggtgagcaagggcgaggagctgtcaccgggtggtgccatcctggtcgagctggacggcgacgtaaa
cggccacaagttcagcgtgtccggcgagggcgagggcgatgccacctacggcaagctgacctgaagttcatctgcaccaccggcaagctg
ccggtgccctggcccaccctcgtga

KanR (circuit integration test):

gggatcgagtggtgagtaaccatgcatcatcaggagtacggataaaatgcttgatggtcgaagaggcataaattccgtcagccagtttagtc
tgaccatctcatctgtaacatcattggcaacgctaccttggcatgtttcagaacaactctggcgatcgggctccatacaagcgatagattgt
cgcacctgattgcccacattatcgcgagcccatttataccatataaatcagcatcattgttgaatttaatcgcgggcctcgacgtttcccgttga
atatggctcat

Supplementary Text 2-2. DNA sequences of the primers used for genomic PCRs.

XD 621 Forward AAVS1 Primer: 5- GGCCCTGGCCATTGTCACTT -3

XD 622 Forward AAVS1 CEL-I Primer: 5- TTCGGGTCACCTCTCACTCC -3

XD 623 Reverse AAVS1 CEL-I Primer: 5- GGCTCCATCGTAAGCAAACC -3

LW06-LP-RV: 5-ACTCGAGAAATTCGGAGCCAAC-3

Supplementary Text 2-3. DNA sequences of the genetic components from our mammalian part library used in this study.

Pars int p_Insulator:

p_Insulator [2xcHS4 core fw]

CCTCGAGACAATTGATTAACATCGATACGGTACCGAGTTGGCGCGCCTGGGAGCTCACGGGGACAGCCCCCCCCAAAGCCC
CCAGGGATGTAATTACGTCCCTCCCCGCTAGGGGGCAGCAGCGAGCCGCCGGGGCTCCGCTCCGGTCCGGCGCTCCCCC
GCATCCCCGAGCCGGCAGCGTGCGGGGACAGCCCGGGCACGGGGAAGGTGGCACGGGATCGCTTTCTCTGAACGTTCTCG
CTGCTCTTTGAGCCTGCAGACACCTGGGGGGATACGGGAAAAAGCTTTAGGCTGAAAGAGAGATTTAGAATGACAGGCGCG
CCTGGCCATACATCGATACGGTACCGAGTTGGCGCGCCTGGGAGCTCACGGGGACAGCCCCCCCCAAAGCCCCAGGGATG
TAATTACGTCCCTCCCCGCTAGGGGGCAGCAGCGAGCCGCCGGGGCTCCGCTCCGGTCCGGCGCTCCCCCGCATCCCCGA
GCCGGCAGCGTGCGGGGACAGCCCGGGCACGGGGAAGGTGGCACGGGATCGCTTTCTCTGAACGTTCTCGCTGCTCTTTG
AGCCTGCAGACACCTGGGGGGATACGGGAAAAAGCTTTAGGCTGAAAGAGAGATTTAGAATGACAGGCGCGCCTGGTGGC
GAACGATCAAC

p_Insulator [2xcHS4 core rev]

CCTCGAGACAATTGATTAACATCGATACGGTACCGAGTTGGCGCGCCTGGGAGCTCACGGGGACAGCCCCCCCCAAAGCCC
CCAGGGATGTAATTACGTCCCTCCCCGCTAGGGGGCAGCAGCGAGCCGCCGGGGCTCCGCTCCGGTCCGGCGCTCCCCC
GCATCCCCGAGCCGGCAGCGTGCGGGGACAGCCCGGGCACGGGGAAGGTGGCACGGGATCGCTTTCTCTGAACGTTCTCG
CTGCTCTTTGAGCCTGCAGACACCTGGGGGGATACGGGAAAAAGCTTTAGGCTGAAAGAGAGATTTAGAATGACAGGCGCG
CCTGGCCATACATCGATACGGTACCGAGTTGGCGCGCCTGGGAGCTCACGGGGACAGCCCCCCCCAAAGCCCCAGGGATG
TAATTACGTCCCTCCCCGCTAGGGGGCAGCAGCGAGCCGCCGGGGCTCCGCTCCGGTCCGGCGCTCCCCCGCATCCCCGA
GCCGGCAGCGTGCGGGGACAGCCCGGGCACGGGGAAGGTGGCACGGGATCGCTTTCTCTGAACGTTCTCGCTGCTCTTTG
AGCCTGCAGACACCTGGGGGGATACGGGAAAAAGCTTTAGGCTGAAAGAGAGATTTAGAATGACAGGCGCGCCTGGTGGC
GAACGATCAAC

p_Insulator [inert sequence]

CCTCGAGACAATTGATTAACATCGATACGGTACCGAGTTGGCGCGCCTGGGAGCTCACGGGGACAGCCCCCCCCAAAGCCC
CCAGGGATGTAATTACGTCCCTCCCCGCTAGGGGGCAGCAGCGAGCCGCCGGGGCTCCGCTCCGGTCCGGCGCTCCCCC
GCATCCCCGAGCCGGCAGCGTGCGGGGACAGCCCGGGCACGGGGAAGGTGGCACGGGATCGCTTTCTCTGAACGTTCTCG
CTGCTCTTTGAGCCTGCAGACACCTGGGGGGATACGGGAAAAAGCTTTAGGCTGAAAGAGAGATTTAGAATGACAGGCGCG
CCTGGCCATACATCGATACGGTACCGAGTTGGCGCGCCTGGGAGCTCACGGGGACAGCCCCCCCCAAAGCCCCAGGGATG
TAATTACGTCCCTCCCCGCTAGGGGGCAGCAGCGAGCCGCCGGGGCTCCGCTCCGGTCCGGCGCTCCCCCGCATCCCCGA
GCCGGCAGCGTGCGGGGACAGCCCGGGCACGGGGAAGGTGGCACGGGATCGCTTTCTCTGAACGTTCTCGCTGCTCTTTG
AGCCTGCAGACACCTGGGGGGATACGGGAAAAAGCTTTAGGCTGAAAGAGAGATTTAGAATGACAGGCGCGCCTGGTGGC
GAACGATCAAC

Promoters:

Ubc

TTCGCCCTTGCTAGCTTAAGTGATCTGGCCTCCGCGCCGGGTTTTGGCGCCTCCCGCGGGCGCCCCCTCTCACGGCGAGCG
CTGCCACGTAGACGAAGGGCGCAGCGAGCGTCTGATCCTTCCGCCCGGACGCTCAGGACAGCGGCCCGTCTCATAAGA
CTCGGCCTTAGAACCCAGTATCAGCAGAAGGACATTTTAGGACGGGACTTGGGTGACTCTAGGGCACTGGTTTTCTTCCAG
AGAGCGGAACAGGGCGAGGAAAAGTAGTCCCTTCTCGGCGATTCTGCGGAGGGATCTCCGTGGGGCGGTGAACGCCGATGATT
ATATAAGGACGCGCCGGGTGTGGCACAGCTAGTCCGTGCGAGCCGGGATTTGGGTGCGGGTCTTGTGTTGTGGATCGCTGTG
ATCGTCACTTGGTGAGTAGCGGGCTGTGGGCTGGCCGGGGCTTTCGTGGCCCGGGCCGCTCGGTGGGACGGAAGCGTGT
GGAGAGACCCCAAGGGCTGTAGTCTGGGTCCGCGAGCAAGTTGCCCTGAACTGGGGGTGGGGGAGCGCAGCAAAATG
GCGGCTGTTCCCGAGTCTTGAATGGAAGACGCTTGTGAGGCGGGCTGTGAGGTCGTTGAAACAAGGTGGGGGCATGGTGGG
CGGCAAGAACCAAGGTCTGAGGCCTTCGCTAATGCGGGAAAGCTTATTTCGGGTGAGATGGGCTGGGACCATCTGGG
GACCCTGACGTGAAGTTTGTCACTGACTGGAGAACTCGGTTTGTCTGTGTTGCGGGGGCGGCAGTTATGGCGGTGCCGTTGG
GCAGTGCACCCGTACCTTTGGGAGCGCGCCCTCGTCTGTGTCGTGACGTCACCCGTTCTGTTGGCTTATAATGCAGGGTGGG
GCCACCTGCCGGTAGGTGTGCGGTAGGCTTTTCTCCGTGCGAGGACGCAGGGTTCGGGCCTAGGGTAGGCTCTCTGAATCGA
CAGGCGCCGACCTCTGGTGAGGGGAGGGATAAGTGAGGCGTCAGTTTCTTGGTTCGTTTTATGTACCTATCTCTTAAGTA
GCTGAAGCTCCGGTTTTGAACTATGCGCTCGGGGTGGCGAGTGTGTTTTGTAAGTTTTTAGGCACCTTTTGAATGTAATC
ATTTGGGTCAATATGTAATTTTCAAGTGTAGACTAGTAAATTGTCGCTAAATTCTGGCCGTTTTTGGCTTTTTTGTAGACGA
AGCTTGGGC

Hefla

TTAATTGGCTCCGGTGCCCGTCAGTGGGCAGAGCGCACATCGCCACAGTCCCCGAGAAGTTGTGGGGAGGGGTCCGCCAATT
GAACCGGTGCCTAGAGAAGGTGGCGCGGGGTAAGTGGGAAAGTGATGTCGTGTACTGGCTCCGCCTTTTTCCCGAGGGTGG
GGGAGAACCGTATATAAGTGCAGTAGTCGCCGTGAACGTTCTTTTTCGCAACGGGTTTGCCGCCAGAACACAGGTAAGTGCC
GTGTGTGGTTCCCGCGGGCCTGGCCTTTTACGGGTATAGGCCCTTGCGTGCCTTGAATTACTTCCACCTGGCTGCAGTACGTG
ATTCTTGATCCCGAGCTTCGGGTTGGAAGTGGGTGGGAGAGTTCGAGGCCCTTGCGCTTAAGGAGCCCTTCGCCTCGTGCTTG
AGTTGAGGCTGGCCTGGGCGCTGGGGCCCGCGTGCGAATCTGGTGGCACCTTCGCGCTGTCTCGCTGCTTTGATAAGT
CTCTAGCCATTTAAAATTTTTGATGACCTGCTGCGACGCTTTTTTTCTGGCAAGATAGTCTTGTAATGCGGGCCAAGATCTGC
ACACTGGTATTTTCGTTTTTTGGGGCCGCGGGCGGCGACGGGGCCCGTGCCTCCAGCGCACATGTTCCGGCAGGCGGGGCCT
GCGAGCGCGCCACCGAGAATCGGACGGGGTAGTCTCAAGCTGGCCGGCCTGCTCTGGTGCCTGGCCTCGCGCCCGCTGT
ATCGCCCCGCCCTGGGCGCAAGGCTGGCCCGTGGCACCAGTTGCGTGAGCGGAAAGATGGCCGCTTCCCGGCCCTGCTG
CAGGGAGCTCAAAATGGAGGACGCGCGCTCGGGAGAGCGGGCGGGTGAAGTACCCACACAAAGGAAAAGGGCCTTCCGT
CCTACGCCGTGCTTCATGTGACTCCACGGAGTACCGGGCGCGTCCAGGCACCTCGATTAGTCTCGAGCTTTGGAGTACG
TCGTCTTTAGGTTGGGGGAGGGGTTTTATGCGATGGAGTTTCCCCACACTGAGTGGGTGGAGACTGAAGTTAGGCCAGCTTG
GCACTTGATGTAATTTCTCCTTGGAAATTTGCCCTTTTTGAGTTGGATCTTGGTTCATTCTCAAGCCTCAGACAGTGTTCAAAG
TTTTTTCTTCCATTTCAAGTGGATCCCAAGG

CAG

GTCGACCTCGAGTTAATTAATGTGCTGCAAGGCGATTAAGTTGGGTAACGCCAGGGTTTTCCAGTCACGACGTTGTA AAAAC
GACGGCCAGTGAATTGTAATACGACTACTATAGGGCGAATTGGAGCTAGGCCTACGTAGCGCGGAGCTCCACCGCGGTGG
ACTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGGTAAA
TGGCCCGCCTGGCTGACCGCCCAACGACCCCGCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGGG
ACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTAC
GCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCTGGCATTATGCCAGTACATGACCTTATGGGACTTTCCTACTTGGC
AGTACATCTACGTATTAGTCATCGCTATTACCATGGTGCAGGTGAGCCCCAGTTCTGCTTACTTCCCCATCTCCCCCCT
CCCCACCCCAATTTGTATTTATTTATTTTAAATTTTTTGTGCAGCGATGGGGCGGGGGGGGGGGGGCGCGCCAGG
CGGGCGGGGCGGGGCGAGGGGCGGGGCGGGGCGAGGCGGAGAGGTGCGGCGGCAGCCAATCAGAGCGGCGGCTCCGAA
AGTTTCCTTTATGGCGAGGCGGGCGGGCGGGCCCTATAAAAAGCGAAGCGCGGGCGGGGAGTGCAGTGCCTGCG
TTGCCCTCGCCCGTGCCCGCTCCGCGCCGCTCGCGCCCGCCCGGGCTCTGACTGACCGGTTACTCCACAGGTGAG

CGGGCGGGACGGCCCTTCTCCTCCGGGCTGTAATTAGCGCTTGGTTAATGACGGCTCGTTCTTTTCTGTGGCTGCGTGAAAG
CCTTAAAGGGCTCCGGGAGGGCCCTTTGTGCGGGGGGAGCGGCTCGGGGGTGCCTGCGTGTGTGTGTGCGTGGGGAGCGC
CGCGTGCGCCCGCGCTGCCCGGGCTGTGAGCGCTGCGGGCGCGCGCGGGGCTTTGTGCGCTCCGCGTGTGCGGAGGG
GAGCGCGCCGGGGGCGGTGCCCGCGGTGCGGGGGGGTGCAGGGGAACAAAGGCTGCGTGCGGGGTGTGTGCGTGGGG
GGGTGAGCAGGGGGTGTGGGCGCGCGGTGCGGGCTGTAACCCCCCTGCACCCCCCTCCCGAGTTGCTGAGCACGGCCCG
GCTTCGGGTGCGGGCTCCGTGCGGGGCGTGGCGCGGGGCTCGCCGTGCCGGGCGGGGGTGGCGGCAGGTGGGGGTGCCG
GGCGGGGCGGGGCCCTCGGGCCGGGAGGGCTCGGGGGAGGGGCGCGGGCCCGGAGCGCCGGGGCTGTCGAGGC
GCGGCGAGCCGAGCCATTGCTTTTATGGTAATCGTGCAGAGGGGCGCAGGGACTTCTTTGTCCAAATCTGGCGGAGCC
GAAATCTGGGAGGCGCCGCCACCCCTCTAGCGGGCGGGCGAAGCGGTGCGGCGCCGCGAGGAAGGAAATGGGCGGG
GAGGGCTTCGTGCGTCCCGCGCCGCTCCCTTCTCCATCTCCAGCCTCGGGGCTGCCGAGGGGGACGGCTGCCTTCGG
GGGGACGGGGCAGGGCGGGTTCGGCTTCTGGCGTGTGACCGGGCTCTAGAGCCTCTGCTAACCATGTCATGCCTTCTT
CTTTTCTACAGCTCCTGGGCAACGTGCTGGTTGTTGTGCTGTCTCATCATTTGGCAAAGAATTCTGTAC

TRE-tight:

ACTAATTCGCCCTTCAGGTCCGCTAGACGAGTTTACTCCCTATCAGTGATAGAGAACGATGTCGAGTTTACTCCCTATCAGTG
ATAGAGAACGATATGTCGAGTTTACTCCCTATCAGTGATAGAGAACGATGTCGAGTTTACTCCCTATCAGTGATAGAGAACGT
ATGTCGAGTTTACTCCCTATCAGTGATAGAGAACGATGTCGAGTTTACTCCCTATCAGTGATAGAGAACGATGTCGAGGTAG
GCGTGTACGGTGGGAGGCCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCAAAGGGCGAATTCGACCCAAGTTG
TACAATCACTCTTCTGGTGAC

Inert 3' UTR

CCACTGGATTGTACAATTAC

Genes (with Kozak sequence):

BxB1 wild-type recombinase

GCCGCCACCATGAGAGCCCTGGTAGTCATCCGCTGTCCCGCTCACCGATGCTACGACTTACCCGGAGCGTCAGCTGGAGTC
TTGCCAGCAGCTCTGCGCCAGCGCGGTGGGACGTCGTCGGGGTAGCGGAGGATCTGGACGTCTCCGGGGCGGTGCGATCCG
TTCGACCGGAAGCGCAGACCGAACCTGGCCCGTGGCTAGCGTTTCGAGGAGCAACCGTTTCGACGTGATCGTGGCGTACCGGG
TAGACCGTTGACCCGATCGATCCGGCATCTGCAGCAGCTGGTCCACTGGGCCGAGGACCACAAGAAGCTGGTCTGCTCCGC
GACCGAAGCGCACTTCGATACGACGACCGCTTTCGCGCGTTCATCGCGCTTATGGGAACGGTGGCGCAGATGGAATTA
GAAGCGATCAAAGAGCGGAACCGTTCGGCTGCGCATTTCAATATCCGCGCCGGGAAATACCGAGGATCCCTGCCCGCTGGG
GATACCTGCCTACGCGGTGGACGGGGAGTGGCGGTGGTGCCGACCTGTGCAGCGAGAGCGCATCCTCGAGGTGTATCA
CCGCGTGCCTGACAACCACGAGCCGCTGCACCTGGTGGCCACGACCTGAACCGCGTGGTGTCTGTCGCCGAAGGACTAC
TTCGCGCAGCTGCAAGGCCGCGAGCCGCGAGGGCCGGGAGTGGTTCGGCTACCGCGCTGAAGCGATCGATGATCTCCGAGGCGA
TGCTCGGGTACGCGACTCTGAACCGTAAGACCGTCCGAGACGACGACGGAGCCCCGCTGGTGCGGGTGAGCCGATCCTGAC
CCGTGAGCAGCTGGAGGCGTGCAGCGTGCAGCTCGTCAAGACCTCCGGGCGAAGCCCGCGGTGTCTACCCCGTGCCTGCTG
CTGCGGGTGTGTTCTGCGCGGTGTGCGGGGAGCCCGGTACAAGTTCCCGGGGGAGGACGTAAGCACCCGCGCTACCGCT
GCCGCTCGATGGGGTTCGGAAGCACTGCGGGAACGGCACGGTGGCGATGGCCGAGTGGGACGCGTTCTGCGAGGAGCAGG
TGCTGGATCTGCTCGGGACGCGGAGCGTCTGGAGAAAGTCTGGGTAGCCGGCTCGGACTCCGCGGTGAACTCGCGGAGGT
GAACGCGGAGCTGGTGGACCTGACGTCGCTGATCGGCTCCCCGGCTACCGGGCCGGCTCTCCGACGCGAGAAGCACTGGAT
GCCCGTATTGCGGGCTGGCCGCGCGGAAGAGGAGCTGGAGGGTCTAGAGGCTCGCCCGTCTGGCTGGGAGTGGCGCGAG
ACTGGGACGCGTTTCGGGACTGGTGGCGGGAGCAGGACACCGCGGCAAAGAACACCTGGCTTCGGTTCGATGAACGTTCCGG
CTGACGTTGACGTCGCGGGGCTGACTCGCACGATCGACTTCGGGGATCTGCAGGAGTACGAGCAGCATCTCAGGCTCG
GCAGCGTGGTCAACGGCTACACACCGGGATGTCGTAG

BxB1 mammalian codon optimized recombinase

GCCGCCACCATGAGAGCACTGGTGGTTCATCCGACTGAGTAGGGTACAGACGCAACAACAAGCCCCGAGAGGCAGCTGGAA
TCATGTCAGCAGCTGTGCGCACAGCGAGGATGGGACGTGGTTCGGAGTGGCAGAGGATCTGGACGTGAGCGGGCGCTGTCGATC

CATTCGACAGAAAGCGGAGGCCAACCTGGCAAGGTGGCTGGCTTTCGAGGAACAGCCCTTGATGTGATCGTCGCCTACAG
AGTGGACAGGCTGACACGCTCTATTTCGACATCTGCAGCAGCTGGTGCATTGGGCCGAGGACCACAAGAACTGGTGGTCAGT
GCAACTGAAGCCCACTTCGATAACCACAACCTCTTTTGGCCGCTGTGGTCACTGCACTGATGGGCACCGTGGCCAGATGGAGCT
GGAAGCTATCAAGGAGCGAAACCGGAGTGCAGCCATTTCAATATTCGGGCCGGGAAATACAGAGGATCACTGCCCCCTTGG
GGCTATCTGCCTACCCGGGTGGATGGGGAGTGGAGACTGGTGCCAGACCCCGTCCAGAGAGAGAGGATTCTGGAAGTGATACC
ACAGGGTGGTCGATAACCACGAACCACTGCATCTGGTCGCCCACGACCTGAATAGGCGCGGCGTGTGAGCCAAAAGATTA
TTTTGCTCAGCTGCAGGGAAGGGAGCCACAGGGACGAGAATGGTCCGCTACCGCCCTGAAGCGGAGCATGATCAGTGAGGCT
ATGCTGGGCTACGCAACTCTGAATGGGAAAACCGTCCGGGACGATGACGGAGCACCCTGGTGAGGGCTGAGCCTATTCTGA
CACGCGAGCAGCTGGAAGCTCTGCGGGCAGAACTGGTGAACCTCCAGAGCCAAACCTGCCGTGAGCACCCAAAGCTGCT
GCTGAGGGTGTGTTCTGCGCCGCTGTGGGGAGCCAGCATAACAAGTTTGGCCGCGGGGAAAGAAAACATCCCCGCTATCGA
TGCCGGTCTATGGGATTCCTAAGCACTGTGGAACCGCACTGTGGCTATGGCCGAGTGGGACGCCTTTTGTGAGGAACAGG
TGCTGGATCTGCTGGGAGACGCCGAGAGGCTGAAAAAGTGTGGTTCGCTGGCAGCGACTCCGCTGTGGAGCTGGCAGAAGT
CAATGCCGAGCTGGTGGATCTGACCTCCCTGATCGGATCTCTGCATATAGGGCAGGCTCACCACAGCGAGAAGCTCTGGAC
GCACGAATTGCTGCACTGGCAGCTCGACAGGAGGAACTGGAGGGGCTGGAAGCACGACCTAGCGGATGGGAGTGGCGAGAA
ACAGGCCAGCGTTTGGGGATTGGTGGAGAGAGCAGGACACAGCAGCCAAGAACAACCTTGGCTGAGAAGTATGAATGTCAGG
CTGACTTTCGATGTGCGCGCGGGCTGACCCGAACAATCGATTTTGGCGACCTGCAGGAGTATGAACAGCACCTGAGACTGG
GGAGCGTGGTCGAAAGACTGCACACTGGGATGTCATAGTAA

BxB1 mammalian codon optimized recombinase with N-terminal NLS

GCCGCCACCATGCCAAAGAAGAAACGCAAAGTTAGAGCACTGGTGGTCATCCGACTGAGTAGGGTTCACAGACGCAACAACA
AGCCCCGAGAGGCAGCTGGAATCATGTGCAGCAGCTGTGCGCACAGCGAGGATGGGACGTGGTTCGAGTGGCAGAGGATCTG
GACGTGAGCGGCGCTGTCGATCCATTCGACAGAAAGCGGAGGCCAACCTGGCAAGGTGGCTGGCTTTCGAGGAACAGCCCT
TTGATGTGATCGTCGCCTACAGAGTGGACAGGCTGACACGCTCTATTTCGACATCTGCAGCAGCTGGTGCATTGGGCCGAGGA
CCACAAGAACTGGTGGTCACTGCAACTGAAGCCCACTTCGATAACCACAACCTCTTTTGGCCGCTGTGGTCACTGATGG
GCACCGTGGCCAGATGGAGCTGGAAGCTATCAAGGAGCGAAACCGGAGTGCAGCCATTTCAATATTCGGGCCGGGAAAT
ACAGAGGATCACTGCCCCCTTGGGGCTATCTGCCTACCCGGGTGGATGGGGAGTGGAGACTGGTGCCAGACCCCGTCCAGAG
AGAGAGGATTCTGGAAGTGTAACACAGGGTGGTCGATAACCACGAACCACTGCATCTGGTTCGCCCACGACCTGAATAGGCGC
GGCGTGTGAGCCAAAAGATTATTTTGTCTCAGCTGCAGGGAAGGGAGCCACAGGGACGAGAATGGTCCGCTACCGCCCTGA
AGCGGAGCATGATCAGTGAGGCTATGCTGGGCTACGCAACTCTGAATGGGAAAACCGTCCGGGACGATGACGGAGCACCAC
TGGTGAGGGCTGAGCCTATTCTGACACGCGAGCAGCTGGAAGCTCTGCGGGCAGAACTGGTGAACCTCCAGACCCAAAC
TGCCGTGAGCACCCAAAGCTGCTGCTGAGGGTGTGTTCTGCGCCGCTGTGGGGAGCCAGCATAACAAGTTTGGCCGCGGG
GGAAGAAAACATCCCCGCTATCGATGCCGGTCTATGGGATTCCTAAGCACTGTGGAACCGCACTGTGGCTATGGCCGAGT
GGGACGCCTTTTGTGAGGAACAGGTGCTGGATCTGCTGGGAGACGCCGAGAGGCTGGAAAAAGTGTGGGTGCTGGCAGCG
ACTCCGCTGTGGAGCTGGCAGAAGTCAATGCCGAGCTGGTGGATCTGACCTCCCTGATCGGATCTCTGCATATAGGGCAGG
CTCACCACAGCGAGAAGCTCTGGACGCACGAATTGCTGCACTGGCAGCTCGACAGGAGGAACTGGAGGGGCTGGAAGCACG
ACCTAGCGGATGGGAGTGGCGAGAAAACAGGCCAGCGGTTTGGGGATTGGTGGAGAGAGCAGGACACAGCAGCCAAGAACA
TTGGCTGAGAAGTATGAATGTCAGGCTGACTTTCGATGTGCGCGCGGGCTGACCCGAACAATCGATTTTGGCGACCTGCAG
GAGTATGAACAGCACCTGAGACTGGGGAGCGTGGTCGAAAGACTGCACACTGGGATGTCATAG

BxB1 mammalian codon optimized recombinase with C-terminal NLS

GCCGCCACCATGAGAGCACTGGTGGTTCATCCGACTGAGTAGGGTTCACAGACGCAACAACAAGCCCCGAGAGGCAGCTGGAA
TCATGTGCAGCAGCTGTGCGCACAGCGAGGATGGGACGTGGTTCGAGTGGCAGAGGATCTGGACGTGAGCGGCGCTGTCGATC
CATTCGACAGAAAGCGGAGGCCAACCTGGCAAGGTGGCTGGCTTTCGAGGAACAGCCCTTTGATGTGATCGTCGCCTACAG
AGTGGACAGGCTGACACGCTCTATTTCGACATCTGCAGCAGCTGGTGCATTGGGCCGAGGACCACAAGAACTGGTGGTCAGT
GCAACTGAAGCCCACTTCGATAACCACAACCTCTTTTGGCCGCTGTGGTCACTGCACTGATGGGCACCGTGGCCAGATGGAGCT
GGAAGCTATCAAGGAGCGAAACCGGAGTGCAGCCATTTCAATATTCGGGCCGGGAAATACAGAGGATCACTGCCCCCTTGG
GGCTATCTGCCTACCCGGGTGGATGGGGAGTGGAGACTGGTGCCAGACCCCGTCCAGAGAGAGAGGATTCTGGAAGTGATACC
ACAGGGTGGTCGATAACCACGAACCACTGCATCTGGTTCGCCCACGACCTGAATAGGCGCGGCGTGTGAGCCAAAAGATTA

TTTTGCTCAGCTGCAGGGAAGGGAGCCACAGGGACGAGAATGGTCCGCTACCGCCCTGAAGCGGAGCATGATCAGTGAGGCT
ATGCTGGGCTACGCAACTCTGAATGGGAAAACCGTCCGGGACGATGACGGAGCACCCTGGTGAGGGCTGAGCCTATTCTGA
CACCGGAGCAGCTGGAAGCTCTGCGGGCAGAAGCTGGTGAACCTCCAGAGCCAAACCTGCCGTGAGACCCCAAGCCTGCT
GCTGAGGGTGTGTTCTGCGCCGTCTGTGGGGAGCCAGCATAACAAGTTGCCGGCGGGGGAAGAAAACATCCCCGCTATCGA
TGCCGGTCTATGGGATTCCCTAAGCACTGTGGAACGGCACTGTGGCTATGGCCGAGTGGGACGCCTTTTGTGAGGAACAGG
TGCTGGATCTGCTGGGAGACGCCGAGAGGCTGGAAAAAGTGTGGGTGCTGGCAGCGACTCCGCTGTGGAGCTGGCAGAAGT
CAATGCCGAGCTGGTGGATCTGACCTCCCTGATCGGATCTCTGCATATAGGGCAGGCTCACACAGCGAGAAGCTCTGGAC
GCACGAATTGCTGCACTGGCAGCTCGACAGGAGAACTGGAGGGGGCTGGAAGCACGACCTAGCGGATGGGAGTGGCGAGAA
ACAGGCCAGCGTTTGGGGATTGGTGGAGAGAGCAGGACACAGCAGCCAAGAACAACCTTGGCTGAGAAGTATGAATGTCAGG
CTGACTTTTCGATGTGCGCGGGGGCTGACCCGAACAATCGATTTTGGCGACCTGCAGGAGTATGAACAGCACCTGAGACTGG
GGAGCGTGGTGCAAAAGACTGCACACTGGGATGTCACCAAAAAAAAAAACGCAAAGTTTAG

B3 Recombinase

GCCGCCACCATGAGCTCGTATATGGATCTTGTTGATGATGAACCAGCGACTTTGTACCATAAGTTCGTGGAGTGCTTGAAAGC
GGGCGAGAACTTCTGCGGAGACAAGCTGAGTGGAAATTATTACCATGGCGATCCTTAAGGCAATCAAGGCGCTGACCGAAGTT
AAAAAGACAACCTTTAACAAATATAAGACAACAATCAAGCAGGGCCTCCAGTATGACGTGGTTTCGTCGACTATCTCGTTT
TGTATCACTTGAAGGACTGTGATGAGCTGTCCAGGGGGCTTGAGCGATGCCTTCGAGCCCTACAAATTCAAATTAAGTCGAAT
AAAGAGGCAACCTCGTTAAGACTCTCTTTCGTGGCCCTCGTTTGGCAGCCAGAAGAAGTGGCGGAAGAAAGAGGTGGACC
GCGAGGTGGATAAAGTTCATAGCACCAGACAGACGAATCGATTTTCAAATTCATCTTGAACAGTTGGATAGTATTGAG
ACACAAACGAAACAGGATCGCCAAAAGACGGTGTGACTTTTCATCTGTTGATGACATTTTTCAAAGTGTAGGAACAATG
ACCTGATGAACGTTGATCCCTCCACATTTAAGATTGTGAAAAACAATTCGTCGGATACCTGCTGCAGGCTGAGGTCAAACA
GACTAAGACACGCAAGTCGAGGAACATTTCTTCTTCCATCCGCGAGAATCGATTGATCTGTTCTGGCCTTGCACGATTT
CTTCCGCACATGCCAGCCTACCCAAAAGTCGCGTCTTTCGGATCAAGTATCGGAGCAGAAGTGGCAGCTTTTCCGAGATTCCA
TGTCATTGATTACAACCGTTTCTTTAGGAAGTTTCAGCTTCGCTATTTTCGAATAAAGCACGGCCCAAGTCCCATCTGG
GCCGGCATCTGATGAACAGCTTTCTGCACAAGAATGAACTGGATTCTGGGCCAACTCCCTGGGCAACTGGAGCTCCTCCAG
AATCAACGCGAGTCCGGTGCAGGACTGGGCTACACCCACGGTGGTGCAGATTTGCCACAGCCACTGTTCCGATTCCTGGCTG
GCTACTGCGTTCGAAATGAAGAGGGCCACATCGTGGGCTTGGGTCTGGAGAAGGACATTAATGATCTGTTGACGGTATTATG
GACCCACTTAATGAGAAGGAGGATACAGAGATTTGTGAAAGCTACGGCGAGTGGGCCAAAATTGTGAGCAAGGATGTTCTG
ATATTTCTGAAGCGATATCATTGGAAGACGCATGCCGGCGATATCAGAATCCACATTGTACGCACGTACGTTCTTAAAGC
AGAGTCCGTCACCTTGAGCGGCTCCAAGGGAAGCGAAGAGCCGAGCAGTCCCGTCAGGATTCCAATACTTAGTATGGGAAAA
GCCTCCCAAGCGAGGGCCGAAAGTTGCGTGCCAGCGAACATGCCAACGATGATAATGAGATTGAGAAGATAGATTCCGATT
CGTCGCAAAGCGAAGAGATACCTATCGAAATGTCCGACTCCGAGGACGAAACAACGGCAAGTAACATTAGCGGAATTTACCT
GGACATGTCGAAAGCCAACCTCAACGTGGTCTATTTCCCTCCGAGTCAGACTGGCCGTGCCGCCGGTGCCGGCCGTAACGT
GGCGTGGGTGGTGCAGCACCGTTGAGTCGAAGCGGCGCGTCTGGCACCAATTAACCGGTGA

EYFP

GCCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCATCCTGGTCGAGCTGGACGGCGACGTAAACG
GCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGG
CAAGCTGCCCCTGCCCTGGCCACCCTCGTGACCACCTTCGGCTACGGCCTGCAGTGCTTCGCCGCTACCCCGACCATGA
AGCAGCACGACTTCTTCAAGTCCGCCATGCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTA
CAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGA
CGGCAACATCCTGGGGCACAAGCTGGAGTACAACATAACAGCCACAACGTCTATATCATGCCGACAAGCAGAAGAAGCG
CATCAAGGTGAAGTTCAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCC
ATCGGCGACGGCCCGTGTCTGCTGCCGACAACCACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGC
GCGATACATGGTCTGCTGGAGTTCGTGACCGCCGCCGGGATCACTCTCGCATGGACGAGCTGTACAAGTGA

mKate2

GCCGCCACCATGGTGAGCGAGCTGATTAAGGAGAACATGCACATGAAGCTGTACATGGAGGGCACCGTGAACAACCACCAC
TTCAAGTGCACATCCGAGGGCGAAGGCAAGCCCTACGAGGGCACCCAGACCATGAGAATCAAGGCGGTGAGGGGCGGCCCT
CTCCCCTTCGCCTTCGACATCCTGGCTACCAGCTTCATGTACGGCAGCAAAACCTTCATCAACCACACCCAGGGCATCCCCGA
CTTCTTTAAGCAGTCCCTCCCGAGGGCTTACATGGGAGAGAGTACCACATACGAAGATGGGGGCGTGTGACCGCTACC
CAGGACACCAGCTCCAGGACGGTGCCTCATCTACAACGTCAAGATCAGAGGGGTGAACTTCCATCCAACGGCCCTGTGA
TGCAGAAGAAAAACTCGGCTGGGAGGCCTCCACCGAGACACTGTACCCCGCTGACGGCGGCCTGGAAGGCAGAGCCGACA
TGGCCCTGAAGCTCGTGGGCGGGGGCCACCTGATCTGCAACCTTAAGACCACATACAGATCCAAGAAACCCGCTAAGAACCT
CAAGATGCCCGGCTACTATGTGGACAGGAGACTGGAAAGAATCAAGGAGCCGACAAAGAGACATACGTGCGAGCAGCA
CGAGGTGCTGTGGCCAGATACTGCGACCTCCCTAGCAAACCTGGGGCACAAACTTAATTCCTGA

EBFP

GCCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCATCCTGGTTCGAGCTGGACGGCGACGTAAACG
GCCACAAGTTCAGCGTGAGGGGCGAGGGGCGAGGGCGATGCCACCAACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCG
GCAAGCTGCCCCTGCCCTGGCCACCCTCGTGACCACCCTGAGCCACGGCGTGCAGTGTTCGCCCCTACCCCGACCACATG
AAGCAGCACGACTTCTCAAGTCCGCCATGCCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCACCT
ACAAGACCCGCGCGAGGTGAAGTTCGAGGGGCGACACCCTAGTGAACCGCATCGAGCTGAAGGGCGTGCAGTTCAGGAGG
ACGGCAACATCCTGGGGCACAAGCTGGAGTACAACCTTCAACAGCCACAACATCTATATCATGGCCGTCAGCAGAAGAACGG
CATCAAGGTGAACTTCAAGATCCGCCACAACGTGGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCC
ATCGGCGACGGCCCCGTGCTGTGCCGACGCCACTACCTGAGCACCCAGTCCGTGCTGAGCAAAGACCCCAACGAGAAGC
GCGATCACATGGTCTGCTGGAGTTCGCGACCGCCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTGA

Cerulean

GCCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCATCCTGGTTCGAGCTGGACGGCGACGTAAACG
GCCACAAGTTCAGCGTGTCCGGGCGAGGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGG
CAAGCTGCCCCTGCCCTGGCCACCCTCGTGACCACCCTGACCTGGGGCGTGCAGTGTTCGCCCCTACCCCGACCACATGA
AGCAGCACGACTTCTCAAGTCCGCCATGCCCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTTCAAGGACGACGGCAACTA
CAAGACCCGCGCGAGGTGAAGTTCGAGGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGA
CGGCAACATCCTGGGGCACAAGCTGGAGTACAACGCCATCAGCGACAACGTCTATATCACCGCCGACAAGCAGAAGAACGG
CATCAAGGCCAACTTCAAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCACTACCAGCAGAACACCCCC
ATCGGCGACGGCCCCGTGCTGTGCCGACAACCACTACCTGAGCACCCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGC
GCGATCACATGGTCTGCTGGAGTTCGTGACCGCCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGTGA

mKate-2A-Puro

GCCGCCACCATGGTGAGCGAGCTGATTAAGGAGAACATGCACATGAAGCTGTACATGGAGGGCACCGTGAACAACCACCAC
TTCAAGTGCACATCCGAGGGCGAAGGCAAGCCCTACGAGGGCACCCAGACCATGAGAATCAAGGCGGTGAGGGGCGGCCCT
CTCCCCTTCGCCTTCGACATCCTGGCTACCAGCTTCATGTACGGCAGCAAAACCTTCATCAACCACACCCAGGGCATCCCCGA
CTTCTTTAAGCAGTCCCTCCCGAGGGCTTACATGGGAGAGAGTACCACATACGAAGATGGGGGCGTGTGACCGCTACC
CAGGACACCAGCTCCAGGACGGTGCCTCATCTACAACGTCAAGATCAGAGGGGTGAACTTCCATCCAACGGCCCTGTGA
TGCAGAAGAAAAACTCGGCTGGGAGGCCTCCACCGAGACACTGTACCCCGCTGACGGCGGCCTGGAAGGCAGAGCCGACA
TGGCCCTGAAGCTCGTGGGCGGGGGCCACCTGATCTGCAACCTTAAGACCACATACAGATCCAAGAAACCCGCTAAGAACCT
CAAGATGCCCGGCTACTATGTGGACAGGAGACTGGAAAGAATCAAGGAGCCGACAAAGAGACATACGTGCGAGCAGCA
CGAGGTGGCTGTGGCCAGATACTGCGACCTCCCTAGCAAACCTGGGGCACAAACTTAATGGAAGCGGAGCTACTAACTTCAGC
CTGCTGAAGCAGGCTGGGCGAGTGGAGGAGAACCCTGGACCTACCATGACCGAGTACAAGCCACGGTGCGCCCTCGCCACCC
GCGACGACGTCCCCGGGCGTACGCACCCTCGCCGCCGCTTCGCCGACTACCCCGCCACGCGCCACACCGTGCACCCGGA
CCGCCACATCGAGCGGGTACCGAGCTGCAAGAACTTCTCTACGCGCGTGGGGCTCGACATCGGCAAGGTGTGGGTGCGG
GACGACGGCGCGGTTGGCGGTCTGGACCACGCCGAGAGCGTCAAGCGGGGGCGGTGTTCCGCCGAGATCGGCCCGCGC
ATGGCCGAGTTGAGCGGTTCCCGGCTGGCCGCGCAGCAACAGATGGAAGGCTCCTGGCGCCGACCCGGCCAAAGAGCCCG
CGTGGTCTGTCGCCACCGTCGGCGTCTCGCCGACCACAGGGCAAGGGTCTGGGCAGCGCCGTCGTGCTCCCCGAGTGGGA

GGCGGCCGAGCGCGCCGGGGTGCCCGCTTCTTGAGACATCCGCGCCCCGCAACCTCCCCTTCTACGAGCGGCTCGGCTTCA
CCGTCACCGCCGACGTCGAGGTGCCCGAAGGACCGCGCACCTGGTGCATGACCCGCAAGCCGGTGCCTGA

EYFP-2A-Hygro

GCCGCCACCATGGTGAGCAAGGGCGAGGAGCTGTTACCGGGGTGGTGCCATCCTGGTGCAGCTGGACGGCGACGTAAACG
GCCACAAGTTCAGCGTGTCCGGCGAGGGCGAGGGCGATGCCACCTACGGCAAGCTGACCCTGAAGTTCATCTGCACCACCGG
CAAGCTGCCCGTGCCTTGCCACCCCTCGTGACCACCTTCGGCTACGGCTGCAGTGCTTCGCCGCTACCCCGACCACATGA
AGCAGCAGACTTCTCAAGTCCGCCATGCCGAAGGCTACGTCCAGGAGCGCACCATCTTCTCAAGGACGACGGCAACTA
CAAGACCCGCGCCGAGGTGAAGTTCGAGGGCGACACCCTGGTGAACCGCATCGAGCTGAAGGGCATCGACTTCAAGGAGGA
CGGCAACATCCTGGGGCACAAGCTGGAGTACAACATAACAGCCACAACGTCTATATCATGGCCGACAAGCAGAAGAACGG
CATCAAGGTGAAGTTCAGATCCGCCACAACATCGAGGACGGCAGCGTGCAGCTCGCCGACCCTACCAGCAGAACACCCCG
ATCGGCGACGGCCCCGTGCTGCTGCCGACAACACTACCTGAGCTACCAGTCCAAGCTGAGCAAAGACCCCAACGAGAAGC
GCGATCACATGGTCTGCTGGAGTTCGTGACCGCCCGGGATCACTCTCGGCATGGACGAGCTGTACAAGGGAAGCGGAGC
TACTAACTTCAGCTGCTGAAGCAGGCTGGCGACGTGGAGGAGAACCCTGGACCTGCCACCATGAAAAAGCCTGAAGTCAAC
GCGACGTCTGTCGAGAAGTTTCTGATCGAAAAGTTCGACAGCGTCTCCGACCTGATGCAGCTCTCGGAGGGCGAAGAATCTC
GTGCTTTCAGCTTCGATGTAGGAGGGCGTGGATATGTCCTGCGGGTAAATAGCTGCGCCGATGGTTTCTACAAAGATCGTTAT
GTTTATCGGCACCTTTCATCGGCCGCGCTCCCGATTCCGGAAGTGTGACATTGGGGAATTCAGCGAGAGCCTGACCTATTG
CATCTCCCGCCGTGCACAGGGTGTACGTTGCAAGACCTGCTGAAACCGAACTGCCCGCTGTTCTGCAGCCGGTTCGCGGAG
GCCATGGATGCGATCGCTGCGGCCGATCTTAGCCAGACGAGCGGGTTCGGCCATTCGGACCGCAAGGAATCGGTCAATACA
CTACATGGCGTGATTCATATGCGCGATTGCTGATCCCCATGTGTATCACTGGCAAACCTGTGATGGACGACACCGTCAAGTGC
TCCGTGCGCAGGCTCTCGATGAGCTGATGCTTTGGGCGGAGGACTGCCCGAAGTCCGGCACCTCGTGCACGGGATTTCCG
CTCAACAATGTCTGACGGACAATGGCCGATAACAGCGGTCACTGACTGGAGCGAGGCGATGTTCCGGGATTCCTCAATAC
GAGGTCGCCAACATCTTCTTCTGGAGGCCGTGGTTGGCTTGTATGGAGCAGCAGACGCGCTACTTCGAGCGGAGGCATCCGG
AGCTTGCAGGATCGCCGCGGCTCCGGGCGTATATGCTCCGCATTGGTCTTGACCAACTCTATCAGAGCTTGGTTGACGGCAAT
TTCGATGATGCAGCTTGGGCGCAGGGTGCATGCGACGCAATCGTCCGATCCGGAGCCGGACTGTCCGGGCTACACAAATCG
CCCCGAGAAGCGCGGCCGTCTGACCGATGGCTGTGTAGAAGTACTCGCCGATAGTGAAAACCGACGCCCCAGCACTCGTCC
GGATCGGGAGATGGGGGAGGCTAACTGA

N-ter H2B localization tag

GCCACCATGCCAGAGCCAGCGAAGTCTGCTCCCGCCCCGAAAAAGGGTCCAAGAAGGCGGTGACTAAGGGCGAGAAGAAA
GGCGGCAAGAAGCGCAAGCGCAGCCGCAAGGAGAGCTATTCCATCTATGTGTACAAGGTTCTGAAGCAGGTCCACCCTGACA
CCGGCATTTCGTCCAAGGCCATGGGCATCATGAATTCGTTTGTGAACGACATTTTCGAGCGCATCGCAGGTGAGGCTTCCCGC
CTGGCGCATTACAACAAGCGCTCGACCATCACCTCCAGGGAGATCCAGACGGCCGTGCGCCTGCTGCTGCCTGGGGAGTTGG
CCAAGCACGCCGTGTCGAGGGTACTAAGGCCATCACCAAGTACACCAGCGCTAAG + linker:
GATCCCCGGGTACCGGTGCCACC

N-ter Smac mitochondria localization tag

GCCGCCACCATGGCGGCTCTGAAGAGTTGGCTGTCGCGCAGCGTAACTTCATTCTCAGGTACAGACAGTGTGTTGTGTGCC
TGTTGTGGCTAACTTTAAGAAGCGGTGTTTCTCAGAATTGATAAGACCATGGCACAACAACTGTGACGATTGGCTTTGGAGTAA
CCCTGTGT + linker: GGATCAGGAGGATCAGGA

C-ter Utrophin localization tag

Linker: GGATCAGGAGGATCAGGA +
TCCGGAACCATGGCCAAGTATGGAGAACATGAAGCCAGTCTGACAATGGGCAGAACGAATTCAGTGATATCATTAAAGTCCA
GATCTGATGAACACAATGACGTACAGAAGAAAACCTTTACCAAATGGATAAATGCTCGATTTTCAAAGAGTGGGAAACACC
CATCAATGATATGTTACAGACCTCAAAGATGGAAGGAAGCTATTGGATCTTCTAGAAGGCCTCACAGGAACATCACTGCCA
AAGGAACGTGGTTCCACAAGGGTACATGCCTTAAATAACGTCAACAGAGTGCTGCAGGTTTTACATCAGAACAATGTGGAAT

TAGTGAATATAGGGGGAACCGACATTGTGGATGGAATCACAACTGACTTTGGGGTTACTTTGGAGCATCATTTCGACTGG
CAGGTGAAAAGATGTCATGAAGGATGTCATGTCGGACCTGCAGCAGACGAACAGTGAGAAGATCCTGCTCAGCTGGGTGCGTC
AGACCACCAGGCCCTACAGCCAAGTCAACGTCCTCAACTTCACCACCAGCTGGACAGATGGACTCGCCTTTAATGCTGTCTC
CACCGACATAAACCTGATCTCTTCAGCTGGGATAAAGTTGTCAAAATGTCACCAATTGAGAGACTTGAACATGCCTTCAGCAA
GGCTCAAACCTATTGGGAATTGAAAAGCTGTTAGATCCTGAAGATGTTGCCGTTCCGGCTCCTGACAAGAAATCCATAATTA
TGTATTTAACATCTTTGTTGAGGTGCTACCTCAGCAAGTCACCATAGACTGA

Recombination sites:

attB BxB1

GGCTTGTGACGACGGCGGTCTCCGTCGTCAGGATCAT

attP BxB1

GTGGTTTGTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCA

Inert 5' UTR

TAAGTTTGTACAAAAAAGAG

Polyadenylation signal:

rb glob polyA

TGAATTCACCTCCTCAGGTGCAGGCTGCCTATCAGAAGGTGGTGGCTGGTGTGGCCAATGCCCTGGCTCACAAATACCACTGAG
ATCTTTTTCCCTCTGCCAAAAATTATGGGGACATCATGAAGCCCCTTGAGCATCTGACTTCTGGCTAATAAAGGAAATTTATTT
TCATTGCAATAGTGTGTTGGAATTTTTGTGTCTCTCACTCGGAAGGACATATGGGAGGGCAAATCATTAAAAACATCAGAAT
GAGTATTTGGTTTAGAGTTTGGCAACATATGCCCATATGCTGGCTGCCATGAACAAAGGTTGGCTATAAAGAGGTCATCAGTA
TATGAAACAGCCCCCTGCTGTCCATTCCCTTATCCATAGAAAAGCCTTGACTTGAGGTTAGATTTTTTTTATATTTTGTGTTG
TTATTTTTTCTTTAACATCCCTAAAATTTTCTTACATGTTTTACTAGCCAGATTTTTCTCTCTCTGACTACTCCAGTCAT
AGCTGTCCCTCTTCTTTATGGAGATCCCTCGAC

Annexes for Chapter 3 - Expanding the site-specific recombinases toolbox for mammalian synthetic biology

Supp Figure 3-1 Nucleotide sequences of the attP and attB sites used in this study

Peaches

attB [AC]	GCGGTCTCCATCGGGATCTGCACATCGAGCAGCATGCCGACCAG
attB [AG]	GCGGTCTCCATCGGGATCTGCAGATCGAGCAGCATGCCGACCAG
attB [CC]	GCGGTCTCCATCGGGATCTGCCCATCGAGCAGCATGCCGACCAG
attB [TC]	GCGGTCTCCATCGGGATCTGCTCATCGAGCAGCATGCCGACCAG
attB [TG]	GCGGTCTCCATCGGGATCTGCTGATCGAGCAGCATGCCGACCAG
attB [TT]	GCGGTCTCCATCGGGATCTGCTTATCGAGCAGCATGCCGACCAG

attP [AC]	GCGGTCTCCATCGGGATCTGCACATCGAGCAGCATGCCGACCAG
attP [AG]	GCGGTCTCCATCGGGATCTGCAGATCGAGCAGCATGCCGACCAG
attP [CC]	GCGGTCTCCATCGGGATCTGCCCATCGAGCAGCATGCCGACCAG
attP [TC]	GCGGTCTCCATCGGGATCTGCTCATCGAGCAGCATGCCGACCAG
attP [TG]	GCGGTCTCCATCGGGATCTGCTGATCGAGCAGCATGCCGACCAG
attB [TT]	GCGGTCTCCATCGGGATCTGCTTATCGAGCAGCATGCCGACCAG

BxB1

attB [AC]	GGCTTGTCGACGACGGCGACCTCCGTCGTCAGGATCAT
attB [AG]	GGCTTGTCGACGACGGCGACCTCCGTCGTCAGGATCAT
attB [CC]	GGCTTGTCGACGACGGCGACCTCCGTCGTCAGGATCAT
attB [TC]	GGCTTGTCGACGACGGCGACCTCCGTCGTCAGGATCAT
attB [TG]	GGCTTGTCGACGACGGCGACCTCCGTCGTCAGGATCAT
attB [TT]	GGCTTGTCGACGACGGCGACCTCCGTCGTCAGGATCAT

attP [AC]	GTGGTTTGTCTGGTCAACCACCGGAGCTCAGTGGTGTACGGTACAAACC CA
attP [AG]	GTGGTTTGTCTGGTCAACCACCGGAGCTCAGTGGTGTACGGTACAAACC CA
attP [CC]	GTGGTTTGTCTGGTCAACCACCGGAGCTCAGTGGTGTACGGTACAAACC CA
attP [TC]	GTGGTTTGTCTGGTCAACCACCGGAGCTCAGTGGTGTACGGTACAAACC CA

attP [TG]	GTGGTTTGTCTGGTCAACCACCGCGAGCTCAGTGGTGTACGGTACAAACC CA
attB [TT]	GTGGTTTGTCTGGTCAACCACCGCGAGCTCAGTGGTGTACGGTACAAACC CA

Rebeuca

attB [AC]	CGGTATTCGGCGCGATCCGCGGCGAGAAGAACATCACCTGAACATCG
attB [AG]	CGGTATTCGGCGCGATCCGCGGCGAGAAGAACATCACCTGAACATCG
attB [CC]	CGGTATTCGGCGCGATCCGCGGCGAGAAGAACATCACCTGAACATCG
attB [TC]	CGGTATTCGGCGCGATCCGCGGCGAGAAGAACATCACCTGAACATCG
attB [TG]	CGGTATTCGGCGCGATCCGCGGCGAGAAGAACATCACCTGAACATCG
attB [TT]	CGGTATTCGGCGCGATCCGCGGCGAGAAGAACATCACCTGAACATCG

attP [AC]	AGAGCATCGGAGCCTTTCGGGGGATGTGATGTTTCGAGGAGAAGAACAT
attP [AG]	AGAGCATCGGAGCCTTTCGGGGGATGTGATGTTTCGAGGAGAAGAACAT
attP [CC]	AGAGCATCGGAGCCTTTCGGGGGATGTGATGTTTCGAGGAGAAGAACAT
attP [TC]	AGAGCATCGGAGCCTTTCGGGGGATGTGATGTTTCGAGGAGAAGAACAT
attP [TG]	AGAGCATCGGAGCCTTTCGGGGGATGTGATGTTTCGAGGAGAAGAACAT
attB [TT]	AGAGCATCGGAGCCTTTCGGGGGATGTGATGTTTCGAGGAGAAGAACAT

Veracruz

attB [AC]	CGGTATTCGGCGCGATCCGCGGCGAGAAGAACATCACCTGAACATCG
attB [AG]	CGGTATTCGGCGCGATCCGCGGCGAGAAGAACATCACCTGAACATCG
attB [CC]	CGGTATTCGGCGCGATCCGCGGCGAGAAGAACATCACCTGAACATCG
attB [TC]	CGGTATTCGGCGCGATCCGCGGCGAGAAGAACATCACCTGAACATCG
attB [TG]	CGGTATTCGGCGCGATCCGCGGCGAGAAGAACATCACCTGAACATCG
attB [TT]	CGGTATTCGGCGCGATCCGCGGCGAGAAGAACATCACCTGAACATCG

attP [AC]	GTATTGGGGGAACGCGATATTCGAGGAGTAGAACATCACCTTCACCAAAT TC
attP [AG]	GTATTGGGGGAACGCGATATTCGAGGAGTAGAACATCACCTTCACCAAAT TC
attP [CC]	GTATTGGGGGAACGCGATATTCGAGGAGTAGAACATCACCTTCACCAAAT TC
attP [TC]	GTATTGGGGGAACGCGATATTCGAGGAGTAGAACATCACCTTCACCAAAT TC

attP [TG]	GTATTGGGGGAACGCGATATTCGAGGAGTAGAACATCACCTTCACCAAAT TC
attB [TT]	GTATTGGGGGAACGCGATATTCGAGGAGTAGAACATCACCTTCACCAAAT TC

Theia

attB [AC]	TGGGTGAACGCAAAGATGGGGAACTCGATGCCGAGCTCGTCGCA
attB [AG]	TGGGTGAACGCAAAGATGGGGAACTCGATGCCGAGCTCGTCGCA
attB [CC]	TGGGTGAACGCAAAGATGGGGAACTCGATGCCGAGCTCGTCGCA
attB [TC]	TGGGTGAACGCAAAGATGGGGAACTCGATGCCGAGCTCGTCGCA
attB [TG]	TGGGTGAACGCAAAGATGGGGAACTCGATGCCGAGCTCGTCGCA
attB [TT]	TGGGTGAACGCAAAGATGGGGAACTCGATGCCGAGCTCGTCGCA

attP [AC]	TTGTCAAAGTCTAAAGATGGGGAACTCAATATTCATGCTTTGCGAA
attP [AG]	TTGTCAAAGTCTAAAGATGGGGAACTCAATATTCATGCTTTGCGAA
attP [CC]	TTGTCAAAGTCTAAAGATGGGGAACTCAATATTCATGCTTTGCGAA
attP [TC]	TTGTCAAAGTCTAAAGATGGGGAACTCAATATTCATGCTTTGCGAA
attP [TG]	TTGTCAAAGTCTAAAGATGGGGAACTCAATATTCATGCTTTGCGAA
attB [TT]	TTGTCAAAGTCTAAAGATGGGGAACTCAATATTCATGCTTTGCGAA

Benedict

attB [AC]	TGGGTGAACGCAAAGATGGGGAACTCGATGCCGAGCTCGTCGCA
attB [AG]	TGGGTGAACGCAAAGATGGGGAACTCGATGCCGAGCTCGTCGCA
attB [CC]	TGGGTGAACGCAAAGATGGGGAACTCGATGCCGAGCTCGTCGCA
attB [TC]	TGGGTGAACGCAAAGATGGGGAACTCGATGCCGAGCTCGTCGCA
attB [TG]	TGGGTGAACGCAAAGATGGGGAACTCGATGCCGAGCTCGTCGCA
attB [TT]	TGGGTGAACGCAAAGATGGGGAACTCGATGCCGAGCTCGTCGCA

attP [AC]	TTCGCAAAGCCTCAAAATCGGGAACTCGATATTCATGCTTTGTGAA
attP [AG]	TTCGCAAAGCCTCAAAATCGGGAACTCGATATTCATGCTTTGTGAA
attP [CC]	TTCGCAAAGCCTCAAAATCGGGAACTCGATATTCATGCTTTGTGAA
attP [TC]	TTCGCAAAGCCTCAAAATCGGGAACTCGATATTCATGCTTTGTGAA
attP [TG]	TTCGCAAAGCCTCAAAATCGGGAACTCGATATTCATGCTTTGTGAA

PhiC31

attB [AC]	GTGCGGGTGCCAGGGCGTGCCCACGGGCTCCCCGGGCGCGTACTCC
attB [AG]	GTGCGGGTGCCAGGGCGTGCCCAGGGGCTCCCCGGGCGCGTACTCC
attB [CC]	GTGCGGGTGCCAGGGCGTGCCCCGGGCTCCCCGGGCGCGTACTCC
attB [TC]	GTGCGGGTGCCAGGGCGTGCCCTCGGGCTCCCCGGGCGCGTACTCC
attB [TG]	GTGCGGGTGCCAGGGCGTGCCCTGGGGCTCCCCGGGCGCGTACTCC
attB [TT]	GTGCGGGTGCCAGGGCGTGCCCTTGGGCTCCCCGGGCGCGTACTCC

attP [AC]	GTGCCCCAACTGGGGTAACCTACGAGTTCTCTCAGTTGGGGG
attP [AG]	GTGCCCCAACTGGGGTAACCTAGGAGTTCTCTCAGTTGGGGG
attP [CC]	GTGCCCCAACTGGGGTAACCTCCGAGTTCTCTCAGTTGGGGG
attP [TC]	GTGCCCCAACTGGGGTAACCTTCGAGTTCTCTCAGTTGGGGG
attP [TG]	GTGCCCCAACTGGGGTAACCTTGGAGTTCTCTCAGTTGGGGG
attB [TT]	GTGCCCCAACTGGGGTAACCTTTGAGTTCTCTCAGTTGGGGG

	TCGTGGTCAAGAGTAAAGACTCAGAGCTGAGCGAGGACGATGTGAAGCAGTTCGTCTC CAAACAGGTGGTCTTTTACAAGAGGATCAACAAAGTGTCTTTACAGAATCAATCCCA AAGGCCCCAGCGCAAGATTCTGAGGAAAGATCTGAGAGCTAAACTGGCAAATGGAC TGTGA
CHS optimized	ATGGTCACAGTCGAGGAAGTCAGAAAAGCACAGAGAGCAGAAGGCCCGCTACCGTGA TGGCTATCGGAACCGCGTCCCCCTAATTGCGTGGACCAGGCAACATACCTGATTA CTATTTCCGGATTACTAACTCTGAGCACAAGGCCGAGCTGAAGGAAAAATTTACAGAGA ATGTGCGACAAAAGTCAGATCAAGAAACGGTACATGTATCTGAATGAGGAAGTGCTG AAGGAAAACCCTAATATGTGTGCTTATATGGCACCAAGTCTGGACGCTCGCCAGGATA TCGTGGTTCGTGGAGGTCCCTAAGCTGGGCAAAGAAGCCGCTGTGAAGGCCATCAAGGA GTGGGGACAGCCAAAGAGTAAAATCACTCATCTGGTGTCTGCACCACATCAGGAGTG GACATGCCAGGGGCCGATTACCAGCTGACCAAGCTGCTGGGGCTGCGACCCAGCGTGA AAAGGCTGATGATGTATCAGCAGGGCTGCTTTGCTGGAGGAACCGTGCTGCGACTGGC AAAGGACCTGGCTGAGAACAACAAGGGAGCTCGAGTCCTGGTCGTGTGTTTCAGAAATT ACCGCAGTGACATTCAGGGGCCCCACAGACACTCACCTGGATAGCCTGGTGGGACAGG CCCTGTTTGGAGACGGGGCAGCCGCTATCATTATCGGCAGCGATCCCATCCCTGAGGT CGAAAAGCCCCTGTTTCGAGCTGGTGTCCGCAGCCCAGACCATTCTGCCTGACTCTGAG GGGGCAATCGATGGCCACCTGAGAGAAGTGGGGCTGACATTTTCATCTGCTGAAGGATG TCCCCGGCCTGATTTCAAAGAACGTGGAGAAAAGCCTGACAGAAGCCTTCAAACCTCT GGGCATCTCTGACTGGAATAGTCTGTTTTGGATTGCACACCCAGGAGGACCAGCAATC CTGGATCAGGTGGAGGCCAAGCTGAGCCTGAAGCCAGAAAAACTGCGGGCTACTCGCC ATGTCCTGTCCGAGTACGGCAACATGAGCTCCGCTTGGCTGCTGTTTCATTCTGGACGA GATGAGGAGAAAAGTCCAAAGAAGATGGCCTGAAAACCACCGGCGAGGGAAATCGAATG GGGAGTGCTGTTCCGATTTGGACCAGGACTGACTGTGAGACAGTCGTGCTGCATTCT GTGGCCATCAATTGA

Supp Text 4-3

Nucleotide sequence of TtgR repressor

TtgR codon optimized	ATGGTCAGAAGAACAAGGAGGAGGCTCAGGAAACAAGGGCTCAGATTATCGAAGCA GCAGAAAGGGCATTTTACAAGCGGGGAGTCGCAAGGACCACACTGGCTGACATCGCAG AGCTGGCCGGCGTGACAAGAGGAGCCATCTACTGGCACTTCAACAATAAGGCCGAGCT GGTCCAGGCTCTGCTGGACAGCCTGCATGAAACTCACGATCATCTGGCACGAGCAAGC GAGTCCGAGGACGAGGTGGATCCACTGGGCTGCATGCGCAAACCTGCTGCTGCAGGTGT TCAACGAACTGGTCCTGGACGCCCAGACTAGGAGAATCAATGAAATTTCTGCACCATAA GTGCGAGTTTACCGACGATATGTGTGAAATCAGACAGCAGCGGCAGTCTGCTGTGCTG GATTGTACAAAGGCATTACCCTGGCTCTGGCAAACGCCGTCGCGCAGGACAGCTGC CAGGAGAACTGGACGCAGAGCGCGCCGCTGTGGCTATGTTTGCATATGTCGATGGACT GATCCGAAGGTGGCTGCTGCTGCCCAGCTCTGTGGATCTGCTGGGAGATGTGGAGAAG TGGGTGACACAGGGCTGGATATGCTGCGCCTGAGTCCTGCTCTGCGAAAATGA
-------------------------	---

Annexes for Chapter 5 - Hijacking lentiviruses to create a multi-channel protein-based intercellular communication system

Supp Text 5-1 Nucleotide sequences of VLP proteins

RSV-Gag-P2A-B3	<p>ATGGAAGCCGTCATAAAGGTGATTTTCGTCCGCGTGTA AACCTATTGCGGCAAGACTA GTCCTTCTAAGAAGGAAATAGGGGCCATGTTGTCCCTCTTACAAAAGGAAGGGTTGCT TATGTCTCCCTCAGACTTATATTTCCCGGGGTCCTGGGATCCCATTACCGCGGCGCTAT CCCAGCGGGCTATGATACTTGGGAAATCGGGAGAGTTAAAAACCTGGGGATTGGTTTT GGGGGCATTGAAGGCGGCTCGAGAGGAACAGGTTACATCTGAGCAAGCAAAGTTTTG GTTGGGATTAGGGGAGGGAGGGTATCTCCCCAGGTCCGGAGTGCATCGAGAAACCA GCAACGGAGCGGCAATCGACAAAGGGGAGGAAGTGGGAGAAACAACCTGTGCAGCGA GATGCGAAGATGGCGCCGGAGGAAACGGCCACACCTAAAACCGTTGGCACATCCTGCT ATCATTGCGGAACAGCTATTGGCTGTAATTGCGCCACAGCCTCGGCTCCTCCTCCTCCT TATGTGGGAGTGGTTTTGTATCCTTCCCTGGCGGGGTGGGAGAGCAGCAGGGCCAGG GGGGTGACACACCTCCGGGGGCGGAACAGTCAAGGGCGGAGCCAGGGCATGCGGGTCA GGCTCCTGGGCCCGCCCTGACTGACTGGGCAAGGGTCAGGGAGGAGCTTGCAGTACT GGTCCGCCCCGTGGTGGCCATGCCTGTAGTGATTAAGACAGAGGGACCCGCTTGGACCC CTCTGGAGCCAAAATTGATCACAAGACTGGCTGATACGGTCAGGACCAAGGGCTTACG ATCCCCGATTACTATGGCAGAAGTGGAAAGCGTTATGTCCCTCCCCGCTGCTGCCGCAT GACGTCACGAATCTAATGAGAGTTATTTTAGGGCTGCCCATATGCCTTATGGATGG ACGCTTGGGGAGTCCA ACTCCAGACAGTTATAGCGGCAGCCACTCGCGACCCCCGACA CCCAGCGAACGGTCAAGGGCGGGGGGAACGGACTAATTTGAATCGTTAAAGGGCTTA GCTGATGGGATGGTGGGCAACCCACAGGGTCAGGCCGATTATTAAGACCGGGGAAT TGTTGCTATTACGGCGTCGGCTCTCCAGGCGTTTAGAGAGGTTGCCCGGCTGGCGGA ACCTGCAGGTCCATGGGCGGACATCATGCAGGGACCATCTGAGTCCTTTGTTGATTTT GCCAATCGGCTTATAAAGGCGGTTGAGGGGTCAGATCTCCCGCCTTCCGCGCGGGCTC CGGTGATCATTGACTGCTTTAGGCAGAAGTCACAGCCAGATATTCAGCAGCTTATACG GACAGCACCTCCACGCTGACCACCCAGGAGAGATAATTAATATGTGCTAGACAGG CAAAGACTGCCCCTCTTACGGATCAAGGCATAGCCGCGGCCATGTCGTCTGCTATCC AGCCCTTAATTATGGCAGTAGTCAATAGAGAGAGGGATGGACAAACTGGGTCGGGTG GTCGTGCCCGAGGGCTCTGCTACACTTGTGGATCCCCGGGACATTATCAGGCGCAGTG CCGAAAAAACGGAAGTCAGGAAACAGCCGTGAGCGATGTCAGTTGTGTAACGGGAT GGGACACAACGCTAAACAGTGTAGGAAGCGGGATGGCAACCAGGGCCAACGCCCAGGA AAAGGTCTATCTTCGGGACCGTGGCCTGGTCCC GGAATCATGAGCTCGTATATGGATC TTGTTGATGATGAACCAGCGACTTTGTACCATAAGTTCGTGGAGTGCTTGAAAGCGGG CGAGAACTTCTGCGGAGACAAGCTGAGTGAATTATTACCATGGCGATCCTTAAGGCA ATCAAGGCGCTGACCGAAGTTAAAAAGACAACCTTTAACAAATATAAGACAACAATC AAGCAGGGCCTCCAGTATGACGTGGGTTTCGTCGACTATCTCGTTTGTGTATCACTTGA AGGACTGTGATGAGCTGTCCAGGGCTTGAGCGATGCCTTCGAGCCCTACAAATTCAA AATTAAGTCAATAAAGAGGCAACCTCGTTTAAGACTCTCTTTCGTGGCCCCCTCGTTT GGCAGCCAGAAGAACTGGCGGAAGAAAGAGGTGGACCGCGAGGTGGATAACTTGTTT CATAGCACCGAGACAGACGAATCGATTTTCAAATTCATCTTGAACACGTTGGATAGTA TTGAGACACAAACGAACACGGATCGCCAAAAGACGGTGCTGACTTTCATCCTGTTGAT</p>
----------------	--

	<p>GACATTTTTCAACTGCTGTAGGAACAATGACCTGATGAACGTTGATCCCTCCACATTT AAGATTGTGAAAAACAAATTCGTGCGGATACCTGCTGCAGGCTGAGGTCAAACAGACT AAGACACGCAAGTCGAGGAACATTTTCTTCTTTCCCATCCGCGAGAATCGATTGATC TGTTCTGGCCTTGCACGATTTCTTCCGCACATGCCAGCCTACCCCAAAGTCGCGTCTT TCGGATCAAGTATCGGAGCAGAAGTGGCAGCTTTTCCGAGATTCCATGGTCATTGATT ACAACCGTTTCTTTAGGAAGTTTCCAGCTTCGCCTATTTTCGCAATAAAGCACGGCCC CAAGTCCCATCTGGGCCGGCATCTGATGAACAGCTTTCTGCACAAGAATGAACTGGAT TCCTGGGCCAACTCCCTGGGCAACTGGAGCTCCTCCCAGAATCAACGCGAGTCCGGTG CGCGACTGGGCTACACCCACGGTGGTCGAGATTTGCCACAGCCACTGTTCCGATTCTT GGCTGGCTACTGCGTTTCGAAATGAAGAGGGCCACATCGTGGGCTTGGGTCTGGAGAAG GACATTAATGATCTGTTTACCGGTATTATGGACCCACTTAATGAGAAGGAGGATACA GAGATTTGTGAAAGCTACGGCGAGTGGGCCAAAATTGTGAGCAAGGATGTTCTGATA TTTCTGAAGCGATATCATTCGAAGAACGCATGCCGGCGATATCAGAACTCCACATTGT ACGCACGTACGTTCTTAAGACAGAGTCCGTCACCTTGAGCGGCTCCAAGGGAAGCGA AGAGCCGAGCAGTCCCGTCAGGATTCCAATACTTAGTATGGGAAAAGCCTCCCCAAGC GAGGGCCGAAAGTTGCGTGCCAGCGAACATGCCAACGATGATAATGAGATTGAGAAG ATAGATTCCGATTCTGTCGAAAGCGAAGAGATACCTATCGAAATGTCCGACTCCGAGG ACGAAACAACGGCAAGTAACATTAGCGGAATTTACCTGGACATGTCGAAAGCCAACTC CAACGTGGTCTATTCCCCTCCGAGTCAGACTGGCCGTGCCGCCGGTGGCCGCCGTA AAC GTGGCGTGGGTGGTGCCTGCACCGTTGAGTCGAAGCGGCGGCGTCTGGCACCAAT TAACCGGTGA</p>
<p>RSV-Gag- P2A-Cre</p>	<p>ATGGAAGCCGTCATAAAGGTGATTTTCGTCCGCGTGTA AACCTATTGCGGCAAGACTA GTCCTTCTAAGAAGGAAATAGGGGCCATGTTGTCCCTCTTACAAAAGGAAGGGTTGCT TATGTCTCCCTCAGACTTATATTTCCCCGGGGTCTGGGATCCCATTACCGCGGCGCTAT CCCAGCGGGCTATGATACTTGGGAAATCGGGAGAGTTAAAAACCTGGGGATTGGTTTT GGGGGCATTGAAGGCGGCTCGAGAGGAACAGGTTACATCTGAGCAAGCAAAGTTTTG GTTGGGATTAGGGGGAGGGAGGGTATCTCCCCAGGTCCGGAGTGCATCGAGAAACCA GCAACGGAGCGGCAATCGACAAAGGGGAGGAAGTGGGAGAAACA ACTGTGCAGCGA GATGCGAAGATGGCGCCGGAGGAAACGGCCACACCTAAAACCGTTGGCACATCCTGCT ATCATTGCGGAACAGCTATTGGCTGTAATTGCGCCACAGCCTCGGCTCCTCCTCCTCCT TATGTGGGGAGTGGTTTTGTATCCTTCCCTGGCGGGGGTGGGAGAGCAGCAGGGCCAGG GGGGTGACACACCTCCGGGGGCGGAACAGTCAAGGGCGGAGCCAGGGCATGCGGGTCA GGCTCCTGGGCCGGCCCTGACTGACTGGGCAAGGGTCAGGGAGGAGCTTGGGAGTACT GGTCCGCCCCGTGGTGGCCATGCCTGTAGTGATTAAGACAGAGGGACCCGCTTGGACCC CTCTGGAGCCAAAATTGATCACAAAGACTGGCTGATACGGTCAGGACCAAGGGCTTACG ATCCCCGATTACTATGGCAGAAGTGAAGCGCTTATGTCCTCCCCGCTGCTGCCGCAT GACGTACGAATCTAATGAGAGTTATTTTAGGGCTGCCCCATATGCCTTATGGATGG ACGCTTGGGGAGTCCA ACTCCAGACAGTTATAGCGGCAGCCACTCGCGACCCCCGACA CCCAGCGAACGGTCAAGGGCGGGGGGAACGGACTAATTTGAATCGCTTAAAGGGCTTA GCTGATGGGATGGTGGGCAACCCACAGGGTCAGGCCGATTATTAAGACCGGGGGAAT TGTTTGCTATTACGGCGTCGGCTCTCCAGCGTTTTAGAGAGGTTGCCCGCTGGCGGA ACCTGCAGGTCCATGGGCGGACATCATGCAGGGACCATCTGAGTCTTTGTTGATTTT GCCAATCGGCTTATAAAGGCGGTTGAGGGGTGAGATCTCCCGCTTCCGCGCGGGCTC CGGTGATCATTACTGCTTTAGGCAGAAGTACAGCCAGATATTCAGCAGCTTATACG GACAGCACCTCCACGCTGACCACCCAGGAGAGATAATTAATATGTGCTAGACAGG CAAAGACTGCCCTCTTACGGATCAAGGCATAGCCGCGGCCATGTCGCTGCTATCC AGCCCTTAATTATGGCAGTAGTCAATAGAGAGAGGGATGGACAAACTGGGTGGGTG GTCGTGCCCGAGGGCTCTGCTACACTTGTGGATCCCCGGGACATTATCAGGCGCAGTG CCCGAAAAACGGAAGTCAGGAAACAGCCGTGAGCGATGTCAGTTGTGTAACGGGAT GGGACACAACGCTAAACAGTGTAGGAAGCGGGATGGCAACCAGGGCCAACGCCCAGGA AAAGGTCTATCTTCGGGACCGTGGCCTGGTCCCAGGAATCATGTCCAACCTGCTGACTG TGCACCAAAACCTGCCTGCCCTCCCTGTGGATGCCACCTCTGATGAAGTCAGGAAGAA CCTGATGGACATGTTTCAAGGACAGGCAGGCCTTCTCTGAACACACCTGGAAGATGCTC CTGTCTGTGTGCAGATCCTGGGCTGCCTGGTGAAGCTGAACAACAGGAAATGGTTCC</p>

	<p>CTGCTGAACCTGAGGATGTGAGGGACTACCTCCTGTACCTGCAAGCCAGAGGCCTGGC TGTC AAGACCATCCAACAGCACCTGGGCCAGCTCAACATGCTGCACAGGAGATCTGGC CTGCCTCGCCCTTCTGACTCCAATGCTGTGTCCCTGGTGATGAGGAGAATCAGAAAGG AGAATGTGGATGCTGGGGAGAGAGCCAAGCAGGCCCTGGCCTTTGAACGCACTGACTT TGACCAAGTCAGATCCCTGATGGAGA ACTCTGACAGATGCCAGGACATCAGGAACCTG GCCTTCTGGGCATTGCCTACAACACCCTGCTGCGCATTGCCGAAATTGCCAGAATCA GAGTGAAGGACATCTCCCGCACCGATGGTGGGAGAATGCTGATCCACATTGGCAGGAC CAAGACCCTGGTGTCCACAGCTGGTGTGGAGAAGGCCCTGTCCCTGGGGGTTACCAAG CTGGTGGAGAGATGGATCTCTGTGTCTGGTGTGGCTGATGACCCCAACA ACTACCTGT TCTGCCGGGTCAGAAAGAATGGTGTGGCTGCCCTTCTGCCACCTCCCAACTGTCCACC CGGGCCCTGGAAGGGATCTTTGAGGCCACCACCCTGATCTATGGTGCCAAGGATG ACTCTGGGCAGAGATACCTGGCCTGGTCTGGCCACTCTGCCAGAGTGGGTGCTGCCAG GGACATGGCCAGGGCTGGTGTGTCCATCCCTGAAATCATGCAGGCTGGTGGCTGGACC AATGTGAACATTGTGATGAACTACATCAGAAACCTGGACTCTGAGACTGGGGCCATGG TGAGGCTGCTCGAGGATGGGGACTGA</p>
VSV-G	<p>ATGAAGTGCCTTTTGTACTTAGCCTTTTATTTCATTGGGGTGAATTGCAAGTTCACCA TAGTTTTTCCACACAACCAAAAAGGAAACTGGAAAAATGTTCCCTTCTAATTACCATTA TTGCCCGTCAAGCTCAGATTTAAATTGGCATAATGACTTAATAGGCACAGCCTTACAA GTCAAAAATGCCAAGAGTCACAAGGCTATTCAAGCAGACGGTTGGATGTGTCATGCTT CCAAATGGGTCACTACTTGTGATTTCCGCTGGTATGGACCGAAGTATATAACACATTC CATCCGATCCTTCACTCCATCTGTAGAACAATGCAAGGAAAGCATTGAACAAACGAAA CAAGGA ACTTGGCTGAATCCAGGCTTCCCTCCTCAAAGTTGTGGATATGCAACTGTGA CGGATGCCGAAGCAGTGATTGTCCAGGTGACTCCTCACCATGTGCTGGTTGATGAATA CACAGGAGAAATGGGTTGATTCACAGTTCATCAACGGAAAATGCAGCAATTACATATG CCCCACTGTCCATAACTCTACAACCTGGCATTCTGACTATAAGGTCAAAGGGCTATGT GATTCTAACCTCATTTCCATGGACATCACCTTCTTCTCAGAGGACGGAGAGCTATCAT CCCTGGGAAAGGAGGGCACAGGGTTCAGAAGTAACTACTTTGCTTATGAAACTGGAG GCAAGGCCTGCAAAATGCAATACTGCAAGCATTGGGGAGTCAGACTCCCATCAGGTGT CTGGTTCGAGATGGCTGATAAGGATCTCTTTGCTGCAGCCAGATTCCCTGAATGCCCA GAAGGGTCAAGTATCTCTGCTCCATCTCAGACCTCAGTGGATGTAAGTCTAATTCAGG ACGTTGAGAGGATCTTGGATTATTCCCTCTGCCAAGAAACCTGGAGCAAAATCAGAGC GGGACTTCCAATCTCTCCAGTGGATCTCAGCTATCTTGCTCCTAAAAACCAGGAACC GGTCTGCTTTCACCATAATCAATGGTACCCTAAAATACTTTGAAACCAGATACATCA GAGTCGATATTGCTGCTCCAATCCTCTCAAGAATGGTCGGAATGATCAGTGGA ACTAC CACAGAAAGGAACTGTGGGATGACTGGGCACCATATGAAGATGTGGAAATTTGGACC CAATGGAGTTCTGAGGACCAGTTCAGGATATAAGTTTCTTTATACATGATTGGACAT GGTATGTTGGACTCCGATCTTCATCTTAGCTCAAAGGCTCAGGTGTTGCAACATCCTC ACATTCAAGACGCTGCTTCGCAACTTCTGATGATGAGAGTTTATTTTTTGGTGATAC TGGGCTATCCAAAATCCAATCGAGCTTGTAGAAGGTTGGTTCAGTAGTTGGAAAAG CTCTATTGCCTCTTTTTTCTTTATCATAGGGTTAATCATTGGACTATTCTTGGTTCTC CGAGTTGGTATCCATCTTTGCATTA AATTAAGCACACCAAGAAAAGACAGATTTAT ACAGACATAGAGATGAACCGACTTGGAAAGTGA</p>

Chapter 9 Manuscripts, reports and presentations

Rapid, modular and reliable construction of complex mammalian gene circuits

Patrick Guye¹, Yinqing Li², Liliana Wroblewska¹, Xavier Duportet^{1,3} and Ron Weiss^{1,2,*}

¹Department of Biological Engineering, Massachusetts Institute of Technology, 40 Ames Street, Cambridge, MA 02142, USA, ²Department of Electrical Engineering and Computer Science, Massachusetts Institute of Technology, 40 Ames Street, Cambridge, MA 02142, USA and ³INRIA Paris-Rocquencourt, Le Chesnay, 78153, France

Received April 28, 2013; Revised June 10, 2013; Accepted June 18, 2013

ABSTRACT

We developed a framework for quick and reliable construction of complex gene circuits for genetically engineering mammalian cells. Our hierarchical framework is based on a novel nucleotide addressing system for defining the position of each part in an overall circuit. With this framework, we demonstrate construction of synthetic gene circuits of up to 64 kb in size comprising 11 transcription units and 33 basic parts. We show robust gene expression control of multiple transcription units by small molecule inducers in human cells with transient transfection and stable chromosomal integration of these circuits. This framework enables development of complex gene circuits for engineering mammalian cells with unprecedented speed, reliability and scalability and should have broad applicability in a variety of areas including mammalian cell fermentation, cell fate reprogramming and cell-based assays.

INTRODUCTION

The ability to introduce into mammalian cells genetic circuits that contain multiple transcription units (TUs) is of great interest for a variety of applications including biotechnology (1), gene therapy (2), systems/synthetic biology (3) and reprogramming cell fate and functions (4), as well as basic biological research (5). Encoding such multi-TU gene circuits on single vectors offers several advantages over using separate vectors (6,7), for example, to improve correlation in gene expression between the different circuit elements and for an integration of the entire circuit into a single genomic locus. However, the construction of such large single-vector circuits is challenging because of long and/or repetitive sequences and the need for genetic elements that impart

robust expression in mammalian cells. Existing DNA assembly methods are often not well suited for manipulating large collections of mammalian sequences. For example, methods that rely on the use of Type II restriction enzymes (8) can be problematic because these restriction sites occur frequently in mammalian promoters and genes. Other methods require multiple rounds of cloning (6,9), cloning in yeast (10) or polymerase chain reaction (PCR) (7). With PCR, the precision of even high-fidelity polymerases is insufficient for reliable and error-free large-scale amplification (Supplementary Table S1). Furthermore, multi-TU gene circuits lacking insulating elements suffer from transcriptional interference and are significantly hampered in their function (11). To address these issues, we developed a new framework for quick and reliable assembly of functional complex mammalian gene circuits. Here, we describe in detail the components, steps and mechanisms underlying the framework. We demonstrate efficient and robust construction of circuits with various sizes and number of assembled parts, and show that assembly works well despite repetitive sequences present in some of the parts. The resulting gene circuits were functionally assessed in transfection as well as stable genomic integration and behaved according to their predicted phenotypes. The framework described here can also prove to be valuable for building large-scale mammalian genetic module libraries, and is well suited for generation of stable cell lines with multielement circuits.

MATERIALS AND METHODS

Strains

Escherichia coli 10G (Lucigen Corp., Middletown, WI, USA) were used for most cloning steps except for the pJazz-based and the bacterial artificial chromosome (BAC)-based carrier vectors. pJazz/BigEasy v2.0 Linear Cloning System and pSMART-BAC (Lucigen Corp.) were purchased and used according to their manual in

*To whom correspondence should be addressed. Tel: +1 617 253 8966; Fax: +1 617 452 2631; Email: rweiss@mit.edu

The authors wish it to be known that, in their opinion, the first two authors should be regarded as joint First Authors.

© The Author(s) 2013. Published by Oxford University Press.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/3.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

their respective strains (BigEasy TSA and BAC-Optimized Replicator v2.0). Antibiotic was used with following concentrations: 100 µg/ml Ampicillin, 50 µg/ml Kanamycin, 25 µg/ml Chloramphenicol. Gel extraction was done with the Qiagen QIAquick Gel Extraction Kit. PCR purification was done using the Qiagen QIAquick PCR Purification Kit. Miniprepping of DNA was done using the Qiagen Qiaprep Spin Miniprep Kit. Some minipreps were automated on a Qiagen Qiacube using the same miniprep kit.

Library of genes and promoters

The Gateway BP reaction was performed according to the manufacturer's manual (Life technologies, Carlsbad, CA, USA). Briefly, to create the promoter library, the promoter sequences of interest were amplified, digested and inserted into pENTR_L4_R1 that was cut with compatible restriction enzymes. To create the gene library, the gene of interest was amplified with attB1-site in the forward primer and attB2-site in the reverse primer. Ten femtomoles of the PCR product was mixed with 5 fmol of pDONR221P2 and incubated with 1 µl of BP clonase II (11789-020, Life technologies) for 1 h. One microliter of the reaction was transformed into ccdB sensitive competent *E. coli* cells. Alternatively, the gene sequences of interest were amplified, digested and inserted into pENTR_L1_L2 cut with compatible restriction enzymes.

Gateway LR reaction

Gateway LR reactions were performed according to the manufacturer's manual (Life technologies, Carlsbad). Briefly, 5 fmol of each of pENTR_L4_R1, pENTR_L1_L2 and pZDonor_Seq(n)-GTW-Seq(n+1) that contains Gateway cassette of pDEST_R4_R2 were mixed and incubated with 1 µl of LR clonase II mix (11971-020, Life technologies) for 16 h. One microliter of the reaction was transformed into ccdB sensitive competent *E. coli* cells.

Digestion of vectors containing the basic assembly units

Seventy femtomoles of each vector containing a basic assembly unit were pooled and digested in a total volume of 20 µl for 6 h using 10 U of the restriction enzyme I-SceI (R0694, NEB Biolabs, MA, USA). Subsequently, this digest was purified using the QIAGEN QIAquick PCR purification kit and eluted in 30 µl of warm Tris-EDTA (TE) buffer. Alternatively, the I-SceI digest was heat-inactivated (65°C, 20 min) and used without further purification.

Digestion of the adaptor vector

Two hundred eighty femtomoles of the adaptor vector required for proper circuit closure was digested in a total volume of 20 µl with 10 U each of restriction enzymes XbaI and XhoI (R0146, R0145, NEB Biolabs) for 3 h, purified using the QIAGEN QIAquick PCR purification kit and eluted in 30 µl of warm TE buffer. Alternatively, the digest was heat-inactivated (65°C, 20 min) and used without further purification.

Digestion of the carrier vector

One hundred forty femtomoles of the carrier vector was digested in a total volume of 20 µl with 4 U of the restriction enzyme FseI or PacI (R0588, R0547 NEB Biolabs) for 3 h, purified using the QIAGEN QIAquick PCR purification kit and eluted in 30 µl of warm TE buffer. Alternatively, the digest was heat-inactivated (65°C, 20 min) and used without further purification.

One-step assembly using exonuclease, ligase and polymerase (Gibson Reaction)

The reaction buffer was made according to Gibson's original recipe (12). Briefly, isothermal reaction buffer (IRB) was prepared on ice and stored at -20°C: 25% PEG-8000, 500 mM Tris-HCl, pH 7.5, 50 mM MgCl₂, 50 mM dithiothreitol (DTT), 1 mM each of dATP, dCTP, dGTP and dTTP and 5 mM NAD. Assembly Master Mix was prepared on ice and stored at -20°C: 320 µl IRB buffer, 0.64 µl of 10 U/µl T5 exonuclease (T5E4111K, Epicentre Biotechnologies, WI, USA), 20 µl of 2 U/µl Phusion polymerase (F-530, NEB Biolabs), 160 µl of 40 U/µl Taq ligase (M0208, NEB Biolabs), deionized water to a 1.2 ml total volume. Seven femtomoles of each part (digested adaptor vector, digested carrier vector and digested pool of assembly units) were combined in a 200 µl PCR reaction tube on ice and filled up to a 5 µl of total volume with deionized water. The mix was then added to 15 µl of the Assembly Master Mix, and the reaction was incubated at 50°C for 1 h. This reaction (2-5 µl) was then transformed into competent *E. coli* cells.

Hierarchical assembly

The assembly units and the adaptor were assembled into the hierarchical pJazz carrier vector. One hundred forty femtomoles of assembled vector was digested in total volume of 20 µl for 6 h using 10 U of the restriction enzyme I-SceI. Seven femtomoles of this digest was combined with 7 fmol of each additional part (digested adaptor vector, digested carrier vector and digested pool of assembly units). Kanamycin resistance adaptor and Tetracycline resistance adaptors were used in an alternating manner. A one-step assembly protocol was applied on this mixture.

Cell culture

HEK293FT (Invitrogen, Carlsbad, CA, USA) cells were cultured in supplemented Dulbecco's modified Eagle's medium according to their manual. Chemical DNA transfection was performed using Qiagen SuperFect Transfection Reagent (QIAGEN, Hilden, Germany) or Metafectene Pro (Biontix, Martinsried, Germany). In brief, 800 000 cells were seeded into a 10-cm² well and immediately transfected with 2 µg of DNA. The medium was replaced 6 h after transfection. To induce the Tet-On system, Doxycycline (Clontech, Mountain View, CA, USA) was added to culture media at 1 µg/ml. To induce the Rheo system, Genostat ligand (EMD Millipore, Burlington, MA, USA) was added to the cell culture at

a final concentration of 5 nM. Targeted integration into the AAVS1 locus: The carrier vector was stably integrated (13) into the chromosome of HEK293 cells and stable clones were selected with 1 µg/ml of Puromycin (Xavier Dupontet *et al.*, submitted for publication).

Microscope imaging

Images were taken using a Leica TCS SP5 II 405UV confocal microscope (Leica Microsystems, Bannockburn, IL, USA). Images were acquired using a sequential scan. First scan: excitation/laser lines: 488 nm, emission: 495–539 nm; second scan: excitation/laser lines: 458 and 543 nm, emission: 462–487 and 547–800 nm, respectively.

Flow cytometry measurement

Flow cytometry measurement was carried out on BD LSR II in Koch Institute Flow Cytometry Core at MIT. Data were collected in BD FACSDiva software and analyzed in Flowjo (Tree Star, Inc. Ashland, OR, USA)

Algorithms

The algorithm and parameters for designing the unique oligonucleotide sequences are detailed in Supplementary Method.

RESULTS

Genetic circuit design and construction

Our assembly method integrates multisite Gateway recombination, Gibson assembly (12) and a nucleotide-addressing system for defining the position of every part in the final overall vector (Figure 1a). First, the user chooses promoter/gene pairs from a sequence-verified library of parts and then determines the circuit position of each of these TUs by Gateway recombination with an appropriate customized Gateway destination vectors. The resulting vectors, called position vectors, contain nucleotide sequences that specify the position of each TU in the final circuit vector. Position vectors are verified by restriction mapping (>90% usually correct) (15), and then digested and assembled together with a carrier vector and an adaptor vector using a Gibson reaction to form the final vector. Our customized Gateway destination vectors contain (from 5' to 3') an I-SceI restriction site, a unique nucleotide sequence (UNS), a tandem repeat of the core cHS4 chromatin insulator (16), a Gateway recombination cassette, a polyadenylation sequence, another UNS and another I-SceI restriction site (Figure 1a, Supplementary Figure S1a). The I-SceI recognition sites and UNSs form the core of the nucleotide-addressing system. Digesting the position vector with I-SceI releases the TU flanked by the two UNSs. The TU should not contain the 18-bp I-SceI site, but the likelihood of this sequence being present in a TU is small—neither the mouse genome nor the human genomes contain this recognition site. UNSs comprise a series of 40-bp nucleotide sequences that are designed using a computational algorithm to maximize the probability of annealing to the complementary UNS during the Gibson reaction and

to minimize hairpin formation when exposed as single-stranded DNA (Supplementary Figure S1b, algorithm described in Supplementary Methods). Chromatin insulator and polyadenylation sequences are included in each position vector for robust mammalian gene expression. Once assembled, chromatin insulators from adjacent assembled transcriptional units form insulation pairs that are used to dampen crosstalk by transcriptional regulators (11) as well as the spreading of genomic silencing (17). The carrier vector contains sequences necessary for propagation in *E. coli* as well as UNS 1 and X, where UNS X is used to link the last TU in the circuit to the carrier vector. Additional genetic elements, for instance episomal sequences or genomic recombination sites, can be added to the carrier vector to obtain other desired functionality. To link the last position vector to the carrier vector at UNS X, an adaptor vector is chosen from an adaptor library. The adaptor also provides a second selection marker (e.g. Kanamycin or Tetracycline) to select against empty vector backbones during Gibson assembly. For example, a 5-TU gene circuit can be assembled from position vectors with UNS pairs 1–2, 2–3, 3–4, 4–5 and 5–6 and an adaptor vector containing UNS 6–X (Figure 1a). A list of components available as part of the platform is available in Supplementary Table S2. To facilitate gel electrophoresis analysis of the large gene circuits, we developed an algorithm to create high-resolution restriction maps (Supplementary Figure S2, restriction map algorithm described in Supplementary Method S2).

Assembly method efficiency

We demonstrate the efficiency of our method by assembling >30 basic parts into single vectors (Figure 1c and d and Supplementary Figure S2). The entire construction process from basic parts (promoters, genes) to a transfection- or integration-ready circuit vector requires <5 days with a workload of a few hours a day (Figure 1b). Starting from preexisting TUs, construction takes only 2 days. Figure 1c shows the percentage of bacterial colonies that contained the correct circuits for assemblies of up to 33 basic parts and a final vector size of 64 kb. Our data suggest that for assemblies with <21 parts or 7 TUs, it is usually sufficient to screen only two colonies (<10% probability of both colonies not containing the correct circuit, Figure 1b).

Reliable, robust expression and exogenous control

We assayed the behavior in mammalian cells of large gene circuits constructed with our method. A 39-kb vector containing a 7-TU circuit that was assembled into a BAC carrier vector (Figure 2a and restriction digest/analysis in Supplementary Figure S2), exhibited robust gene expression on induction with Doxycycline and/or Rheo ligand when transfected into HEK293FT cells (Figure 2a). We also compared gene expression from a single transfected plasmid to gene expression from three co-transfected plasmids that contain the same circuit elements. The circuit comprises constitutive expression of Enhanced Blue Fluorescent Protein (EBFP),

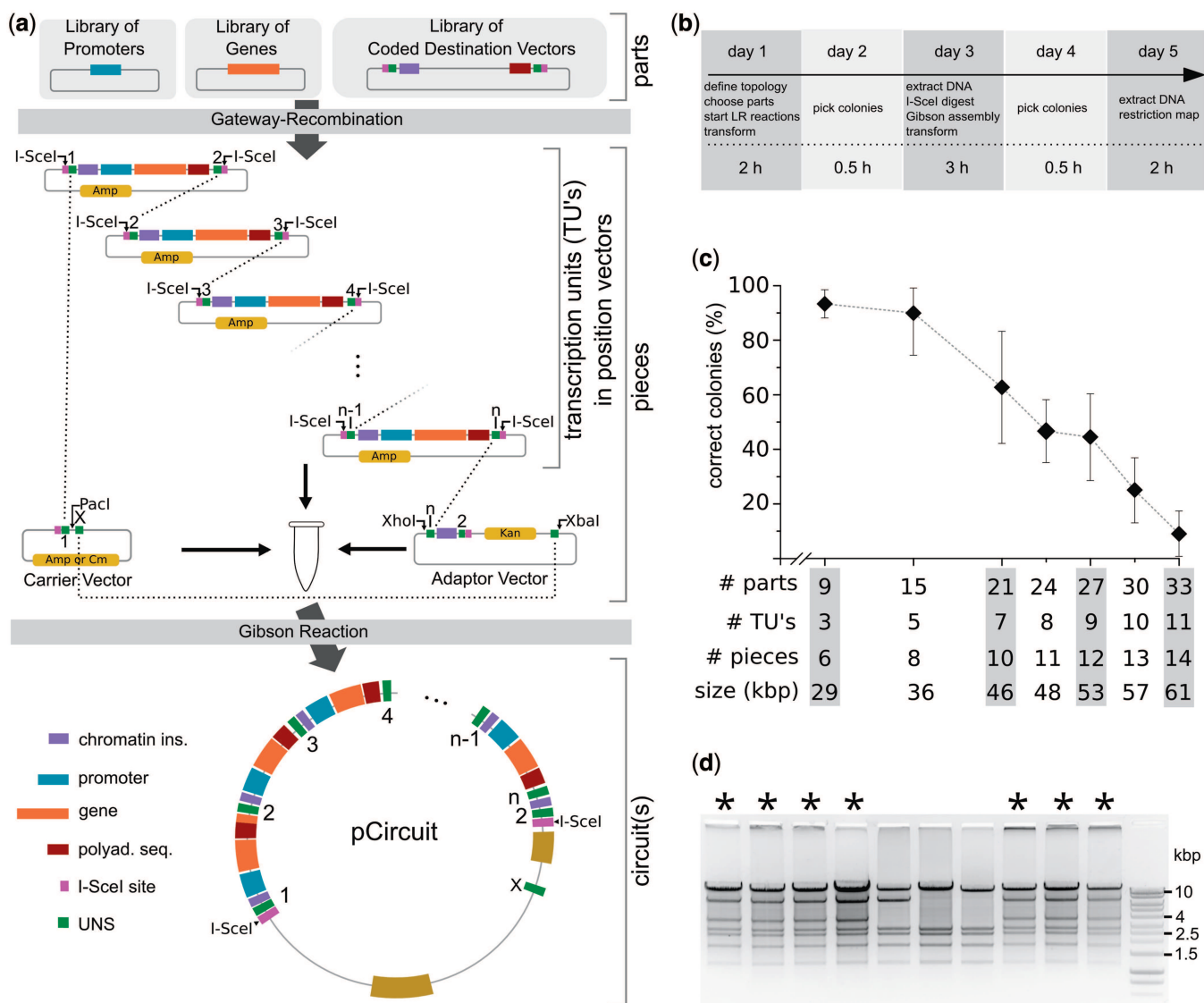


Figure 1. Overview of the DNA assembly process and its efficiency. (a) Schematics of the assembly process. For every TU, the chosen promoter, gene and destination vector are recombined using the Gateway LR reaction. The UNS in the destination vector determines the TU's position within the gene circuit. The resulting position vectors are pooled, digested with I-SceI and combined with equimolar amounts of predigested carrier vector and adaptor vector. The Gibson reaction mix exposes the UNS, permitting annealing, extension and ligation to form a single large vector (pCircuit). pCircuit itself can be reused for a further round of assemblies by digesting it with I-SceI, freeing the 5-TU circuit flanked by the UNS 1 and 2. (b) Timeline and steps for circuit assembly starting from libraries of parts. If all TUs are already available, the assembly process starts from day 3 and takes 2 days to finish. The approximate hours of benchwork needed each day is indicated in the bottom row (h = hours). (c) TUs were assembled into a linear carrier vector (customized pJazz) (14). The percentage reflects the number of colonies containing the correct vector vs. the number of analyzed colonies. Size in kilobases is the average size (multiple different circuits tested per # of parts) (d) Accuracy of the assembly process for a 7-TU circuit. Asterisk: correct restriction digest.

Enhanced Yellow Fluorescent Protein (EYFP) and a reverse tetracycline transactivator (rtTA3) and Doxycycline-inducible mKate (Figure 2b). EYFP serves as a transfection marker. The results show that having the entire circuit in a single plasmid significantly reduces EBFP and mKate variance over a wide range of transfection levels, based on flow cytometry analysis (Supplementary Figure S3). To determine if circuits assembled with this framework perform well with stable chromosomal integration, we integrated the above 3-TU circuit into the AAVS1 locus of HEK293FT cells and quantified the resulting fluorescence. After 17 days,

>85% of polyclonal cells co-expressed EBFP and EYFP with similar levels of expression and 87% of these EBFP-EYFP double-positive cells expressed mKate in an inducible manner (Figure 2c and d).

Hierarchical circuit assembly

Our method supports hierarchical assembly of gene circuits (Supplementary Figure S4). Hierarchical construction and reuse of genetic modular circuits should facilitate design and implementation of high-order gene circuits and larger systems (18). With this approach to

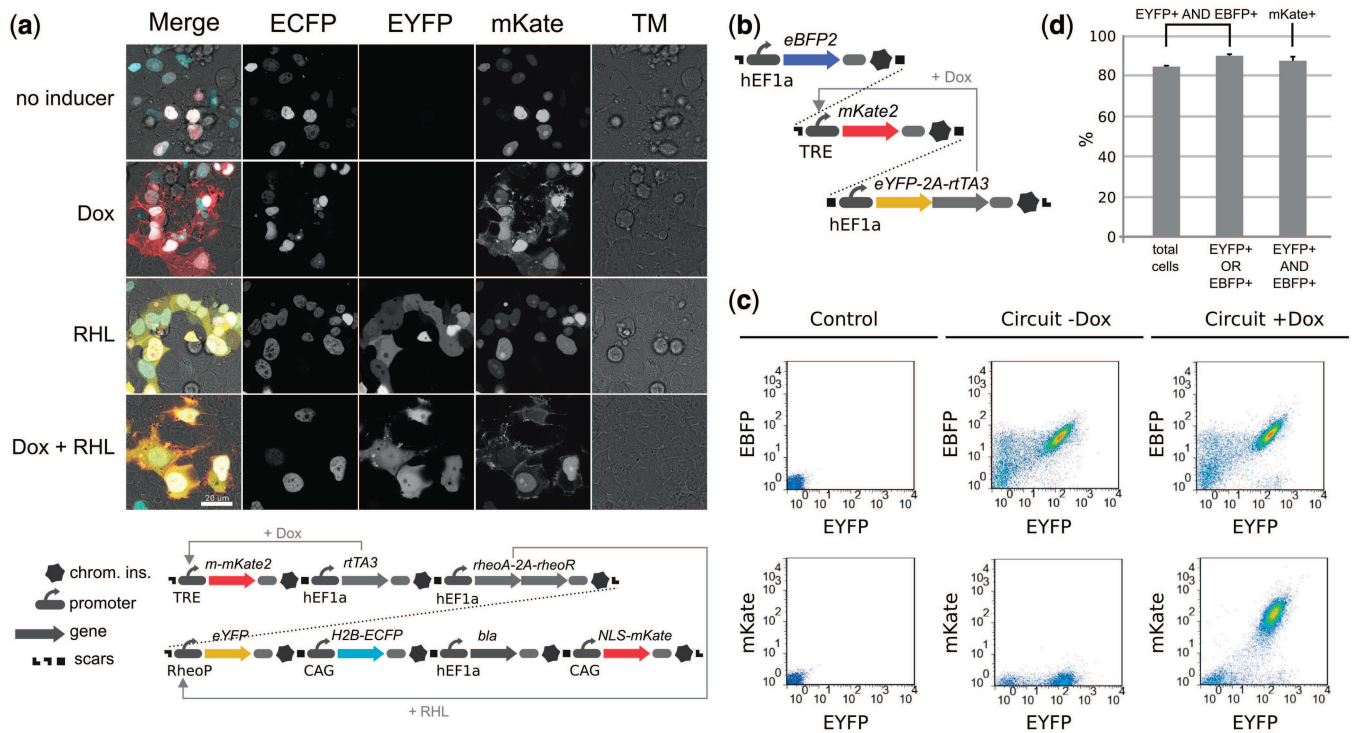


Figure 2. Gene circuit function in mammalian cells. **(a)** Microscope images from transfection of a 7-TU, 39-kb gene circuit. The circuit constitutively expresses nuclear-localized ECFP, nuclear-localized mKate, rtTA3, RheoAct/RheoRec and Blasticidin. On addition of Doxycycline, expression of membrane-localized mKate2 is induced. EYFP is expressed on addition of Genostat ligand. Chrom. ins., chromatin insulator. Scale bar: 20 μ m **(b)** A 3-TU, 17-kb gene circuit that was stably integrated into HEK293FT cells. **(c)** Flow cytometry dot-plots for cells not containing the circuit (Control), and cells containing the chromosomally integrated circuit with or without Doxycycline (\pm Dox). **(d)** Assessing the functionality of the gene circuit described in (b). Percentage of cells expressing EYFP and EBFP in (first column) total population, (second column) EYFP or EBFP expressing population and (third column) percentage of cells expressing mKate in EYFP and EBFP double-positive cells on addition of Doxycycline.

the creation of sophisticated systems, one can first assemble and verify simpler modules (e.g. bistable switches, cascades, biosynthesis pathways, etc.) and also obtain such modules from other sources, before proceeding to the construction of the complete systems. Recently, a plug-and-play method (19) based on restriction enzymes was developed that supports reuse and repurposing of existing genetic circuits for the construction of new bacterial synthetic gene networks. In our framework, each assembled circuit vector contains an *I-SceI* site upstream of the vector's 5' UNS 1 and an *I-SceI* site downstream of UNS 2 at the 3'-end (Figure 1a). Digestion with *I-SceI* creates a position 1–2 vector piece that can be assembled with other position vectors and an appropriate adaptor vector. Two different adaptor vectors, containing either Kanamycin or Tetracycline resistance, are used in an alternating fashion to select against the parental modules. To demonstrate recursive assembly of gene circuits, we first designed and built a 7-TU, 45-kb module. The module contains three inducible promoters and four constitutive promoters expressing inducible regulators, selection markers and fluorescent reporters. After constructing this module, we then used it to create a larger 12-TU circuit (36 parts, 63 kb, Supplementary Figure S4). The additional 5 TUs that were added in the second step

include parts that are already present in the 7-TU circuit, allowing us to determine whether our hierarchical assembly method is robust to repetitive sequences, and indeed the construction was successful (Supplementary Figure S4).

DISCUSSION

The framework described here for assembling complex genetic circuits uses readily available reagents and enzymes and is reliable, efficient, modular and supports a hierarchical construction scheme. We assemble large and complex gene and demonstrate robust regulation of gene expression within these multigene circuits using small molecule inducers in human cells with both transient transfection and stable chromosomal integration. The assembly of repetitive sequences (e.g. multiple repeats of the same promoter or gene) did not result in undesirable recombination events or genetic stability issues (confirmed by sequencing, data not shown). Because our framework does not rely on restriction enzymes, it is highly flexible and can be used for assembly of components constructed using other cloning methods, (e.g. Golden Gate assembly). Also, gene circuit modules with specialized functions can be validated and stored separately, and combined when needed based on hierarchical system design, yielding large

circuits with complex phenotypes. We anticipate that our approach will be valuable for building large-scale gene circuit libraries with reliable gene expression, and will be suitable for the generation of stable cell lines with functional multielement circuits. This will greatly benefit the rapidly growing field of mammalian synthetic biology as well as facilitate genetic engineering of mammalian cells with complex multigene circuits. The method described here is not restricted to mammalian cell engineering. By extending the library of parts (genes, promoters) and appropriate modifications of inter-TU regions, the approach can support rapid genetic engineering of many other organisms.

SUPPLEMENTARY DATA

Supplementary Data are available at NAR Online, including [20–27].

ACKNOWLEDGEMENTS

We thank Dr G. Felsenfeld (National Institute of Diabetes and Digestive and Kidney Diseases, USA) for the HS4 insulator sequences.

FUNDING

Defence Advanced Research Projects Agency [HR0011-12-C-0067, DARPA-BAA-11-23]; National Science Foundation [CBET-0939511]; Synthetic Biology Engineering Research Center (SynBERC) [SA5284-11210]; National Institutes of Health [5-R01-CA155320-02]; Swiss National Science Foundation [PA00P3_131492 to P.G.]; the Ernst Schering Foundation (to P.G.). Funding for open access charge: SynBERC [SA5284-11210].

Conflict of interest statement. None declared.

REFERENCES

- Weber, W. and Fussenegger, M. (2012) Emerging biomedical applications of synthetic biology. *Nat. Rev. Genet.*, **13**, 21–35.
- Ruder, W.C., Lu, T. and Collins, J.J. (2011) Synthetic biology moving into the clinic. *Science*, **333**, 1248–1252.
- Ausländer, S., Ausländer, D., Müller, M., Wieland, M. and Fussenegger, M. (2012) Programmable single-cell mammalian biocomputers. *Nature*, **487**, 123–127.
- Vierbuchen, T., Ostermeier, A., Pang, Z.P., Kokubu, Y., Südhof, T.C. and Wernig, M. (2010) Direct conversion of fibroblasts to functional neurons by defined factors. *Nature*, **463**, 1035–1041.
- Sprinzak, D., Lakhanpal, A., LeBon, L., Santat, L.A., Fontes, M.E., Anderson, G.A., Garcia-Ojalvo, J. and Elowitz, M.B. (2010) Cis-interactions between Notch and Delta generate mutually exclusive signalling states. *Nature*, **465**, 86–90.
- Kriz, A., Schmid, K., Baumgartner, N., Ziegler, U., Berger, I., Ballmer-Hofer, K. and Berger, P. (2010) A plasmid-based multigene expression system for mammalian cells. *Nature Commun.*, **1**, 120.
- Schmid-Burgk, J.L., Xie, Z., Frank, S., Virreira Winter, S., Mitschka, S., Kolanus, W., Murray, A. and Benenson, Y. (2012) Rapid hierarchical assembly of medium-size DNA cassettes. *Nucleic Acids Res.*, **40**, e92.
- Sarrion-Perdigones, A., Falconi, E.E., Zandalinas, S.I., Juárez, P., Fernández-del-Carmen, A., Granell, A. and Orzaez, D. (2011) GoldenBraid: an iterative cloning system for standardized assembly of reusable genetic modules. *PLoS One*, **6**, e21622.
- Li, M. (2007) Harnessing homologous recombination *in vitro* to generate recombinant DNA via SLIC. *Nat. Methods*, **4**, 251–256.
- Gibson, D.G., Benders, G.A., Axelrod, K.C., Zaveri, J., Algire, M.A., Moodie, M., Montague, M.G., Venter, J.C., Smith, H.O. and Hutchison, C.A. (2008) One-step assembly in yeast of 25 overlapping DNA fragments to form a complete synthetic *Mycoplasma genitalium* genome. *Proc. Natl Acad. Sci. USA*, **105**, 20404–20409.
- Yahata, K., Maeshima, K., Sone, T., Ando, T., Okabe, M., Imamoto, N. and Imamoto, F. (2007) cHS4 insulator-mediated alleviation of promoter interference during cell-based expression of tandemly associated transgenes. *J. Mol. Biol.*, **374**, 580–590.
- Gibson, D., Young, L., Chuang, R. and Venter, J. (2009) Enzymatic assembly of DNA molecules up to several hundred kilobases. *Nat. Methods*, **6**, 12–16.
- Brown, W.R.A., Lee, N.C.O., Xu, Z. and Smith, M.C.M. (2011) Serine recombinases as tools for genome engineering. *Methods*, **53**, 372–379.
- Godiska, R., Mead, D., Dhodda, V., Wu, C., Hochstein, R., Karsi, A., Usdin, K., Entezam, A. and Ravin, N. (2010) Linear plasmid vector for cloning of repetitive or unstable sequences in *Escherichia coli*. *Nucleic Acids Res.*, **38**, e88.
- Sasaki, Y. (2004) Evidence for high specificity and efficiency of multiple recombination signals in mixed DNA cloning by the Multisite Gateway system. *J. Biotechnol.*, **107**, 233–243.
- Ishihara, K., Oshimura, M. and Nakao, M. (2006) CTCF-dependent chromatin insulator is linked to epigenetic remodeling. *Mol. Cell*, **23**, 733–742.
- Bushey, A.M., Dorman, E.R. and Corces, V.G. (2008) Chromatin insulators: regulatory mechanisms and epigenetic inheritance. *Mol. Cell*, **32**, 1–9.
- Slusarczyk, A.L., Lin, A. and Weiss, R. (2012) Foundations for the design and implementation of synthetic genetic circuits. *Nature Rev. Genet.*, **13**, 406–420.
- Litcofsky, K.D., Afeyan, R.B., Krom, R.J., Khalil, A.S. and Collins, J.J. (2012) Iterative plug-and-play methodology for constructing and modifying synthetic gene networks. *Nat. Methods*, **9**, 1077–1080.
- Drummond, A.J., Ashton, B., Buxton, S., Cheung, M., Cooper, A., Duran, C., Field, M., Heled, J., Kearse, M., Markowitz, S. *et al.* (2011) *Geneious v5.4*. Biomatter Ltd, Auckland, New Zealand.
- Botev, Z.I., Grotowski, J.F. and Kroese, D.P. (2010) Kernel density estimation via diffusion. *Ann. Stat.*, **38**, 2916–2957.
- Shortreed, M.R., Chang, S.B., Hong, D., Phillips, M., Campion, B., Tulpan, D.C., Andronescu, M., Condon, A., Hoos, H.H. and Smith, L.M. (2005) A thermodynamic approach to designing structure-free combinatorial DNA word sets. *Nucleic Acids Res.*, **33**, 4965–4977.
- Andronescu, M., Zhang, Z.C. and Condon, A. (2005) Secondary structure prediction of interacting RNA molecules. *J. Mol. Biol.*, **345**, 987–1001.
- Andronescu, M. (2003) RNAsoft: a suite of RNA secondary structure prediction and design software tools. *Nucleic Acids Res.*, **31**, 3416–3422.
- Hockemeyer, D., Soldner, F., Beard, C., Gao, Q., Mitalipova, M., DeKaveler, R.C., Katibah, G.E., Amora, R., Boydston, E.A., Zeitler, B. *et al.* (2009) Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases. *Nat. Biotechnol.*, **27**, 851–857.
- Xie, Z., Wroblewska, L., Prochazka, L., Weiss, R. and Benenson, Y. (2011) Multi-input RNAi-based logic circuit for identification of specific cancer cells. *Science*, **333**, 1307–1311.
- Livet, J., Weissman, T.A., Kang, H., Draft, R.W., Lu, J., Bennis, R.A., Sanes, J.R. and Lichtman, J.W. (2007) Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature*, **450**, 56–62.



Towards the engineering of genomes on demand

High-throughput discovery of new recombination sites for large serine recombinases



- Introduction
- Project plan
- Timeline and milestones
- Why synthetic DNA libraries?
- Applications

Xavier Duportet

Yinqing Li

Ron Weiss



NEEDS AND LIMITATIONS

We believe Synthetic Biology hold promise to revolutionize the world in which we are living.

As one of the very few Synthetic Biology lab focusing on application in mammalian cells, we are facing problems for which technologies have yet to be discovered.

A MODULAR AND EFFICIENT ASSEMBLY OF DNA LIBRARY PARTS

Implementing new biological/therapeutic functions in mammalian cells, as well reconstructing metabolic pathways in bacteria require the design and assembly of large genetic regulatory networks. As our predictive abilities on the circuit functioning are still low, it is critical to be able to build a large number of possible combinations for the circuit in the shortest amount of time possible. Therefore, we need a technology to allow a rapid and modular assembly of dozens of genetic parts.

Assembling a high number of DNA pieces is still very inefficient and the error rate increases with the number of pieces
Gibson, Golden-gate, Gateway

B EFFICIENT TOOLS FOR MAMMALIAN GENOME ENGINEERING

To enable therapeutical application of synthetic biology based on DNA circuits, we will need to come up with a technology which not only allows the stable integration of the circuit into the host genome, but also a safe and copy-controlled integration process.

Homologous-recombination based techniques to integrate circuits into mammalian cells are extremely inefficient even for large DNA fragment, present many off-target effects, and can't control the scar it will make in the genome and are often toxic to the cells
TALEN, Zinc-Finger Nuclease, Meganuclease

Integration of therapeutic circuits in mammalian cells with viral vectors is compromised given the size constraints due to the small size of the viral particle capsid, plus it becomes very hard to control the integration locus and the number of integration events
Lentivirus

➡ We propose to focus on Large Serine Recombinases as a tool to solve most of these problems.



- Introduction
- Project plan
- Timeline and milestones
- Why synthetic DNA libraries?
- Applications

- Needs and limitations
- Current limitations

EXPLORING LARGE SERINE RECOMBINASES

Large Serine recombinase combine extremely interesting properties

- **specific directionality:** they mediate unidirectional recombination reaction between two different DNA sequences: attP (phage attachment site) and attB (bacterial attachment site), which prevent the reversion of the reaction
=> small tyrosine recombinase favor the excision reaction and therefore are very inefficient for integration (Cre, Flp)
- **high efficiency across organisms:** they have been proven to work in a very broad range of organisms (bacteria, yeast, insect, bird, mammals)
=> less constraints than the large tyrosine recombinases which need a host-specific cofactor and therefore are very inefficient outside of their native host

And have been successfully applied for interesting applications

- A few biotech companies are using PhiC31 to integrate functional genes for gene therapy, or for recombinant protein production.
- Bonnet et al (PNAS, 2012) used BxB1 to create a memory-device in bacterial cells.

We recently discovered 3 new large serine recombinase, and tested their properties for different SynBio applications:

- **in-vitro results:** - we found their native recombination sites and created up to 16 orthogonal mutant recombination site pairs;
- in-vitro recombination assays showed that 99% of the recombination events happened in 5min whereas it takes 3-5h for state of the art multi-fragment recombination systems (i.e. Gateway)
- **in-vivo results:** - integration of large DNA fragments (genetic circuits) in a single locus of human genome was between 25x and 40x more efficient compared to tyrosine recombinases
- specificity of the integration events reached 95% vs 30-40% with other recombinases



Introduction
Project plan
Timeline and milestones
Why synthetic DNA libraries?
Applications

Needs and limitations
Large serine recombinases

TIME FOR HIGH-THROUGHPUT DISCOVERY

Based on these exciting foundational results, we are looking forward to expanding the library of large serine recombinases available for genome engineering and DNA manipulation

➔ Having at our disposal a large library of site-specific and orthogonal large serine recombinase will indeed allow us to develop the following applications:

- **Highly efficient multi-fragment DNA assembly technology**
We can develop an assembly framework relying on a cocktail of orthogonal recombinases, instead of using a single enzyme to assemble multiple fragments (Gateway system, inefficient for more than 3 piece-assembly).
Moreover, no optimization of the reaction conditions is needed (unpublished data) given that the recombinase does not need any cofactor.
It could therefore become enable the assembly of 2 to 40 DNA fragments of any size within minutes.
- **Gene and cell therapy**
With a large number of recombination sites in hand, we can find a safe and highly expressed locus in the human genome to integrate large therapeutical circuits.
We can evolve natural recombinase to target new recombination sites using the knowledge gained from this large library of serine recombinases.
- **Artificial chromosomes design**
We will use a panel of orthogonal recombination sites on an artificial chromosome to allow the easy plug-in of genetic circuits into the artificial chromosome.

BUT there are only 4 available pairs of recombinase/recombination sites known.

Therefore, we propose to synthesize a large number of candidate recombinases to discover their target sites using high-throughput methods.



Introduction
Project plan
Timeline and milestones
Why synthetic DNA libraries?
Applications

Our vision of SynBio
Current limitations
Large serine recombinases

PROJECT PLAN OVERVIEW

- ✓ Identification of a library of interesting predicted recombinases
- ① DNA synthesis of recombinases and their attP site region
- ① Assembly of functional integrative vectors
- ② Integration experiments in the phage's host cell
- ③ Analysis of the integration site and attB site inference
- ④ Validation in non-native other hosts: bacterial and mammalian cells

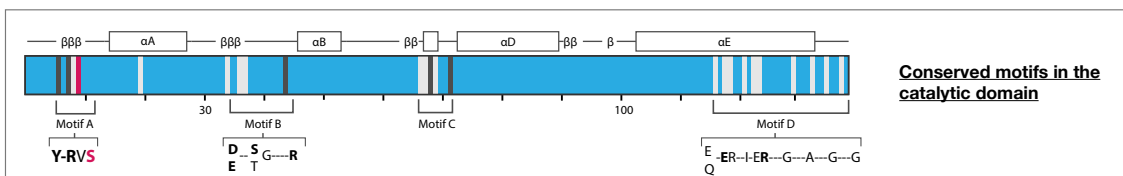


Introduction
Project plan
Timeline and milestones
Why synthetic DNA libraries?
Applications

Our vision of SynBio
Current limitations
Our proposal

LIBRARY IDENTIFICATION

Based on conserved motifs of the available large serine recombinases, we have found more than 3000 other putative candidates.



However, some of these recombinases are produced by phage infected bacteria species hard to handle in the lab.

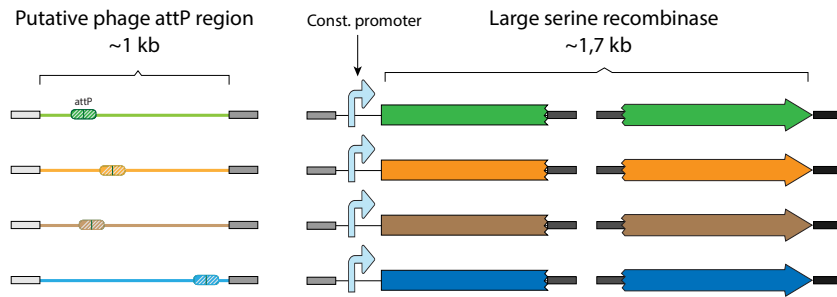
Therefore, we reduced this number to 214 optimal predicted recombinases that we would like to synthesize.



Introduction
Project plan
Timeline and milestones
Why synthetic DNA libraries?
Applications

Our vision of SynBio
Current limitations
Large serine recombinases

High-throughput DNA library synthesis



We plan to have synthesized a library of predicted recombinases coding sequences with their putative attP site sequence. We have proven that we can restrict the location of the attP site region to a ~1kb region from the phage genome.

As the current size limitation for DNA synthesis oligos is around 1kb, we will need to synthesize two separate DNA fragments to recreate a functional recombinase coding sequence, using the highly efficient Golden Gate assembly method.

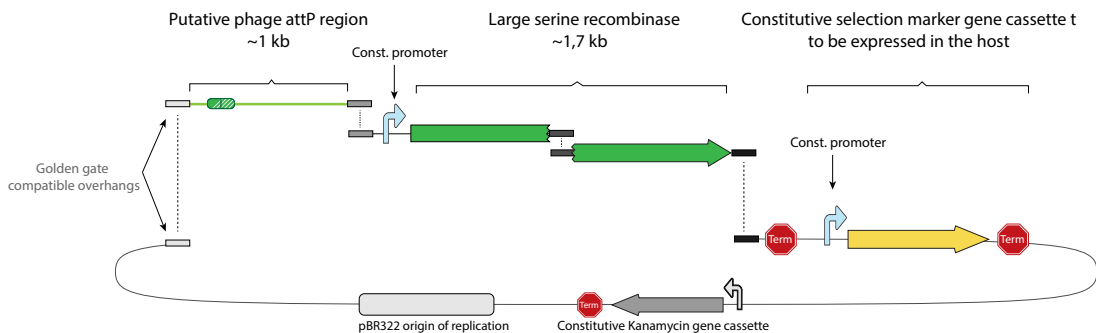
In our design, a constitutive promoter will be placed in front of the recombinase sequence to allow expression of the recombinase in the corresponding host.



- Introduction
- Project plan**
- Timeline and milestones
- Why synthetic DNA libraries?
- Applications

- Overview
- DNA library synthesis**
- Oligo assembly
- Integration in the host
- Identification of attB sites
- Validation in other hosts

Integrative vectors creation



We will create a library of integrative vectors composed of:

- the putative phage attP region
- a gene cassette to allow expression of the corresponding large serine recombinase in the host;
- a gene cassette to allow expression of a resistance marker to screen for resistant host cells upon integration;
- a gene cassette to allow expression of a resistance marker to screen correct assembly products in E.coli;
- an origin of replication allowing plasmid maintenance in E.coli but not in the host cells.

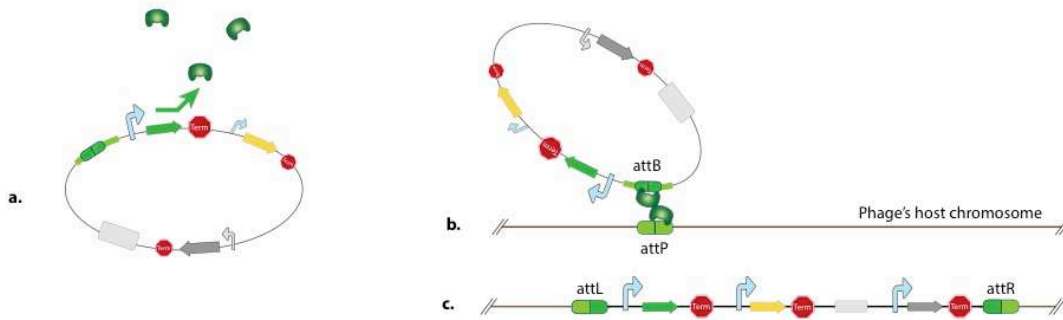
* Done using high-throughput liquid handling robot*



- Introduction
- Project plan**
- Timeline and milestones
- Why synthetic DNA libraries?
- Applications

- DNA library synthesis
- Oligo assembly**

Integration of the vectors into the host



We will transform each assembled integration vector into its corresponding host and screen for integration events.

a. Upon transformation into the phage's host cell, the recombinase is expressed.

b. The recombinase will then bind the attP site contained within the putative region cloned on the integrative vector and the unique attB on the genome of the phage's host cell.

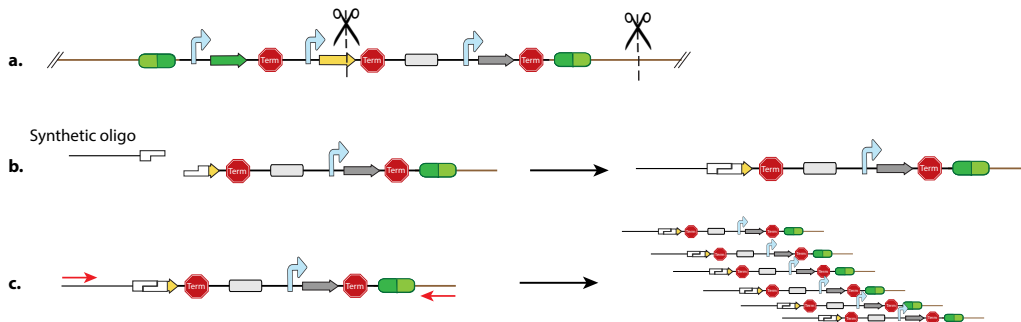
c. Once the recombination happened, the plasmid will be stably integrated within the phages's host genome and will allow the cell to survive antibiotic selection given that the integrative plasmid provides the antibiotic resistance gene.



Introduction
Project plan
 Timeline and milestones
 Why synthetic DNA libraries?
 Applications

DNA library synthesis
 Oligo assembly
Integration in the host

Identification of attB sites



a. Digestion: Once the genomic DNA has been extracted from resistant cells, we digest it with a specific restriction enzyme cutting only once in the vector and multiple times in the genome to generate small fragments of DNA.

b. Ligation: We will then ligate a small oligo with a complementary overhang to the digestion products.

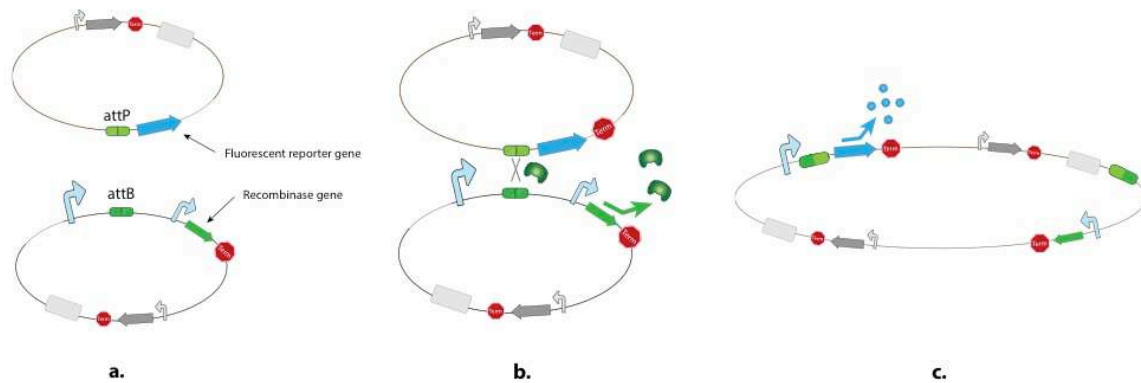
c. PCR and sequencing: Using a primer pairs binding on the ligated oligo and on the integrative vector, we specifically amplify the correct ligation products. Sequencing the PCR products will provide us with the attR site. We will then be able to infer from this result the original attB site recognized by the recombinase to integrate the whole vector.



Introduction
Project plan
 Timeline and milestones
 Why synthetic DNA libraries?
 Applications

DNA library synthesis
 Oligo assembly
 Integration in the host
Identification of attB sites
 Further work

VALIDATION OF TARGET SITES IN OTHER HOSTS



- a. Transformation** of a reporter plasmid harbouring the attP site placed before a promoterless fluorescent reporter gene; and the recombinase expressing plasmid harbouring the attB site downstream of a constitutive promoter.
- b. Recombination** of the two sites leading to plasmid fusion which places the fluorescent reporter downstream of the constitutive promoter.
- c. Screening** of fluorescent cells. The percentage of fluorescent cells will be compared to the previous results obtained with the known recombinases.

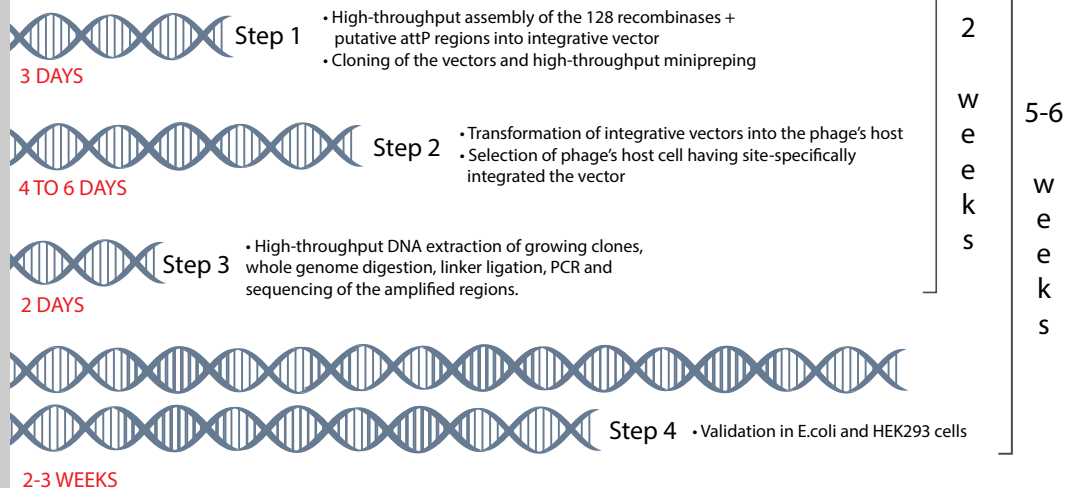


Introduction
 Project plan
 Timeline and milestones
 Why synthetic DNA libraries?
 Summary

DNA library identification and synthesis
 DNA assembly
 Integration in the host
 Identification of attB sites
 Validation in other hosts

TIMELINE AND MILESTONES

Synthetic DNA sequences delivery



Introduction
 Project plan
 Timeline and milestones
 Why synthetic DNA libraries?
 Applications

WHY SYNTHETIC DNA LIBRARIES?

- Quickly test a large number of recombinase coding sequences
=> physical DNA from rare phages sequenced in the past decades is barely available
- Codon optimization of the coding sequence to improve expression and recombination efficiency in non-native host
=> native nucleotide sequence results in lower recombination efficiency in non-native host (cf manuscript in preparation)

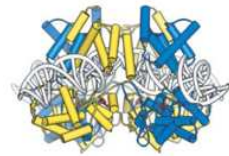
Required synthetic sequences:

- 130 wild-type recombinases => 260 sequences to synthesize
- 130 corresponding attP regions
- 50 codon optimized recombinase for mammalian expression
=> 100 sequences to synthesize

TOTAL of 490 required synthetic sequences



Introduction
Project plan
Timeline and milestones
Why synthetic DNA libraries?
Summary

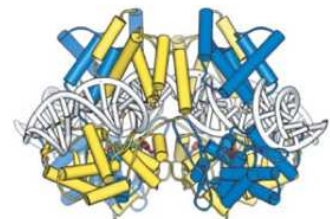


SUMMARY

- We identify large serine recombinase as enabling technology for many SynBio applications
DNA assembly, Metabolic Engineering, Genome Engineering, Therapeutics
- We have ready-to-publish results demonstrating the usage of large serine recombinases for genome engineering
- We propose a high-throughput workflow to expand our very limited library of large serine recombinase in a short period of time
- The success of this project relies on the synthesis of a large amount of synthetic DNA sequences
- The generated data will provide fundamental knowledge to ease the engineering of recombinases



Introduction
Project plan
Timeline and milestones
Why synthetic DNA libraries?
Summary



COMMISSION INNOVATION 2030



LA BIOLOGIE
DE SYNTHÈSE
EN FRANCE

Panorama 2013

Xavier Duportet et Aleksandra Nivina

TABLE DES MATIÈRES

AVANT-PROPOS	1
INTRODUCTION	1
Définitions de la Biologie de Synthèse	2
Les grandes applications de la Biologie de Synthèse	2
La place de la France dans le monde	3
RECHERCHE PUBLIQUE	5
Laboratoires dont l'un des principaux axes de recherche est la Biologie de Synthèse, et qui ont déjà publié dans ce domaine	5
Laboratoires qui travaillent/envisagent travailler sur des projets de Biologie de Synthèse	7
Autres personnalités éminentes en Biologie de Synthèse en France	8
Chercheurs en Biologie de Synthèse français, à l'étranger	9
RECHERCHE PRIVEE	10
Sociétés	10
Pôles de compétitivité	13
Conseil	14
EDUCATION	15
Les formations	15
Les équipes étudiantes iGEM	15
ANNEXES	18
Publications des équipes de recherche françaises (liste non exhaustive)	18
Carte des acteurs de la Biologie de Synthèse en France	19

AVANT-PROPOS

Ce rapport a pour but de recenser le plus exhaustivement possible les différents acteurs de la Biologie de Synthèse en France en 2013, et non d'apprécier l'importance de la discipline. Nous avons décidé d'utiliser un ordre alphabétique pour les différents classements dans le rapport.

Nous avons rédigé ce rapport en tant que jeunes scientifiques, doctorants en Biologie de Synthèse et impliqués dans le développement de la discipline en France mais aussi à l'étranger. Lors de l'écriture de ce rapport, nous avons pris conscience de deux points importants, que nous souhaitons mentionner dans cet avant-propos:

- La France possède quelques laboratoires de Biologie de Synthèse d'envergure internationale, et l'intérêt pour cette discipline est de plus en plus important au sein des chercheurs français. Il est toutefois important de souligner que le terme Biologie de Synthèse est devenu très «à la mode» et de nombreux laboratoires, dans le monde entier, indiquent appartenir à cette discipline sans toutefois réellement ni en utiliser les outils, ni en développer.
- Dans le cadre de la recherche privée, l'appartenance à cette discipline est au contraire quelque peu crainte en France. En effet, la plupart des startups n'apprécient guère être associées à la Biologie de Synthèse par peur de l'image négative que lui donnent plusieurs groupes associatifs.

INTRODUCTION

La Biologie de Synthèse a pour but de concevoir et de construire des nouveaux systèmes biologiques, ou de modifier ceux déjà existant, afin de réaliser des nouvelles tâches utiles pour l'homme. Selon l'Office Gouvernementale pour la Science du Royaume-Uni, la «Biologie de Synthèse a le potentiel de transformer l'industrie, la recherche et les emplois (dans le domaine des sciences du vivant) d'une manière encore plus compétitive que l'a fait l'informatique». Cela comprend un large panel de technologies en cours de développement, depuis la modification génétique avancée de bactéries jusqu'à la création de nouvelles espèces ou de nouveaux organes. La Biologie de Synthèse permet dès à présent de produire des molécules chimiques et des médicaments à prix réduits, et pourrait très rapidement ouvrir des nouveaux horizons dans des domaines très variés tels que la production de nourriture, la production d'énergie, la décontamination de l'eau, les biomatériaux voir même le traitement d'informations.

En France, le domaine se structure depuis environ cinq ans, et la Stratégie nationale de recherche et d'innovation a défini la Biologie de Synthèse comme « prioritaire » en 2009 et mandaté un groupe de travail pour évaluer les développements, les potentialités et les défis associés. Le rapport publié en 2011 (tinyurl.com/SNRI-rapport-BS) identifie la Biologie de Synthèse comme une "priorité de la stratégie nationale de recherche et d'innovation", et estime que "il existe en France un gisement de compétences à mobiliser, permettant de viser une position mondiale de second ou troisième".

En 2012, l'office Parlementaire D'évaluation Des Choix Scientifiques Et Technologiques a rédigé un rapport sur les enjeux de la Biologie de Synthèse (tinyurl.com/OPECST-rapport-BS), qui conclue qu' "Investir dans la Biologie de Synthèse est donc essentiel pour mieux comprendre le fonctionnement du vivant", et qui émet plusieurs recommandations pour le développement maîtrisé, en toute transparence, de la Biologie de Synthèse.

Définitions de la Biologie de Synthèse

Le domaine de la Biologie de Synthèse évolue tellement rapidement qu'il n'a pas encore été trouvé une définition acceptée et partagée par toute la communauté. Toutefois, on retrouve dans les différentes définitions la même notion que la Biologie de Synthèse est une application des principes d'ingénierie aux composants fondamentaux de la biologie.

Trois caractérisations de la Biologie de Synthèse permettent d'appréhender la substance des débats.

- De Lorenzo et Danchin décrivent la Biologie de Synthèse comme un cadre de travail théorique et pratique qui aborde les systèmes biologiques à travers: 1) les outils et le langage propre aux circuits électroniques et 2) les processus de manufacture mécanique; ceci afin de parvenir à combiner d'une manière rationnelle des briques biologiques standardisées et découplées de leur contexte naturel. Ces briques deviennent alors des éléments fonctionnels individuels, qui peuvent être combinés dans de nouvelles configurations afin de modifier les propriétés existantes ou d'en créer de nouvelles. (De Lorenzo and Danchin 2008: 822).
- Endy divise les Biologistes de Synthèse en 4 groupes: les biologistes, les chimistes, les ingénieurs et ceux qui ré-écrivent. Pour les biologistes, la Biologie de Synthèse offre le moyen de mieux comprendre les systèmes biologiques naturels. Pour les chimistes, c'est une extension de la chimie de synthèse pour le développement de nouvelles molécules. Pour les ingénieurs, c'est une technologie qui permet de simplifier la conception et la construction de systèmes biologiques modifiés. Enfin, pour les ré-écrivains, la Biologie de Synthèse permet d'optimiser les systèmes biologiques principalement par le recodage des génomes naturels. (Endy, 2005)
- Benner et Sismour identifient deux classes de «Biologistes de synthèse». La première classe s'intéresse à l'assemblage de composants synthétiques afin de créer des systèmes chimiques qui sont soumis à l'évolution Darwinienne. La deuxième classe utilise des principes d'ingénierie afin d'extraire des composants interchangeables des systèmes vivants existant pour créer des nouvelles unités de construction qui peuvent ne pas être analogues aux systèmes biologiques que nous connaissons. Les deux classes s'intéressent à la synthèse chimique des composants biologiques, depuis les circuits de simples gènes jusqu'à l'assemblage de génomes entiers. Toutefois, la première classe s'en sert pour comprendre la biologie alors que la deuxième tente de la modifier, voire de la recréer. (Benner and Sismour, 2005)

Les grandes applications de la Biologie de Synthèse

Dans le domaine du biomédical, les systèmes développés par Biologie de Synthèse sont utilisés à la fois pour comprendre le mécanisme des maladies et pour développer de nouveaux traitements préventifs et thérapeutiques. Ainsi, des biocapteurs capables de détecter des stimuli variés tels que des marqueurs de maladie ont été développés. Un des exemples de cette approche est l'ingénierie d'un circuit d'ARN capable de détecter un certain type de cellules cancéreuses, et d'induire leur mort programmée (Xie et al. Science Sept. 2, 2011 333: 1307-11) ou encore la conception du système cellulaire activé par la lumière qui produit des peptides atténuant les symptômes du diabète (Ye et al. Science Jun. 24, 2011 332: 1565-68). La production de vaccins a été également facilitée grâce à un séquençage et une synthèse d'ADN plus rapide et moins chère. En 2009 par exemple, des gènes synthétiques codant pour des protéines d'enveloppe du virus H1N1 ont pu être produits dans un délai de 5 jours seulement. La Biologie de Synthèse permet aussi de **découvrir de nouvelles molécules thérapeutiques, ou bien d'améliorer leur production**, comme ceci a été fait pour l'artémisinine, utilisée dans le traitement du paludisme (Paddon et al. Nature Apr. 25, 2013 496: 528-32). Il est donc fortement envisageable que dans un futur proche, la Biologie de Synthèse permette de mettre au point des thérapies personnalisées en fabriquant des médicaments adaptés au génome du patient, ou encore d'effectuer des transplantations en utilisant des cellules ou des tissus modifiés génétiquement pour le patient.

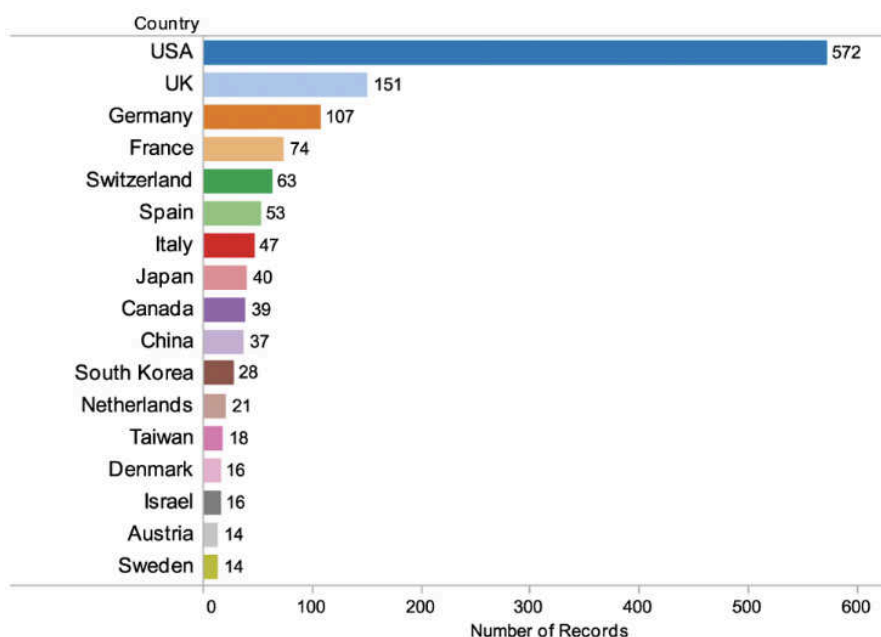
Dans le domaine de l'agro-alimentaire, les efforts de recherche tendent à optimiser l'utilisation des ressources (eau, engrais) ou la réutilisation des matières biologiques pour la production de nouveaux composés, comme par exemple l'ingénierie des voies métaboliques des bactéries pour la production des surfactants à partir de cellulose (Reznik et al. Appl. Microbiol. Biotechnol. May 2010 86: 1387-97). La Biologie de Synthèse permet également de développer des plantes aux propriétés nouvelles: meilleur rendement par optimisation de la voie de photosynthèse ou modification des voies métaboliques, résistance accrue à certains ions, à la sécheresse etc.. Il est important de souligner que ces efforts se développent en parallèle d'une amélioration de la sûreté de l'ingénierie des Organismes Génétiquement Modifiés.

Dans le domaine des biotechnologies blanches, la Biologie de Synthèse ouvre des possibilités de bioproduction de nombreuses molécules et matériaux, tels que la soie, les plastiques (Jung and Lee J.Biotechnol. Jan. 10, 2011 151: 94-101), d'autres polymères, mais aussi des biocarburants (Atsumi et al. Nature Jan. 3, 2008 451: 86-89). En effet, de très nombreuses équipes travaillent sur le développement de nouvelles sources d'énergie qui, grâce à la construction de nouvelles voies de synthèse dans des microorganismes, améliorent les rendements et diminuent les coûts de production.

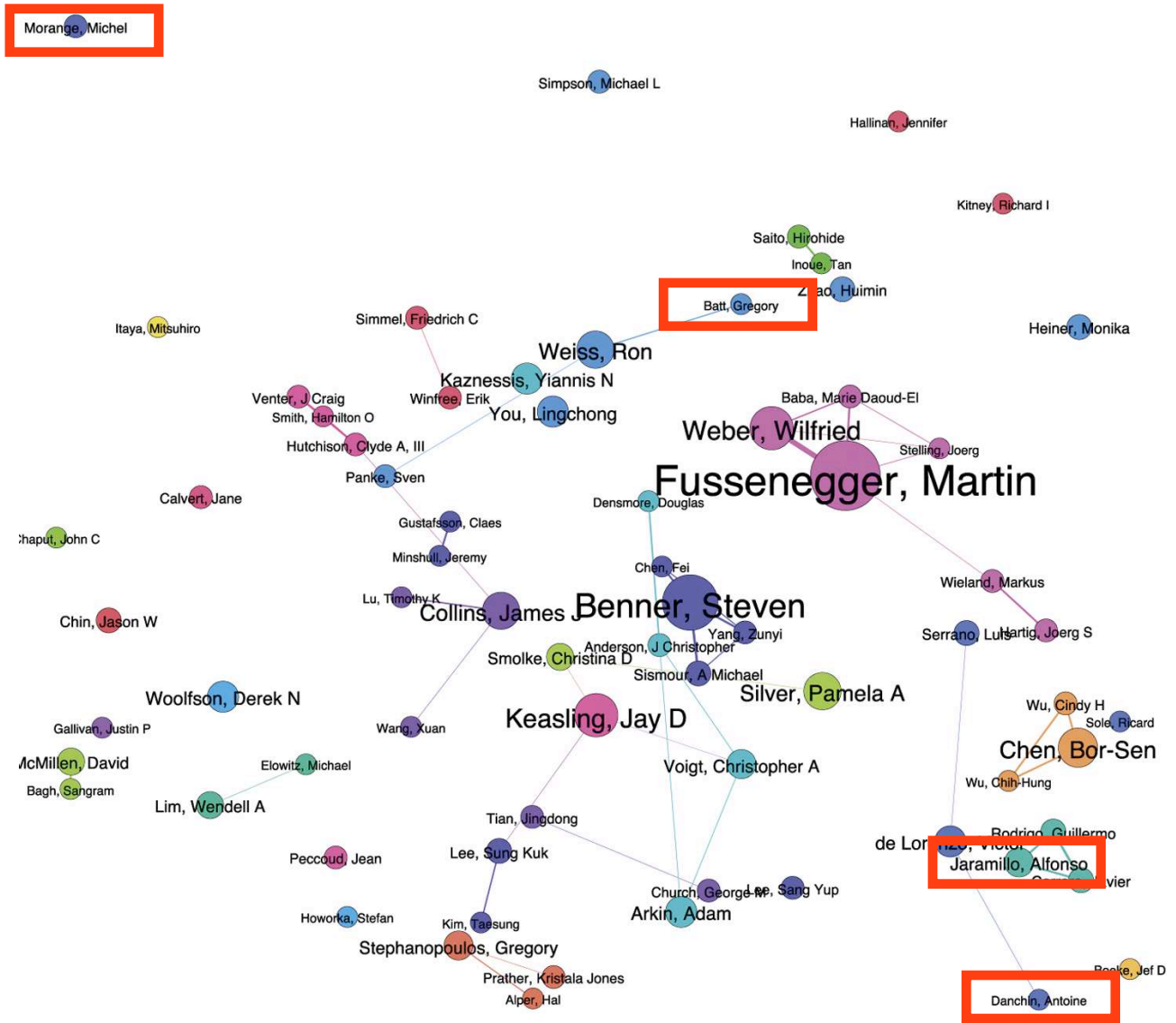
Et enfin, dans le secteur de l'environnement, de nombreux chercheurs travaillent à créer des microorganismes capable d'éliminer des polluants et de purifier les eaux, en tenant compte des problèmes de la sûreté biologique de telles approches (Schmidt et al. FEBS Lett. Jul. 16, 2012 586: 2199-206). On assiste surtout au développement de systèmes biocapteurs capables de détecter des métaux lourds ou des molécules nocives, qui sont couplés à des voies d'assimilation ou de dégradation afin d'assurer la purification de l'environnement.

La place de la France dans le monde

En Janvier 2012, on totalisait 1255 publications sur la Biologie de Synthèse et Génomes de Synthèse, regroupant des travaux de recherche scientifiques, sociaux et éthiques, provenant de plus de 40 pays. La Biologie de Synthèse est une discipline largement dominée par les Etats-Unis, suivis par le Royaume Uni, l'Allemagne puis la France et la Suisse (Oldham P, Hall S, Burton G. PLoS ONE. 2012 7(4): e34368.).



Sur plus de 2934 auteurs identifiés comme ayant publié au moins un article dans le domaine de la Biologie de Synthèse, 4 chercheurs travaillant en France se placent dans le top 40: Gregory Batt, Antoine Danchin, Alfonso Jaramillo, et Michel Morange.



RECHERCHE PUBLIQUE

Laboratoires dont l'un des principaux axes de recherche est la Biologie de Synthèse, et qui ont déjà publié dans ce domaine

- **Architecture and evolution of eukaryotic genetic circuits - SyntheCell** (CNRS IGDR, Rennes)

Utilisation de la Biologie de Synthèse comme outil pour étudier la régulation du cycle cellulaire chez la levure.

Contact Biologie de Synthèse: damien.coudreuse@univ-rennes1.fr

Site web: <http://goo.gl/gT5PJ>

- **Biologie du Fruit et Pathologie** (INRA, Bordeaux)

Modification du métabolisme de *Mycoplasma* - Ingénierie de génomes minimaux.

Contacts Biologie de Synthèse: carole.lartigue-prat@bordeaux.inra.fr, alain.blanchard@bordeaux.inra.fr

Site web: <http://www6.bordeaux-aquitaine.inra.fr/bfp>

- **BioRetroSynth** (CNRS ISSB, Evry)

Ingénierie métabolique des composés thérapeutiques par bio-retrosynthèse et développement des outils bioinformatiques sous-jacents.

Contact Biologie de Synthèse: Jean-Loup.Faulon@univ-evry.fr

Site web: <http://www.issb.genopole.fr/~faulon/>

- **Contraintes** (INRIA, Paris Rocquencourt)

Design et modélisation *in-silico* de circuits génétiques et tissus artificiels.

Contacts Biologie de Synthèse: gregory.batt@inria.fr

Site web: <http://contraintes.inria.fr/~batt/home.html>

- **Hersen Lab** (CNRS MSC, Paris)

Contrôle de l'expression des gènes en temps réel à l'échelle de la cellule unique ou d'une population (bactéries, levures, cellules mammifères).

Contact Biologie de Synthèse: pascal.hersen@univ-paris-diderot.fr

Site web: <http://www.msc.univ-paris-diderot.fr/~phersen/513/Lab.php>

- **Heterogeneous systems and microsystems** (CNRS INESS, Strasbourg)

Développement d'outils bio-informatiques d'aide à la conception pour la Biologie de Synthèse.

Contact Biologie de Synthèse: christophe.lallement@ensps.u-strasbg.fr

Site web: <http://www-iness.c-strasbourg.fr/-Heterogeneous-systems-and->

- **Ibis** (INRIA & CNRS, Grenoble)

Mesure de l'expression des gènes - Modélisation des réseaux de régulation génétique et étude de l'adaptation des bactéries aux stress environnementaux.

Contacts Biologie de Synthèse: Hidde.de-Jong@inria.fr, hans.geiselman@ujf-grenoble.fr

Site web: <http://ibis.inrialpes.fr/>

- **Ingénierie et Evolution des Voies Métaboliques chez les Procaryotes** (INSA LISBP, Toulouse)

Analyse du métabolisme des procaryotes et création de nouvelles voies métaboliques de composés organiques utilisables en tant que carburant.

Contact Biologie de Synthèse: isabelle.meynial-salles@insa-toulouse.fr

Site web: http://www.lisbp.fr/fr/la_recherche/axe_physiologie_et_metabolisme_microbiens/equipe_ead3.htm

- **Ingénierie Moléculaire et Métabolique** (INSA & CNRS, Toulouse)

Ingénierie métabolique chez la levure - Ingénierie combinatoire de biocatalyseurs et de structures protéo-nucléiques artificielles.

Contact Biologie de Synthèse: denis.pompon@insa-toulouse.fr

Site web: http://www.lisbp.fr/fr/la_recherche/axe_biocatalyse/ingenierie_moleculaire_et_metabolique.html

- **Laboratoire de Métagénomique des procaryotes** (CEA Genoscope, Evry)

Recherche des solutions biologiques pour la chimie de synthèse et travail sur le développement de nouveaux systèmes hôtes utilisables en Biologie de Synthèse.

Contact Biologie de Synthèse: jean.weissenbach@genoscope.cns.fr

Site web: <http://www.genoscope.cns.fr/spip/Cloaca-Maxima.html>

- **MEGA** (CRNS ISSB, Evry)

Design et construction de circuits génétiques - Etude de l'organisation des génomes

Contact Biologie de Synthèse: francois.kepes@epigenomique.genopole.fr

Site web: <http://www.issb.genopole.fr/Research/teams/mega>

- **Plasticité de Genome Bactérien** (Institut Pasteur, Paris)

Etude et Ingénierie des chromosomes bactériens - Développement d'une plateforme de recombinaison pour assemblage de gènes synthétiques.

Contact Biologie de Synthèse: didier.mazel@pasteur.fr

Site web: <http://openwetware.org/wiki/Mazel>

- **Projet SYNBIOTIC** (IBISC, Evry)

Développement d'un langage de programmation pour la Biologie de Synthèse, afin de construire des réseaux génétiques *in-silico*.

Contact Biologie de Synthèse: franck.delaplace@ibisc.univ-evry.fr

Site web: <http://synbiotic.spatial-computing.org/start>

- **Synth-Bio** (CNRS ISSB, Evry)

Automatisation de la conception et de la caractérisation des réseaux génétiques et réseaux d'ARN; automatisation des processus d'évolution dirigée.

Contact Biologie de Synthèse: Alfonso.Jaramillo@issb.genopole.fr

Site web: <http://jaramillolab.issb.genopole.fr/display/sbsite/Jaramillo+Lab>

- **SysDiag** (CNRS et BioRad, Montpellier)

SysDiag est un laboratoire né en 2007 de l'alliance de deux partenaires aux ambitions complémentaires : le CNRS et Bio-Rad, dédié à la recherche, découverte et innovation dédié au diagnostic en santé humaine.

Site web: <http://www.sysdiag.cnrs.fr/>

Contact : franck.molina@sysdiag.cnrs.fr

- **TaMaRa** (INSERM, Paris)

Etude du vieillissement, de la diversité phénotypique, et de la coopération chez les bactéries.

Contacts Biologie de Synthèse: ariel.lindner@inserm.fr, francois.taddei@inserm.fr

Site web: <http://www.necker.fr/tamara/index.html>

- **Xenome** (CNRS ISSB, Evry)

Développement d'acides nucléiques synthétiques, orthogonaux aux acides nucléiques naturels.

Contacts Biologie de Synthèse: Philippe.Marliere@issb.genopole.fr, piet.herdewijn@issb.genopole.fr

Site web: <http://www.issb.genopole.fr/Research/teams/xenome>

Laboratoires qui travaillent/envisagent travailler sur des projets de Biologie de Synthèse

- **Biotechnologie et signalisation cellulaire** (CNRS IREBS, Strasbourg)

Biotechnologie et signalisation cellulaire; exploration de futures stratégies thérapeutiques synthétiques.

Contact Biologie de Synthèse: georges.orfanoudakis@unistra.fr

Site web: <http://irebs.cnrs.fr/spip.php?article57&lang=fr>

- **Contrôle par l'ARN de l'expression génétique** (CNRS IBPC, Paris)

Etude du contrôle de la traduction chez les bactéries.

Contact Biologie de Synthèse: springer@ibpc.fr

Site web: http://www.ibpc.fr/UPR9073/equipe_Mathias/AccueilMSpringer.htm

- **D-Phy-Chloro** (CEA iRTSV/PCV, Grenoble)

Etude et modélisation des régulations dans le chloroplaste.

Contacts Biologie de Synthèse: nrolland@cea.fr, Jerome.Garin@cea.fr

Site web: <http://goo.gl/rAoYI>

- **Equipe Preuves, Programmes, Systèmes** (CNRS PPS, Paris)

Développement des langages de programmation et modélisation pour la Biologie de Synthèse

Contact Biologie de Synthèse : vincent.danos@pps.jussieu.fr

Site web: <http://www.pps.univ-paris-diderot.fr/~danos/>

- **Fonctionnalité et Ingénierie des Protéines** (CNRS UFIP, Nantes)

Ingénierie de protéines afin de concevoir des protéines possédant des acides aminés non naturels.

Contact Biologie de Synthèse: charles.tellier@univ-nantes.fr

Site web: <http://www.ufip.univ-nantes.fr/presentation/?lang=fr>

- **Laboratoire Colloïdes et Matériaux Divisés** (ESPCI, Paris)

Développement d'outils pour la Biologie de Synthèse: microfluidique et liposomes.

Contact Biologie de Synthèse: jerome.bibette@espci.fr

Site web: <http://www.lcmd.espci.fr/fr/Equipe>

- **LCAB** (CEA Genoscope, Evry)

Génomique fonctionnelle pour la découverte de nouveaux biocatalyseurs.

Contact Biologie de Synthèse: vberard@genoscope.cns.fr

Site web: <http://www.cns.fr/spip/Equipe-LCAB.html>

- **Metamorphosys** (CNRS ISSB, Evry)

Etude des xénopes en tant que modèle de la structure, de l'expression et de l'évolution du génome.

Contact Biologie de Synthèse: nicolas.pollet@issb.genopole.fr

Site web: <http://www.issb.genopole.fr/Research/teams/metamorphosys>

- **Métabolisme Energétique des *Streptomyces*** (CNRS, Orsay)

Etude et modification du le métabolisme des *Streptomyces* afin de produire du bio-diesel.

Contact Biologie de Synthèse: marie-joelle.virolle@igmors.u-psud.fr

Site web: <http://www.igmors.u-psud.fr/spip.php?rubrique155&lang=fr>

- **Microbiologie moléculaire des Actinomycètes** (CNRS IGM, Orsay)

Etude du métabolisme des *Streptomyces* et production de composés chimiques.

Contact Biologie de Synthèse: jean-luc.pernodet@igmors.u-psud.fr

Site web: <http://www.igmors.u-psud.fr/spip.php?rubrique140&lang=fr>

- **Résistance aux médicaments** (CNRS BMSSI, Lyon)

Régulation artificielle des transporteurs de drogues des cellules cancéreuses.

Contact Biologie de Synthèse: attilio.dipietro@ibcp.fr

Site web: <http://www.ibcp.fr/bmssi/-Resistance-aux-medicaments-A-Di-?lang=fr>

- **Structure des macromolécules biologiques & mécanismes de reconnaissance** (CNRS IBMC, Strasbourg)

Etude des relations entre la séquence, la structure, l'évolution et les fonctions d'ARN à fonctions enzymatiques.

Contact Biologie de Synthèse: E.Westhof@ibmc.u-strasbg.fr

Site web: <http://www-ibmc.u-strasbg.fr/upr9002/westhof/index.html>

Autres personnalités éminentes en Biologie de Synthèse en France

- **Antoine Danchin**

Scientifique et entrepreneur, Prof. Antoine Danchin est connu pour ses réflexions dans le domaine de la Biologie de Synthèse, sur ses défis et son avenir. Il est le fondateur d'une société de Biologie Symplectique: AMAbiotics.

Contact : antoine.danchin@normalesup.org

Site web: http://www.normalesup.org/~adanchin/AD/CVenglish_AD.html

- **Michel Morange**

Scientifique et philosophe, Prof. Michel Morange s'intéresse à la question des transformations de la biologie au XX^{ème} siècle, et de l'émergence de nouvelles disciplines telles que la Biologie de Synthèse et de Systèmes.

Contact : morange@biologie.ens.fr

Site web: http://www-ihpst.univ-paris1.fr/108,michel_morange.html

- **Joël de Rosnay**

Scientifique, prospectiviste, conférencier et écrivain français, Joël de Rosnay est spécialiste des origines du vivant. Il s'intéresse particulièrement aux technologies avancées et aux applications de la théorie des systèmes. Il a notamment rédigé un livre de vulgarisation sur la Biologie de Synthèse: *Et l'homme créa la vie* (édition Les Liens qui Libèrent, 2010).

Contact : derosnay@club-internet.fr

Site web: <http://www.carrefour-du-futur.com/>

Chercheurs en Biologie de Synthèse français, à l'étranger

- **David Bikard** (Rockefeller University, USA)

Post-doc dans l'équipe de Luciano Marraffini

Contact: david.bikard@rockefeller.edu

Site web: <http://www.rockefeller.edu/research/faculty/labheads/LucianoMarraffini/>

- **Jérôme Bonnet** (University of California Stanford, USA)

Post-doc dans l'équipe de Drew Endy

Contact: bonnet@stanford.edu

Site web: <http://openwetware.org/wiki/Endy:Lab>

- **Guillaume Cambray** (University of California Berkeley, USA)

Post-doc dans l'équipe de Adam Arkin

Contact: cambray.guillaume@gmail.com

Site web: <http://genomics.lbl.gov/>

- **Xavier Duportet** (co-tutelle entre l'INRIA, France et le Massachusetts Institute of Technology, USA)

Doctorant chez Gregory Batt à l'INRIA et chez Ron Weiss au MIT.

Contact: duportet@mit.edu

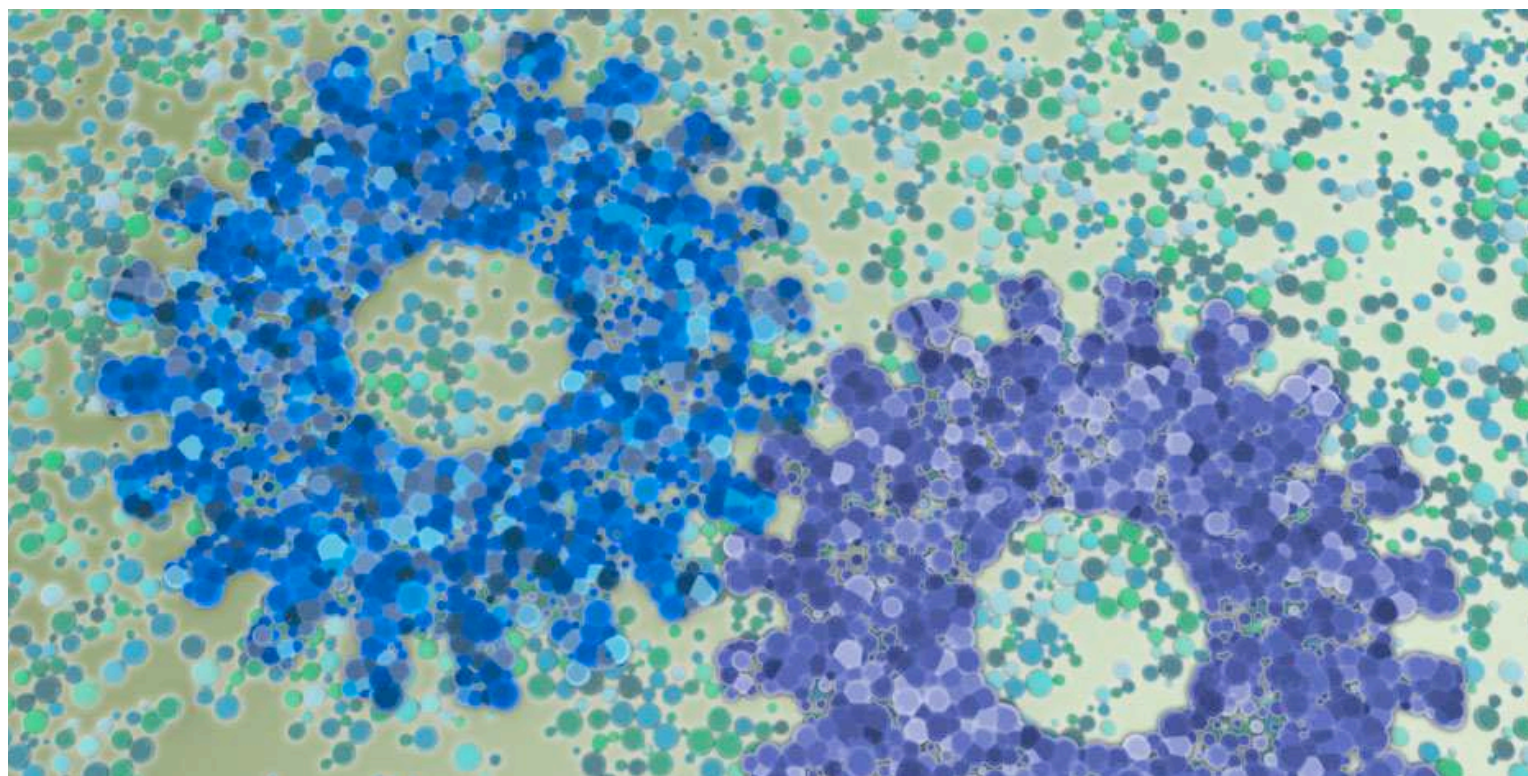
Sites web: <http://groups.csail.mit.edu/synbio/>, <http://contraintes.inria.fr/~batt/home.html>

- **Jean Peccoud** (Virginia Bioinformatics Institute, USA)

Associate Professor, Synthetic Biology Research Group

Contact: jpeccoud@vbi.vt.edu

Site web: <http://peccoud.vbi.vt.edu/>



RECHERCHE PRIVEE

Sociétés

• Alderys (Orsay)

AlderysSAS est une entreprise de chimie biologique spécialisée dans le design et le développement de **nouveaux procédés pour produire par voie biologique des molécules traditionnellement issues de la pétrochimie**. Alderys développe des nouveaux procédés de synthèse de composés destinés à l'alimentation animale, à la nutrition humaine, et au traitement de l'eau potable. Alderys a développé une plateforme intégrée de création et de sélection de micro-organismes dédiés à la production industrielle de composés chimiques par voie biologique. Alderys développe et commercialise ses nouveaux procédés de production auprès des industriels de la chimie.

- Créée en 2010 par Dominique Thomas et Philippe Marlière
- CEO & CSO: **D. Thomas**
- Revenus 2011 : 2.5M€
- Personnel : 28 personnes dont 7 PhD
- Localisation : Orsay (France)
- Site web: www.alderys.fr
- Contact: dt@alderys.fr



• Biométhodes (Evry)

Biométhodes développe un **procédé compétitif de bioraffinerie de 2^{ème} génération**, permettant la conversion complète et optimale de la biomasse non alimentaire lignocellulosique (bois, déchets végétaux agricoles, forestiers ou issus de l'agrochimie) en carburants et composés chimiques (sucres en C6, sucres en CR5 et lignine) utilisables par l'industrie en substitution du pétrole. Il intègre un prétraitement chimique de la lignocellulose, et l'action d'enzymes optimisées grâce à la plateforme de Biométhodes, OPTAZYME.

- Créée en 1997 par Marc Delcourt
- CEO: **G. Amsallem**, CSO: **N. Langvad**
- Personnel: 30 personnes dont 12 en France (dont 13 PhD)
- Localisation: Evry (France), 1 filiale OptaFuel US en Virginie (USA)
- Site web: www.biomethodes.com
- Contact: contact@biomethodes.com



• **Collectis** (Paris)

Collectis S.A. est le **pionnier mondial en ingénierie des génomes**. La société conçoit et commercialise des outils innovants, des nucléases telles que les méganucléases et les TALENs®, permettant d'intervenir de façon maîtrisée sur l'ADN, ainsi que d'autres produits résultant de cette technologie. Collectis a pour but à exploiter tout le potentiel de cette technologie en médecine, agriculture et dans d'autres secteurs de la recherche scientifique et industrielle.

- Créée en 1999 par André Choulika et David Sourdiv
- CEO: A. Choulika, CSO: P. Duchateau
- Revenus: 19M€ (dont 73% du CA à l'international)
- Cotée sur le marché Alternext de NYSE-Euronext à Paris depuis 2007
- Personnel: 220 personnes dont 74 PhD
- 4 filiales: Collectis biosearch,
Collectis therapeutics,
Collectis plant sciences
Collectis stem cells, composée d'Ectycell & Cellartis
- Localisation: Paris (France), Evry (France), Goteborg (Suède),
Saint-Paul (USA) et Cambridge (USA)
- Site web: www.collectis.com
- Contact: contact@collectis.com



• **Eukarÿs** (Beauvais)

La société Eukarÿs SAS a développé et breveté la technologie C3P3 qui permet l'expression ou l'inhibition virtuelle de tous gènes, dans toutes espèces eucaryotes et dans tous systèmes biologiques. Le premier marché visé est celui de la génomique fonctionnelle. Le second marché est celui des services de recherche, développement et production de protéines thérapeutiques pour l'industrie pharmaceutique.

- Créée en 2010 par B. Bertrand, P. Jais et F. Thomas
- CEO: B. Bertrand, CSO: P. Jais
- Localisation: Beauvais, Evry (France)
- Personnel: 3 personnes dont 2 PhD
- Site web: <http://eukarys.collectio.org>
- Contact: contact@eukarys.com



• **Eviagenics** (Villejuif)

Eviagenics est une société de biotechnologie qui applique sa **technologie propriétaire de recombinaison in vivo pour créer de nouveaux produits et des procédés de production améliorés et durables pour l'industrie pharmaceutique et chimique**.

Basée au sein de la pépinière Paris Santé Cochin, Eviagenics a mis au point une technologie innovante, la recombinaison in vivo, permettant d'insérer et d'améliorer des voies métaboliques complètes directement dans des cellules.

- Créée en 2004 par R. Pandjaitan
- CEO: R. Pandjaitan
- Localisation: Villejuif, (France)
- Site web: www.eviagenics.com
- Contact: contact@eviagenics.com



• Global Bioenergies (Evry)

Global Bioenergies développe un **procédé unique de production biologique d'isobutène** à partir de ressources renouvelables qui repose sur une voie métabolique artificielle constituée d'enzymes réalisant des réactions inédites.

La société se focalise maintenant sur l'industrialisation du procédé, c'est-à-dire sur l'amélioration de son rendement et sur sa mise à l'échelle industrielle, et s'attache également à la réplique de ce succès à la production biologique de certaines autres molécules centrales de la pétrochimie.

- Créée en 2008 par Marc Delcourt et Philippe Marlière,
- CEO: **M. Delcourt**, VP Metabolic Engineering: **C. Nakamura**, VP Chemical Engineering: **R. Bockrath**
- Revenus: 1,8M€
- Cotée sur le marché Alternext de NYSE-Euronext à Paris depuis 2011
- Personnel: 40 personnes dont 10 Phd
- Localisation: Evry (France), Iowa (USA)
- Site web: www.global-bioenergies.com
- Contact: info@global-bioenergies.com



• METabolic EXplorer (Saint-Beauzire)

METabolic EXplorer, entreprise de chimie biologique, développe et brevète des **procédés industriels fondés sur le principe de la fermentation**. Face à l'épuisement des ressources fossiles, ces procédés ouvrent la voie à l'utilisation des ressources naturelles renouvelables et pérennes. Economiquement performants et respectueux de l'environnement, ils offrent aux industriels une alternative sérieuse à la pétrochimie pour produire des composés chimiques entrant dans la composition de nombreux produits essentiels au quotidien.

- Créée en 1999 par Benjamin Gonzales
- CEO: **B. Gonzalez**, CSO: **P. Soucaille**
- Cotée sur le marché Alternext de NYSE-Euronext à Paris depuis 2007
- Revenus: 2,1M€
- Personnel: 80 personnes
- Localisation: Saint-Beauzire (France)
- Site web: www.metabolic-explorer.com
- Contact: contact@metabolic-explorer.com



• Michelin

Amyris, spécialiste des carburants et des produits chimiques issus de matières premières renouvelables, et Michelin, leader de l'innovation dans l'industrie pneumatique, collaborent en vue de développer et de commercialiser l'isoprène renouvelable No Compromise®, l'isoprène étant un composé chimique de base du caoutchouc synthétique ou naturel utilisé dans la fabrication des pneumatiques et d'autres produits.

Aux termes de cet accord, Michelin et Amyris contribueront ensemble au financement et aux ressources techniques permettant le développement de la technologie d'Amyris afin de produire de l'isoprène à partir de matières premières renouvelables.

Contact: amyris@schwartzcomm.com

• Sanofi-Aventis

Amyris, spécialiste des carburants et des produits chimiques issus de matières premières renouvelables, et le groupe pharmaceutique Sanofi-Aventis sont partenaires pour le développement d'une artémisinine d'hémi-synthèse, élément-clé dans les traitements antipaludiques de première ligne. Ce partenariat sera basé sur la technologie créée par le Professeur Jay Keasling à l'Université de Californie à Berkeley.

L'objectif de cette collaboration est de générer en complément de la fourniture d'artémisinine botanique, une source d'artémisinine additionnelle, de haute qualité, à un prix abordable, qui ne soit pas tributaire des conditions climatiques et permette à des millions de personnes atteintes de paludisme d'avoir un accès régulier à des associations à base d'artémisinine (ACTs), d'une importance vitale, et à un moindre coût.

Contact: francois.bompart@sanofi-aventis.com

• Total

Dans sa démarche d'engagement dans de projets de recherche et développement permettant une meilleure valorisation de la plante, Total a noué de nombreux partenariats internationaux avec des laboratoires universitaires ou privés travaillant dans le domaine des biotechnologies. Un de tels partenariats est formé avec une compagnie américaine Amyris, qui dispose d'une plate-forme de biologie synthétique de pointe. Celle-ci permet de construire et sélectionner des microorganismes aptes à transformer économiquement du sucre en diverses molécules d'intérêt. Total est aussi un des membres fondateurs de Toulouse White Biotech (France), plate-forme préindustrielle pour les biotechnologies.

Site web: <http://www.total.com/fr/nos-energies/energies-renouvelables/biomasse/>

Pôles de compétitivité

• Génopole (Evry)

Ce bioparc français dédié à la recherche en génomique, en génétique et aux biotechnologie regroupe plusieurs startups de Biologie de Synthèse, notamment Global Bioenergies, Eukarys et Biométhodes.

Site web: <http://www.genopole.fr/>

• Toulouse White Biotechnology (Toulouse)

TOULOUSE WHITE BIOTECHNOLOGY (TWB) est un démonstrateur préindustriel qui conçoit et construit des outils biologiques (enzymes, microorganismes, consortia microbiens...) ouvrant ainsi de nouvelles voies de production de molécules chimiques (synthons), de biopolymères, de biomatériaux, de biocarburants basées sur l'utilisation du carbone renouvelable.

Site web: <http://www.toulouse-white-biotechnology.com/>

Conseil

- **Biotics** (Paris)

La société BIOTICS a été créée en 1992, à l'initiative de Joël de Rosnay, dans le but de promouvoir le conseil stratégique en matière de nouvelles technologies, l'approche systémique appliquée à la prospective et à l'éducation, notamment dans les secteurs de l'Internet et des biotechnologies.

Site web: <http://www.biotics.fr/>

- **SynBioConsulting**, (Paris)

Société de conseil spécialisée en Biologie de Synthèse créée en 2013 par Camille Delebecque.

Site web: <http://www.synbioconsulting.com/>



EDUCATION

Les formations

- **Master 1 AIV**, Paris Descartes/Paris Diderot (Paris) <http://www.aiv-paris.org/en/master-aiv/>
- **Master 2 Biomédical “Systems and Synthetic Biology, Information and Interactions”** associé au Master 2 AIV, Paris Descartes (Paris) <http://www.aiv-paris.org/en/master-aiv/>
- **Master 2 “Master in Systems and Synthetic Biology”**, iSSB (Evry) <http://www.issb.genopole.fr/Education>
- **Master “Biotechnologie spécialisé Biologie Synthétique”**, ESBS et Réseau RBS-RhinSup (Strasbourg) http://www-esbs.u-strasbg.fr/esbs/?page_id=2838
- **Cours de Première Année «Biologie synthétique: une introduction»**, Mines ParisTech (Paris) <https://sgs.mines-paristech.fr/prod/sgs/ensmp/catalog/course/detail.php?code=S3212&lang=FR&year=1A>

Les équipes étudiantes iGEM

L'iGEM est une compétition étudiante internationale de Biologie de Synthèse organisée par le MIT (Cambridge, USA). Elle réunit depuis plusieurs années plus de 160 équipes d'étudiants venant des universités du monde entier, qui travaillent pendant 6 mois sur un projet de recherche et d'ingénierie pour répondre aux enjeux de demain.

La première équipe Française a été créée en 2007 au Centre de Recherche Interdisciplinaire de Paris, et fut rapidement rejointe par de nombreuses nouvelles équipes pour atteindre un total de 6 équipes françaises qui participent à l'édition iGEM 2013.

Au total, ces équipes ont remporté 10 médailles d'or, 1 médaille d'argent et 5 médailles de bronzes.

Récompenses:

- 2007
 - Prix de la recherche fondamentale (Paris Bettencourt)
- 2009
 - Prix des approches éthiques et sociétales (Paris Bettencourt)
- 2010
 - Prix de la recherche fondamentale (Paris Bettencourt)
- 2011
 - Prix du meilleur module génétique (INSA Lyon)
- 2012
 - Prix des approches éthiques et sociétales (Evry)
 - Prix du meilleur modèle (Evry)
 - Prix de la biosécurité (Grenoble)
 - Grand prix - 3ème au classement général (Paris Bettencourt)

- **Equipe iGEM Bordeaux (Université Bordeaux Segalen, Bordeaux)**
 - Participations: 2012, 2013
 - Contact: igem-bordeaux2-2012@googlegroups.com

- **Equipe iGEM Evry (ISSB, Evry)**
 - Participations: 2012, 2013
 - Contact: igem@issb.genopole.fr
 - Site web: <http://www.issb.genopole.fr/Education/igem-competition>

- **Equipe iGEM Grenoble (INP, Grenoble)**
 - Participations: 2011, 2013
 - Contact: igem.grenoble2012@gmail.com

- **Equipe iGEM INSA Lyon (INSA, Lyon)**
 - Participations: 2010, 2011, 2012
 - Contact: lyon.biosciences.igem@gmail.com

- **Equipe iGEM INSA Toulouse (INSA, Toulouse)**
 - Participation: 2013

- **Equipe iGEM Paris Bettencourt (Centre de Recherche Interdisciplinaire, Paris)**
 - Participations: 2007, 2008, 2009, 2010, 2011, 2012, 2013
 - Contact: contact@igem-paris.org
 - Site web: <http://www.igem-paris.org/>

- **Equipe iGEM Paris Saclay (Paris Sud Saclay, Saclay)**
 - Participations: 2012, 2013
 - Contact: igem.parissaclay@gmail.com
 - Site web: <http://www.igem-paris-saclay.u-psud.fr>

- **Equipe iGEM Paris Starsbourg (ENSPS, Strasbourg)**
 - Participations: 2008, 2009, 2010, 2011, 2012, 2013

- **Equipe iGEM SUPBiotech Paris (SUPBiotech, Paris)**
 - Participation: 2009



AUTRES ORGANISMES ET ASSOCIATIONS

- **Aviesan**, Paris

L'alliance nationale pour les sciences de la vie et de la santé a montré un intérêt à la Biologie de Synthèse, et organisé en 2012 un colloque « Perspectives en Biologie de Synthèse » qui a réuni les acteurs de ce domaine, ce qui pourrait donner lieu prochainement à l'organisation d'un Groupe de Recherche en Biologie de Synthèse.

Site web: <http://www.aviesan.fr>

Contact: corinne.brachet-ducos@aviesan.fr

- **Centre de Recherches Interdisciplinaires (CRI)**, Paris

Le Centre de Recherches Interdisciplinaires accueille et soutient le club de Biologie de Synthèse (SynBC), ouvert à tous, qui a pour but de rassembler les jeunes intéressés par la Biologie de Synthèse et de stimuler les discussions autour ce domaine. Ils offrent aussi une formation Master1 et Master2 en Biologie de Synthèse.

Site web : <https://groups.google.com/forum/?hl=fr.#!forum/synbc>

Contact: synbc@googlegroups.com

- **Institut de Biologie Systémique et Synthétique**, Paris

Cet institut français dédié à la recherche en Biologie des Systèmes et Biologie Synthétique accueille 5 équipes de recherche qui se concentrent sur ces thématiques. Ils proposent un Master2 en Biologie des Systèmes et Biologie Synthétique.

Site web: <http://www.issb.genopole.fr>

Contact: Jean-Loup.Faulon@issb.genopole.fr

- **La Paillasse**, Vitry sur Seine

La Paillasse est un laboratoire communautaire pour les biotechnologies citoyennes, ouvert à tous, et qui entre autres développe des projets de Biologie de Synthèse

Site web: <http://www.lapaillasse.org>

Contact: thomas.landrain@lapaillasse.org

- **Le Laboratoire**, Paris

Le Laboratoire est un lieu d'art et de design aux frontières de la science faisant partie du réseau international ArtScience Labs. En 2012 Le Laboratoire a accueilli des équipes de jeunes pour la compétition artscience sur le sujet de la Biologie de Synthèse.

Sites web: <http://www.laboratoire.org/> , <http://www.artscienceprize.org/paris>

Contact: olivier.borgeaud@laboratoire.org

- **Observatoire de la biologie de synthèse**, Paris

L'objectif de l'Observatoire est de réaliser une cartographie des activités, des discussions et des réflexions menés dans ce domaine, ainsi qu'assurer un dialogue avec la société, principalement vis-à-vis le site internet.

Site web: <http://biologie-synthese.cnam.fr/>

Contact: ali.saib@cnam.fr

- **SynBioFrance**, Paris

SynBioFrance est la première association française de promotion de la Biologie de Synthèse, qui vise à regrouper étudiants, chercheurs et entrepreneurs de cette discipline, autour d'événements d'envergure nationale.

Sites web: <http://www.synbiofrance.org>

Contact: contact@synbiofrance.org

ANNEXES

Publications des équipes de recherche françaises (liste non exhaustive)

2010

Bikard D, Julié-Galau S, Cambray G, Mazel D. The synthetic integron: an in vivo genetic shuffling device. *Nucleic Acids Res.* 2010 Aug;38(15):e153. doi: 10.1093/nar/gkq511. Epub 2010 Jun 9.

Delaplace F, Kludel H, Cartier-Michaud A. Discrete Causal Model View of Biological Networks CMSB 2010 Tech. Rep.

2011

Carbonell P, Planson AG, Fichera D, Faulon JL. A retrosynthetic biology approach to metabolic pathway design for therapeutic production. *BMC Syst Biol.* 2011 Aug 5;5:122. doi: 10.1186/1752-0509-5-122.

Delebecque CJ, Lindner AB, Silver PA, Aldaye FA. Organization of intracellular reactions with rationally designed RNA assemblies. *Science.* 2011 Jul 22;333(6041):470-4. doi: 10.1126/science.1206938. Epub 2011 Jun 23.

Jaramillo A, Faulon JL. Editorial: Synthetic Biology - applying new paradigms at the interface of fundamental research and innovation. *Biotechnol J.* 2011 Jul;6(7):766-7. doi: 10.1002/biot.201100254.

Laisne A, Ewald M, Ando T, Lesniewska E, Pompon D. Self-assembly properties and dynamics of synthetic proteo-nucleic building blocks in solution and on surfaces. *Bioconjug Chem.* 2011 Sep 21;22(9):1824-34. doi: 10.1021/bc2002264. Epub 2011 Sep 6.

Randall A, Guye P, Gupta S, Duportet X, Weiss R. Design and connection of robust genetic circuits. *Methods Enzymol.* 2011;497:159-86. doi: 10.1016/B978-0-12-385075-1.00007-X.

2012

Basso-Blandin A, Delaplace F. GUBS, a Behavior-based Language for Open System Dedicated to Synthetic Biology. *Mecbic 2012 - CONCUR*

Batt G, Besson B, Ciron PE, de Jong H, Dumas E, Geiselman J, Monte R, Monteiro PT, Page M, Rechenmann F, Ropers D. Genetic network analyzer: a tool for the qualitative modeling and simulation of bacterial regulatory networks. *Methods Mol Biol.* 2012;804:439-62. doi: 10.1007/978-1-61779-361-5_22.

Delebecque CJ, Silver PA, Lindner AB. Designing and using RNA scaffolds to assemble proteins in vivo. *Nat Protoc.* 2012 Oct;7(10):1797-807. doi: 10.1038/nprot.2012.102. Epub 2012 Sep 6.

Planson AG, Carbonell P, Grigoras I, Faulon JL. A retrosynthetic biology approach to therapeutics: from conception to delivery. *Curr Opin Biotechnol.* 2012 Dec;23(6):948-56. doi: 10.1016/j.copbio.2012.03.009. Epub 2012 Apr 2.

Rodrigo G, Jaramillo A. AutoBioCAD: Full Biodesign Automation of Genetic Circuits. *ACS Synth Biol.* 2012 Nov 26. [Epub ahead of print]

Rodrigo G, Landrain TE, Jaramillo A. De novo automated design of small RNA circuits for engineering synthetic riboregulation in living cells. *Proc Natl Acad Sci U S A.* 2012 Sep 18;109(38):15271-6. Epub 2012 Sep 4.

Singh V, Chaudhary DK, Mani I. Gene network analysis of *Aeromonas hydrophila* for novel drug target discovery. *Syst Synth Biol.* 2012 Jun;6(1-2):23-30. doi: 10.1007/s11693-012-9093-z. Epub 2012 May 22.

Uhlendorf J, Miermont A, Delaveau T, Charvin G, Fages F, Bottani S, Batt G, Hersen P. Long-term model predictive control of gene expression at the population and single-cell levels. *Proc Natl Acad Sci U S A*. 2012 Aug 28;109(35):14271-6. doi: 10.1073/pnas.1206810109. Epub 2012 Aug 14.

Val ME, Skovgaard O, Ducos-Galand M, Bland MJ, Mazel D. Genome engineering in *Vibrio cholerae*: a feasible approach to address biological issues. *PLoS Genet*. 2012 Jan;8(1):e1002472. doi: 10.1371/journal.pgen.1002472. Epub 2012 Jan 12.

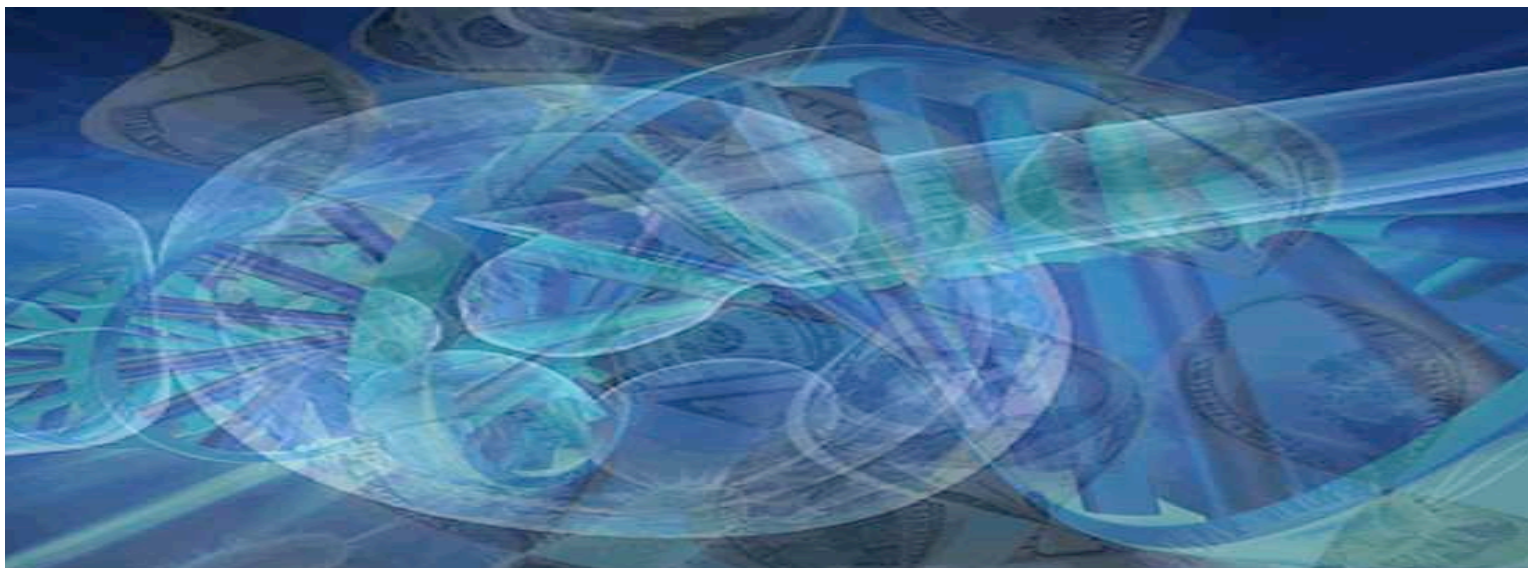
2013

Belič A, Pompon D, Monostory K, Kelly D, Kelly S, Rozman D. An algorithm for rapid computational construction of metabolic networks: a cholesterol biosynthesis example. *Comput Biol Med*. 2013 Jun 1;43(5):471-80. doi: 10.1016/j.combiomed.2013.02.017. Epub 2013 Mar 16.

Carbonell P, Planson AG, Faulon JL. Retrosynthetic design of heterologous pathways. *Methods Mol Biol*. 2013;985:149-73. doi: 10.1007/978-1-62703-299-5_9.

Carte des acteurs de la Biologie de Synthèse en France

<http://goo.gl/agaHg>



ECONOMIE & SOCIÉTÉ

RECHERCHE & INNOVATION INNOVATION

1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 | 13 | 14 | 15 | 16 | 17 | 18 | 19 | 20 | 21 | 22 | 23 | 24 | 25 | 26 | 27 | 28 | 29 | 30 | 31 | 32 | 33 | 34 | 35 | 36 | 37 | 38 | 39 | 40 | 41 | 42 | 43 | 44 | 45 | 46 | 47 | 48 | 49 | 50 | 51 | 52 | 53 | 54 | 55 | 56 | 57 | 58 | 59 | 60 | 61 | 62 | 63 | 64 | 65 | 66 | 67 | 68 | 69 | 70 | 71 | 72 | 73 | 74 | 75 | 76 | 77 | 78 | 79 | 80 | 81 | 82 | 83 | 84 | 85 | 86 | 87 | 88 | 89 | 90 | 91 | 92 | 93 | 94 | 95 | 96 | 97 | 98 | 99 | 100

11/03/2013 | [Gilles Bouchard](#) | [Xavier Dupret](#) | [Innovation](#) | [Le Cercle les Echos](#) | [6 commentaires](#)

Comment développer science, innovation et entrepreneuriat en France

LE CERCLE. Pourquoi se crée-t-il en moyenne trois fois moins de startups en France qu'aux Etats-Unis ?

Tout simplement parce que nous n'avons pas encore su réconcilier et articuler pleinement l'innovation avec la créativité d'un côté, la recherche et l'entrepreneuriat de l'autre.

Innovation et créativité

Le discours ambiant semble parfois confondre innovation et créativité. Question de culture et de système probablement.

Et pourtant, l'innovation n'est pas une invention. L'innovation est très souvent mal identifiée et difficilement mesurable. Elle est toujours à la fois un processus d'émergence complexe et un résultat. Elle n'existe qu'à partir du moment où le marché la valide. Elle ne produit de la valeur que pour les entreprises qui prennent des risques.

La créativité est quelque part plus simple, plus concrète et plus facilement palpable. Elle conduit directement à l'invention. La susciter et l'apprendre répond à un processus plus simple et plus rapide. La créativité ne doit pas son existence à une validation par le marché. Elle consiste tout simplement à ouvrir les cerveaux en amont de l'apprentissage des connaissances. A leur faire percevoir la transversalité existant entre des disciplines et des manières de penser différentes. Le risque en est une dimension intrinsèque et naturelle.

En privilégiant les connaissances au détriment de la créativité, nos systèmes stérilisent l'inventivité faite de sensibilité, d'observation, d'expériences répétées, bridant ainsi ces catalyseurs indispensables de l'innovation. En privilégiant le savoir au détriment de l'inventivité, ils rendent difficile la création naturelle de passerelles entre les métiers (recherche - industrie - finance) et les structures (laboratoires - startups - grandes entreprises).

Ceci est d'autant plus handicapant que les rapports entre science, innovation et entrepreneuriat ne sont pas forcément de type amont-aval. Les innovations peuvent être stimulées par la recherche scientifique et les technologies qui en découlent tout autant que par les entreprises qui les développent. Elles peuvent également être structurées par celles-ci, le dénominateur commun restant la créativité qui demeure le parent pauvre de notre système d'éducation, à quelque niveau que ce soit.

Recherche et entrepreneuriat

Comme le souligne très justement le SIE Network [sie-network.com](#), « trop de jeunes scientifiques et d'ingénieurs décident de mettre leurs talents au service de grandes entreprises, dans des domaines parfois très éloignés de leurs passions et de leurs expertises. Même si ils y excellent souvent, ils ne créent pas une valeur ajoutée aussi passionnante et durable que celle issue de leurs propres recherches. Les carrières scientifiques ne sont aujourd'hui pas assez valorisées en France. Le monde de demain est pourtant rempli de challenges techniques et scientifiques auxquels il faudra apporter des solutions. Et les réponses à ces questions fondamentales résident dans l'innovation technologique ».

L'Institut Montaigne dresse le même constat lorsqu'il souligne « qu'un trop

ÉCRIT PAR

Gilles Bouchard
 CEO Executive
 Business Accelerator
 & Up4Startup
 VP Harvard Angels
 France



VOIR SON
 PROFIL >

SES 2 DERNIERS ARTICLES

16/03/2013 | 07:30

Portrait de "Tentpreneur transformateur"

21/03/2013 | 08:00

Des clés pour une stratégie réussie

11/03/2013 | 13:00

Pour un nouveau modèle d'accélérateur ?

TOUS SES ARTICLES >

Xavier Dupret
 Doctorant
 INRIA & MIT



VOIR SON
 PROFIL >

SUR LE CERCLE



11/03/2013 | 09:00

Innovation et dynamique de croissance

[la suite](#)



07/03/2013 | 12:00

8 priorités pour dynamiser l'innovation en France

[la suite](#)

SUR INTERNET

PROSPEKTIV

Harvard Angels France

PUBLIEZ VOS ARTICLES

LES ÉCONOMISTES DE PROJECT SYNDICATE

KENNETH ROGOFF

Professeur d'économie et de politique publique à l'Université d'Harvard. Il a été économiste en chef au Fonds Monétaire International.

[L'Esprit de what's Hot](#)

LE CERCLE LES ECHOS SUR TWITTER

grand nombre d'ingénieurs formés et diplômés en France s'éloignent non seulement du métier d'ingénieur stricto sensu mais également des métiers de l'industrie, en leur préférant notamment la finance ».

Pourquoi tant de nos étudiants brillants vont-ils compléter leur parcours aux Etats-Unis ? Tout simplement pour aller y chercher une véritable pédagogie de la créativité et de l'innovation qui manque à l'enseignement souvent abstrait qu'ils ont reçu en France.

Pourquoi restent-ils ensuite aux Etats-Unis ? Tout simplement parce que cette pédagogie de l'expérience, associée à un niveau scientifique souvent bien supérieur à leurs homologues étrangers à même niveau d'étude, en fait des profils très recherchés. Des ingénieurs à même d'accompagner efficacement les transferts de technologies de la paillasse de laboratoire à l'application industrialisable et commercialisable. Des Doctorants qui ont l'opportunité d'alterner recherche, enseignement et création de startup dans un cursus où la sensibilisation et l'accompagnement à l'entrepreneuriat sont très forts. Comme le dit avec beaucoup de justesse Stéphane Mallat, Professeur à Polytechnique : « La recherche, aussi bizarre que cela puisse paraître, est une excellente formation pour la création de startup. Les premières phases sont très similaires : recherche de financement, levée de fonds... ». Mais surtout, « créer une startup technologique, c'est décider qu'on veut être le meilleur dans le monde dans son domaine grâce à une innovation de rupture ». Alors, que faire ?

Des solutions simples

Renverser notre culture de l'entrepreneuriat. Il faut « reformater » nos mentalités et « dédramatiser » la recherche publique à visée applicative. Ceci afin de permettre une revalorisation du transfert de technologie issue des organismes publics vers les startups. Seulement alors le profil de chercheur-entrepreneur émergera et deviendra partie intégrante de la recherche. François Nemo a parfaitement analysé la question lorsqu'il rappelle que notre culture économique, fondée sur les Lumières et l'influence Saint-simonienne, s'appuie sur le savoir plutôt que sur l'expérience. « C'est la France des grandes inventions, des grands projets, des grandes écoles, des grandes administrations qui génèrent des esprits hautement rationnels, cartésiens, méthodiques, d'excellents ingénieurs, gestionnaires ou administrateurs. Des dirigeants qui s'appuient sur des modèles de hiérarchie verticaux et cloisonnés qui ont fait leurs preuves pendant de nombreuses générations mais qui aujourd'hui trouvent leur limite dans ce nouveau monde ouvert, multiculturel, modelé par les réseaux et les nouvelles technologies. Une culture élitiste et sélective profondément ancrée dans les institutions politiques et économiques et qui bloque l'émergence des idées créatives souvent portées par les entrepreneurs ».

Rééquilibrer savoir et expérience dans nos cursus d'études, et ce à tous les niveaux. C'est en ce sens qu'il faut intégrer la proposition de l'Institut Montaigne de créer « des centres scientifiques expérimentaux », avec du matériel de laboratoire en libre accès pour les élèves, qui pourraient ainsi travailler sur le tas, à l'instar de leurs homologues des grandes universités américaines telles que le MIT, Harvard ou encore Stanford, qui doivent valider leur Master par une *Master's thesis* préparée au sein des laboratoires, véritables pépinières de futures startups où se côtoient chercheurs, étudiants et entreprises. Il faut ainsi promouvoir les initiatives très récentes telles que le FacLab de l'Université de Cergy Pontoise ou encore La Paillasse, laboratoire issu du mouvement français « *Do It Yourself* » centré sur la biologie et ouvert au public.

Inverser culture technologique et culture de l'idée. C'est souvent par le biais des moyens, de l'organisation et du financement que l'on traite de l'innovation en France, alors que « la véritable problématique ne réside ni dans le financement, ni dans la technologie ». Elle réside dans la créativité du business model, dans la proposition de valeur qui fonde le projet, dans l'investissement immatériel pour lequel il existe une véritable excellence française.

Il s'agit bien là de renforcer la créativité par rapport à l'innovation en évitant de confondre l'une avec l'autre.

Ces quelques propositions n'excluent pas, bien entendu, la nécessité de structurer la filière de la recherche, de renforcer la collaboration entre grandes entreprises et startups innovantes tout en consolidant la chaîne de financement qui doit les lier.

N'oublions qu'il y a un grand gisement de jeunes scientifiques et ingénieurs très compétents pour lesquels l'entrepreneuriat est désormais partie intégrante de leur culture. Favorisons les conditions de leur épanouissement. Qu'ils puissent être à l'origine d'un AMGEN, cette société de biotech américaine fondée en 1980 par trois chercheurs, qui compte aujourd'hui près de 18.000 collaborateurs et génère plus de 14 milliards de dollars de chiffre d'affaires. Qu'ils puissent marcher sur les pas de Roland Moreno, l'inventeur de la carte à puce, mais avec une dimension de *business development* supplémentaire. Celle que les Pouvoirs Publics auraient dû lui apporter pour créer un Apple d'une autre envergure avec plus de 7 milliards de cartes actuellement en circulation. Rappelons au passage qu'il lui aura fallu près de huit ans pour convaincre banquiers et industriels de l'utilité de son invention...

Gilles Bouchard - Harvard Angels
Xavier Duportet - PhD student - Constraints group (INRIA) - Weiss Lab for Synthetic Biology (MIT)

À LIRE ÉGALEMENT

	Gérard Petit Fukushima : tout dire et bien le dire, ne peut pas nuire		Philippe du Fresnay Jeu de Go en mer de Chine : des îlots à fleur de peau
	Damien Voces Olego L'innovation au service des salariés		Paul Derreumaux Afrique Francophone : le Maroc à l'offensive

[Suivre](#)

Le Cercle Les Echos @CercleLesEchos 5h

Pour conquérir, pourquoi faut-il se débarrasser de certains mots vulgaires et préjudiciables ? bit.ly/1jpz17Y

Étendre

Le Cercle Les Echos @CercleLesEchos 5h

Fukushima : tout dire et bien le dire, ne peut pas nuire bit.ly/1jpz17R

Étendre

Le Cercle Les Echos @CercleLesEchos 6h

Jeu de Go en mer de Chine : des îlots à fleur de peau bit.ly/1dCJSJh

Étendre

Le Cercle Les Echos @CercleLesEchos 7h

L'innovation au service des salariés bit.ly/1gh4Q3

Étendre

Le Cercle Les Echos @CercleLesEchos 21 Mars

Burn-out : au-delà de l'individu, un enjeu collectif bit.ly/1dxwh67

Étendre

Tweeter à @CercleLesEchos

AUJOURD'HUI, LES ARTICLES LES PLUS...

...LUS	...COMMENTÉS
Paul Derreumaux Afrique Afrique Francophone : le Maroc à l'offensive	
Mathieu Gabai RH Tribune Les MOOC, nouveaux outils de recrutement et de management pour les entreprises ?	
Francis Boyer RH Tribune Les 4 syndromes qui empêchent les managers de manager	
Thierry PICARD Tendances-innovation De la business intelligence à la business analytics	
Jean-Luc Soulé Europe Tribune Ukraine 1914-2014 : le conflit ukrainien et les leçons de la Grande Guerre	

ABONNEZ-VOUS AU FLUX RSS
SUIVEZ-NOUS AVEC TWITTER

Le Cercle Les Echos
Like

8,081 people like Le Cercle Les Echos.



Facebook social plugin