Multigene Delivery to Mammalian Cells and its Applications in Cell Biology

Inauguraldissertation

zur

Erlangung der Würde eines Doktors der Philosophie

vorgelegt der

Philosophisch-Naturwissenschaftlichen Fakultät

der Universität Basel

Von

Maysam Mansouri

aus Tehran, Iran

Basel, 2016

Genehmigt von der Philosophisch-Naturwissenschaftlichen Fakultät

auf Antrag von:

Prof. Dr. Kurt Ballmer-Hofer

Prof. Dr. Martin Spiess

Basel, 21 Juni 2016

Prof. Dr. Jörg Schibler, Dekan

This thesis is dedicated to my parents and Zahra

With love,
Maysam

Doctoral Thesis

Multigene Delivery to Mammalian Cells and its Applications in Cell Biology

A dissertation submitted to the

UNIVERSITÄT BASEL

Philosophisch Naturwissenschaftliche Fakultät

For the degree of
Doctor of Philosophy

In the subject of
Genetics

Presented by

Maysam Mansouri

Born on 23.03.1984
Citizen of Iran

Accepted on the recommendation of

Prof. Dr. Kurt Ballmer-Hofer
Prof. Dr. Martin Spiess
Prof. Dr. Verdon Taylor
Dr. Philipp Berger

Paul Scherrer Institute
Villigen, July 2016

Contents

Summary **1**

Zusammenfassung **3**

Résumé **5**

Aims of the study **7**

1 Introduction **9**

1.1 Strategies for multigene expression in eukaryotic cells 10

1.2 Multigene delivery in mammalian cells: advances and recent applications ... 16

2 MultiPrime unlocks multigene delivery to primary mammalian cells **47**

3 Simple protocol for efficient MultiPrime virus generation **75**

4 Software for image analysis **92**

5 Conclusion **107**

6 Acknowledgements **109**

7 Curriculum vitae **111**

SUMMARY

Multigene delivery systems are emerging as key technologies required in diverse research fields in both academia and industry. Examples include genome editing and synthetic biology, cellular reprogramming and stem cell biology, production of complex proteins in structural biology and many others. Many strategies for assembly and delivery of multigene constructs exist but all of them have their limitations. Genetic constructs are typically introduced into mammalian cells by transient transfection. However, the low efficiency of transient transfection and its limited applicability to a variety of important mammalian cells significantly restrict the utility of this approach. Viral systems infect many cell types but usually don't have enough capacity for multiple genes. Therefore, highly efficient delivery of multigene plasmid in a variety of mammalian cells is a challenge which needs to be addressed. We developed MultiPrime, a baculovirus-mediated multigene expression system, allowing simultaneous expression of several genes from a single virus for mammalian cells. MultiPrime is a modular, non-cytotoxic, non-integrating system and works with a variety of promoters. MultiPrime viruses efficiently transduced a wide range of cell types, including non-dividing primary neurons and induced-pluripotent stem cells (iPS). We showed that MultiPrime can be used for reprogramming, and for genome editing and engineering by CRISPR/Cas9. Moreover, we implemented dual-host-specific cassettes enabling multiprotein expression like full length human antibodies in insect and mammalian cells using a single reagent. Also, MultiPrime-infected Zebrafish embryos showed expression of all expected genes in vivo. Our experiments establish

MultiPrime as a powerful and highly efficient tool, to deliver multiple genes for a wide range of applications in primary and established mammalian cells.

The thesis is organized in five chapters. The first chapter provides an introduction to available and common strategies for multigene expression in mammalian cells as well as some of the most important applications of multigene delivery in biological research. The chapters two, three, and four present original research reports covering our established system, an optimized protocol necessary for production and utilization of the system and development of some software for analysis of acquired images by our system. In chapter two, Mansouri et al. (2016) present the main project including establishment of MultiPrime and its diverse applications in cell biology. In chapter three, Mansouri et al. (manuscript in preparation) describe an optimized protocol to generate MultiPrime baculovirus particles that harbor complete gene constructs. In chapter four, Rizk et al. (2014 and 2015) present Squassh, Squassh3C and SquasshAnalyst, three user-friendly software modules that enable segmentation and quantification of subcellular structures of fluorescence microscopy images. In the last Chapter, I discuss and summarize our results and refer to the limitations of our system and I propose possible future directions for genetic manipulation of cells.

ZUSAMMENFASSUNG

Systeme zum Einbringen von mehreren Genen in eukaryotische Zellen spielen eine wichtige Rolle in der industriellen und akademischen Forschung. Sie werden z.B benötigt, um das Genom zu editieren, um den Zelltyp zu verändern, um Proteinkomplexe zu produzieren oder in der synthetischen Biologie. Einige Systeme existieren bereits, wobei diese Nachteile besitzen. Konstrukte mit genetischer Information werden üblicherweise mittels Transfektion in die Zellen gebracht. Die Effizienz ist allerding schlecht und diese Strategie ist auf wenige Zelltypen beschränkt. Virale Systeme können zwar viele Zelltypen infizieren, ihre Kapazität für fremde DNAS ist jedoch beschränkt. Deshalb ist die Entwicklung eines Systems mit hoher Kapazität für fremde DNS, das das in vielen Zelltypen aktiv ist, wünschenswert. Wir haben „MultiPrime“ entwickelt. Dieses System basiert auf einem baculoviralen Expressionsystem und erlaubt die gleichzeitige Expression von mehreren Genen von einem einzigen Virus. MultiPrime ist modular aufgebaut, nicht cytotoxisch, integriert nicht ins Genom der Zielzelle uns arbeitet mit verschiedenen Promotoren. MultiPrime Viren infizieren viele Zelltypen wie z.B. sich nicht teilende primäre Nervenzellen oder induzierte pluripotente Stamm Zellen. Wir zeigten auch, dass sich MultiPrime zur Umprogrammierung von Zellen eignet und dass man damit das Genome mittels CRISPR/Cas9 editieren kann. Zusätzlich verwendeten wir Promotoren, die in zwei verschieden Wirtszellen aktiv sind. Damit konnten wir Proteine in Insekten- und Säugerzellen exprimieren. Wir infizierten Zebrafisch Embryos um zu zeigen, dass unser Virus auch in vivo verwendet werden kann. Diese Experimente zeigen, dass unser MultiPrime System ein wertvolles Werkzeug für zellbiogische Experimente ist.

Diese Doktorarbeit enthält fünf Kapitel. Das erste Kapitel gibt eine Einführung in bekannte Strategien zur Expression von mehreren Genen und deren Anwendung. Die Kapitel zwei bis vier enthalten Publikationen meiner Arbeit. Kapitel zwei beschreibt mein Hauptprojekt, die Entwicklung und Validierung von MultiPrime (Mansouri et al., 2016). Kapitel drei beschreibt ein optimiertes Protokoll zur Herstellung von Viren (Mansouri et al., in preparation). Kapitel vier enthält zwei Publikationen an denen ich mitgearbeitet habe (Rizk et al., 2014, 2015). Diese beiden Publoation beschreiben Squassh, Squassh3C und SquasshAnalyst, drei Programme zur Segmentierung und Quantifizierungvon subzellulären Strukturen in Mikroskopie Bildern. Im letzten Kapitel fasse ich meine Ergebnisse zusammen und diskutiere mögliche Einschränkungen und schlage Experiment für die Zukunft vor.

RÉSUMÉ

Le transfert et l'expression de plusieurs gènes dans des cellules eucaryotes sont des techniques cruciales dans de nombreux domaines de la recherche académique et industrielle. Le transfert de plusieurs gènes est par exemple utilisé dans les domaines de la biologie synthétique, de la reprogrammation cellulaire, pour l'édition du génome ou pour la production de complexes protéiques pour la biologie structurale. Il existe actuellement de nombreuses techniques pour la création et le transfert de constructions d'ADN contenant plusieurs gènes, mais toutes ont des limitations. Les constructions génétiques sont habituellement introduites dans les cellules de mammifères par transfection transitoire. Cependant, ces techniques ont un rendement faible et ne peuvent être utilisés que pour certains types de cellules. Les systèmes de transfert basés sur des virus sont capables d'infecter de multiples types cellulaires mais n'ont habituellement pas une capacité d'emport suffisante pour contenir plusieurs gènes. Par conséquent, un système de transfert à haute capacité utilisable dans de nombreux types de cellules est très recherché. Nous avons pour cette raison développé MultiPrime, un système basé sur les baculovirus qui permet le transfert et l'expression simultanée de plusieurs gènes dans des cellules de mammifères en utilisant un unique virus. MultiPrime est un système modulaire, non toxique, qui ne s'intègre pas dans le génome et qui fonctionne avec une grande variété de promoteurs. Nous avons montré que les virus MultiPrime permettent une transduction efficace d'une grande variété de types de cellules, dont les cellules souches pluripotentes induites ou les cellules qui ne se divisent pas comme les neurones. Nous avons également montré que MultiPrime peut être utilisé pour la reprogrammation cellulaire ainsi que pour

l'édition du génome par CRISPR/Cas9. De plus, nous avons implémenté des cassettes d'expression actives dans plusieurs types cellulaires permettant l'expression de complexes protéiques dans des cellules de mammifères ou d'insectes avec le même virus. Enfin, nous avons montré en infectant des embryons de poisson-zèbre que MultiPrime peut être utilisé in vivo et que tous les gènes transférés sont alors exprimés. Ces expériences établissent que MultiPrime est un outil efficace de transfert de gènes pour de nombreuses applications de biologie cellulaire.

Cette thèse est organisée en cinq chapitres. Le premier chapitre présente les techniques usuelles pour le transfert et l'expression de plusieurs gènes dans les cellules de mammifères ainsi que les applications les plus importantes en biologie. Les chapitres deux, trois et quatre sont des publications décrivant notre système et ses applications. Le chapitre deux constitue la partie centrale de ce travail et présente le développement, la validation et les applications du système MultiPrime (Mansouri et al. 2016). Le chapitre trois décrit un protocole pour la production optimisée de particules baculovirales MultiPrime (Mansouri et al., en préparation). Le chapitre quatre présente deux publications résultant d'une collaboration utilisant MultiPrime pour créer et analyser par microscopie des cellules contenant plusieurs marqueurs fluorescents (Rizk et al. 2014, 2015). Plus précisément ces publications décrivent Squassh et SquasshAnalyst, deux programmes pour la segmentation et la quantification de structures cellulaires dans des images de microscopie par fluorescence. Le dernier chapitre est une discussion sur les résultats et perspectives futures du système MultiPrime.

AIMS OF THE STUDY

The introduction of heterologous genetic information, e.g. multiple genes, into mammalian cells is a key technology in current cell biology. These genetic components are typically introduced into mammalian cells by transient transfection or viral transduction. However, the low efficiency of transient transfection and its limited applicability to a variety of important mammalian cells significantly restrict the utility of this approach. On the other hand, most of the common used viruses for gene delivery suffer from low cargo capacity or inability to infect dividing and non-dividing mammalian cells. Therefore, development of an efficient baculoviral system, termed MultiPrime, for the delivery of multiple gene cassettes to mammalian cells, both dividing established lines and non-dividing primary cells, would significantly facilitate genetic manipulation of cells for multiple cell biological applications.

My aims in this project are to;

1. Establish a baculovirus-based multigene delivery system for mammalian cells
2. Evaluate the capability of MultiPrime viruses, harboring multigene cassettes to transduce a variety of mammalian cells such as established cell lines, primary cells and stem cells
3. Assess gene delivery efficacy and expression duration in MultiPrime transduced cells
4. Develop an optimized protocol for efficient generation MultiPrime virus stocks
5. Optimize the protocol to achieve high transduction rates in cell lines and primary cells
6. Generate VSV-G protein pseudotyped MultiPrime baculovirus vectors and compare them to wildtype virus
7. Create multicomponent biosensors for quantitative monitoring of specific cellular parameters through snapshot and live-cell imaging

8. Modulate the gene expression level with alternative promoters in MultiPrime-infected mammalian cells
9. Express multiple proteins in mammalian and insect cells by a single MultiPrime virus
10. Apply MultiPrime system for CRISPR/Cas9-mediated genome editing, reprogramming and antibody production
11. Deliver multiple genes to a living organism like zebrafish

CHAPTER 1

INTRODUCTION

This chapter aims to summarize available strategies and also recent applications of multigene delivery systems for mammalian cells. The chapter contains two review articles.

The first review is focused on the most common strategies for assembly of multigene constructs. I describe co-transfection/co-infection strategies, Internal ribosomal entry sites (IRES) systems, polyproteins systems, direct assembly of multiple expression cassettes (e.g. Gibson assembly, BioBrick and etc), and Cre recombination-based systems such as MultiLabel. Advantages and drawbacks of each system are provided (Mansouri et al., 2014).

The second review, which is in preparation, summarizes recent applications of multigene delivery systems not only for cell biology purposes using RNA polymerase II promoters but also for U6-driven expression cassettes and homology constructs which are required for genome editing though CRISPR/Cas9 (Mansouri et al., in preparation).

Plasmid 75 (2014) 12–17

Contents lists available at ScienceDirect

Plasmid

journal homepage: www.elsevier.com/locate/yplas

Review

Strategies for multigene expression in eukaryotic cells

Maysam Mansouri, Philipp Berger *

Paul Scherrer Institute, Biomolecular Research, Molecular Cell Biology, CH-5232 Villigen PSI, Switzerland

ARTICLE INFO

Article history:
Received 26 May 2014
Accepted 7 July 2014
Available online 15 July 2014
Communicated by Saleem Khan

Keywords:
Multiprotein expression systems
Cell biology
Structural biology

ABSTRACT

Multigene delivery systems for heterologous multiprotein expression in mammalian cells are a key technology in contemporary biological research. Multiprotein expression is essential for a variety of applications, including multiparameter analysis of living cells *in vitro*, changing the fate of stem cells, or production of multiprotein complexes for structural biology. Depending on the application, these expression systems have to fulfill different requirements. For some applications, homogenous expression in all cells with defined stoichiometry is necessary, whereas other applications need long term expression or require that the proteins are not modified at the N- and C-terminus. Here we summarize available multiprotein expression systems and discuss their advantages and disadvantages.

Contents

1. Introduction . . . 12
2. Cotransfection and coinfection . . . 14
3. Internal ribosomal entry sites (IRES). . . . 14
4. Direct assembly of multiple expression cassettes . . . 15
5. Release from polyproteins . . . 15
6. Cre recombination . . . 16
7. Concluding remarks . . . 16
Acknowledgments . . . 16
References . . . 16

1. Introduction

Multigene expression systems are central tools for many applications in biology. Therefore many new strategies for multigene expression have been established in recent years (Fig. 1). Multigene expression systems were initially developed for production of protein complexes needed for structural and biochemical analysis. Bacterial and insect cell systems were mainly used for this purpose because they are fast, cost-efficient and easy to handle (Vijayachandran et al., 2011). Mammalian systems are of course also suitable for protein production, but they are mainly used to study or to manipulate physiological processes within cells (Geisse and Kocher, 1999; Perrakis and Romier, 2008). Therefore, additional requirements have to be fulfilled by these systems.

The expression of multiple fluorescently-tagged sensors is a useful strategy to monitor several parameters simultaneously in living cells. For example, we use our MultiLabel

* Corresponding author. Address: Paul Scherrer Institute, Molecular Cell Biology, OFLC 101, CH-5232 Villigen, Switzerland. Fax: +41 56 3105288.
E-mail address: Philipp.Berger@psi.ch (P. Berger).

http://dx.doi.org/10.1016/j.plasmid.2014.07.001
0147-619X/

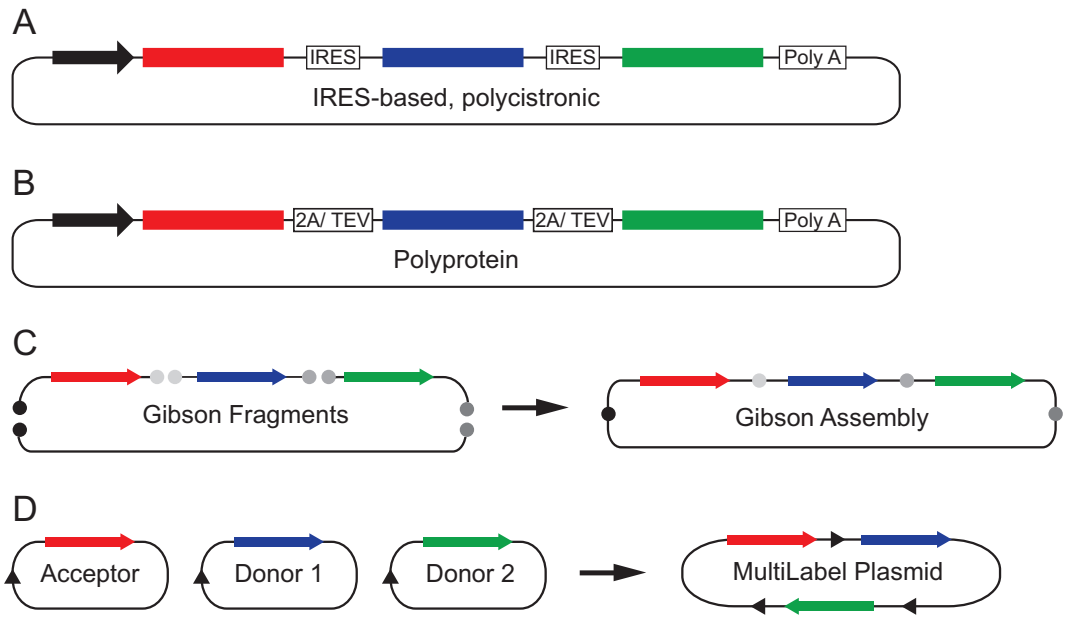


Fig. 1. Strategies for multigene expression. (A) IRES-based: a single transcript can lead to multiple proteins if the coding regions (colored boxes) are separated by an IRES element. (B) Polyprotein: a single transcript leads to a long polyprotein. The individual proteins are then released either by "self-cleavage" (see text) or by a coexpressed protease such as TEV. (C) Gibson assembly allows the assembly of independent expression cassettes (colored arrows). Assembly occurs at homologous ends of the fragments (indicated by grey-shaded circles). (D) cre/LoxP-based: Independent expression cassettes on Acceptor and Donor plasmids are assembled by a cre/LoxP reaction to yield a single plasmid. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

expression system to express Rab GTPases and phosphoinositide sensors to follow trafficking of activated receptors in living cells (Ballmer-Hofer et al., 2011; Kriz et al., 2010). Sensors for monitoring protein–protein binding are often composed of two subunits that transiently interact (Massoud et al., 2007). Fluorescence resonance energy transfer (FRET) and bioluminescence resonance energy transfer (BRET) sensors require coexpression of two fluorescent proteins or of luciferase together with a fluorescent protein (Sun et al., 2011). For bimolecular fluorescence complementation (BiFC) assays, two parts of a fluorescent protein reconstitute upon binding (Kerppola, 2013). Multiprotein expression systems allow in these assays the expression of the two proteins in a defined stoichiometric ratio.

The manipulation of cell fates is another interesting application area for multiprotein expression systems. It was shown that the expression of the transcription factors Oct3/4, Sox2, c-Myc, and Klf4 allows the reprogramming of somatic cells into pluripotent stem cells (Takahashi and Yamanaka, 2006). In the meantime several other combinations of transcription factors have been used for the reprogramming of cells. For example, it was shown that the coexpression of Asc1, Brn2, and Myt1l allows direct conversion of fibroblasts into functional neurons (Vierbuchen et al., 2010). So far these studies mainly involved coinfecting several Lentiviruses, but the coexpression from a single plasmid might be an interesting strategy to overcome the low efficiency of the process.

Drug development is another important application field for multiprotein expression systems. First, multiprotein complexes themselves might act directly as potential drugs. Virus-like particles (VLPs) consisting of components of a virus without its genetic material can be used to produce safe vaccines (Roy and Noad, 2008). VLPs were shown to stimulate the immune response and could replace attenuated viruses that are currently used for vaccination (French et al., 1990; Gheysen et al., 1989; Noad and Roy, 2003). Second, multiprotein expression systems can be used to develop screening platforms for drug discovery. Libraries of small molecules are usually used to identify new agonists or antagonists of GPCRs and receptor tyrosine kinases. Very often, receptors form heterodimeric complexes and a suitable primary cell type is not available. For example, the serotonin receptor 2A (5-HT2A) forms either a complex with the dopamine receptor D2 (D2DR) or the metabotropic glutamate receptor-2 (mGluR2) (Borroto-Escuela et al., 2010; Gonzalez-Maeso et al., 2008). In this situation, a heterologous cell line is a suitable tool. In such a cell line it is important that all cells express both receptors, since otherwise a mixed response is measured in the assay.

As described above, mammalian expression systems have to fulfill special requirements depending on the application. Many strategies have been developed in the last years to fulfill these requirements, and all systems have their advantages and disadvantages. Viral systems usually lead to high infection rates (up to 100%), but their capacity as carriers of foreign DNA is limited and the expression is usually transient. The translation of a polyprotein followed by proteolytic cleavage allows a short mRNA, but the N- and C-terminus of the mature protein are then modified which may lead to problems with proteins that need a native N- and C-terminus (e.g. Rab GTPases). Here we

review strategies for multigene expression in eukaryotic cells and we discuss their applications in cell and structural biology.

2. Cotransfection and coinfection

The most straightforward way to obtain coexpression in mammalian cells is by cotransfection of several vectors each bearing a single, independent expression cassette. The relative expression levels can be titrated by adjusting the ratio between the plasmids or by the use of different promoters. This approach is efficient if only three to four genes have to be expressed and if the cell line is easy to transfect. In our experience, the approach fails in cells that are difficult to transfect (*e.g.*, primary cells). The generation of stable cell lines is possible if plasmids contain different selection markers, but this is tedious. In addition, the approach leads to a heterogeneous cell population with different expression levels and it is not guaranteed that all cells express all proteins. For example, in a co-transfection experiment containing two plasmids A and B, a fraction of cells will be transfected only with either A or B, and only a part of the cell population is transfected with both A and B (Kriz et al., 2010). In addition, the expression ratio between protein A and B is variable. This might be an advantage when single cells are selected for analysis (*e.g.*, by microscopy) but is a problem when the whole cell population is analyzed with a biochemical assay. An atypical but successful example of this strategy is the generation of influenza A viruses for vaccination using an eight-plasmid transfection system in HEK293T cells. This strategy profits from the fact that a replication competent virus is produced. It is therefore sufficient that only a small portion of the cells take up all eight plasmids (Hoffmann et al., 2000; Nakamura et al., 2008).

Coinfection with multiple lentiviruses is a similar strategy. Lentiviral vectors are suitable for overexpression of a gene of interest or for downregulation using short hairpin RNA technology (Dull et al., 1998; Naldini et al., 1996). The vectors stably integrate into the genome of dividing and non-dividing cells. As for cotransfections, this yields a heterogeneous but stable cell population. This effect is used in the so called RGB labeling system. Three lentiviral gene ontology (LeGO) vectors encoding red, green or blue fluorescent proteins are simultaneously used to transduce cells. Individual cells are thereby marked by different combinations of inserted vectors, resulting in the generation of numerous mixed colors. This allows monitoring clonal expansion of cells *in vitro* and *in vivo* (Weber et al., 2011). A similar approach was used to identify factors that are able to convert fibroblast to neurons. Here, a library of 19 lentiviruses was used to identify combinations that allow this conversion (Vierbuchen et al., 2010). If a homogenous cell population is needed, appropriate selection markers must be used. This can be achieved by coexpressing a fluorescent marker proteins followed by FACS sorting or by coexpression of an antibiotic resistance gene (Weber et al., 2010). Lentiviruses have the disadvantage that they have limited capacity as carriers of foreign DNA, and that their production is time consuming. In addition, they are usually classified at Biosafety Level 2, which requires a special laboratory infrastructure.

Coinfection of baculoviruses with mammalian expression cassettes (BacMam) can also be used for multigene expression. This strategy was successfully used to replace the generation of stable cell lines in the drug discovery process. In contrast to retroviruses, baculoviruses do not integrate into the genome of host cells. The expression is therefore only transient (Davenport et al., 2009; Kost et al., 2010).

3. Internal ribosomal entry sites (IRES)

The use of IRES sequences is the most straightforward approach to extending the above strategy. IRES sequences were initially identified in picornaviruses, and they allow the translation initiation from internal sites from mRNAs in a 5'cap independent manner (Jang et al., 1988; Pelletier and Sonenberg, 1988). In the meantime, IRES sequences were found in other viruses and in eukaryotic genes (Mokrejs et al., 2006). The use of IRES elements allows the design of polycistronic mRNAs for the expression of two or more proteins, which is rather unusual for mammalian cells. This strategy allows relatively short constructs, since IRES sequences are typically 150–500 bp long (Mokrejs et al., 2010). In comparison, an independent expression cassette requires an additional promoter and polyadenylation signal, which together encompass approximately 1000 bp. The translation of the first ORF is cap dependent, whereas the translation of the following ORFs depends on IRES sequences. An advantage of IRES elements is that they are active in situations where cap-dependent translation is inhibited, such as in the reduction in overall protein synthesis in response to stress during cell cycle or apoptosis (Holcik et al., 2000). The coexpression with IRES elements is very efficient, meaning that typically more than 90% of transfected cells express both proteins (Ghattas et al., 1991). However, problems arise when the expression levels of the genes upstream and downstream of the IRES are compared. The imbalance of protein expression seen with the use of different IRES sequences makes it difficult to predict protein expression levels. Hotta, et al., showed for example that it is difficult to produce heavy and light chains of an IgG at similar levels with an IRES-based system (Hotta et al., 2004). Studies comparing the expression levels of two cDNA sequences separated by an IRES have shown that genes cloned downstream of the IRES were expressed at significantly lower levels (10–50% of the upstream gene, (Zhu et al., 1999). In these studies, this appeared to be independent of the gene sequence, as the same gene placed before the IRES resulted in high expression. However, other groups have reported that the coding sequence can affect IRES translation efficiency (Hennecke et al., 2001). Mutated IRES elements were developed that allow different expression levels (Poyry et al., 2001).

In conclusion, IRES sequences are a useful tool for many applications even though their effectiveness is controversially discussed (see e.g. Kozak, 2005). The main application is the coexpression of a selectable marker for the

generation of a stable cell line or the coexpression of a fluorescent protein to identify transfected cells. In both cases, the expression level is not very important. Commercial cloning vectors are available for this purpose. It is also possible to combine several IRES sequences to express more than two proteins or to add elements that allow tetracycline-inducible expression or direct generation of stable cell lines (Fussenegger et al., 1997). IRES sequences are also functional *in vivo*. For example, Sangiorgi et al. introduced an IRES-CreERT2 element into the 3′ untranslated region of the *Bim1* gene. Crossing these mice with an appropriate reporter mouse line (ROSA26-LacZ) allowed lineage tracing of Bim1-positive cells. These mice were also used to conditionally express diphtheria toxin to delete this cell lineage (Sangiorgi and Capecchi, 2008).

4. Direct assembly of multiple expression cassettes

Assembly of multiple expression cassettes by classical cloning with restriction enzymes and ligases is another strategy to obtain multigene expression. This approach has two major technical disadvantages, namely that it is difficult to find unique cloning sites in large DNA fragments and that cloning with large DNA fragments is not very efficient. The first problem can be overcome by using homing endonucleases (*e.g.,* I-CeuI) in combination with restriction enzymes with compatible cohesive ends (*e.g.*, BstXI). This approach also permits the reconstitution of the homing endonuclease site, allowing iterative insertion of additional genes (Kriz et al., 2010).

Several methods such as ligase-independent cloning (LIC) (Aslanidis and de Jong, 1990), sequence and ligation-independent cloning (SLIC) (Li and Elledge, 2007), or Gibson assembly (Gibson et al., 2009) were developed to allow cloning of large fragments independent of restriction enzymes. In the Gibson isothermal assembly technique, the exonuclease digests back the ends of each fragment, leaving 3′ single-stranded overhangs. Fragments anneal to each other through their complementary overhangs, the gaps are filled in by DNA polymerase, and the nicks are sealed by DNA ligase. This strategy facilitates assembly of multiple fragments to form large expression constructs or even small bacterial genomes (Lartigue et al., 2009). Based on Gibson assembly, Guye, et al., described a similar method for assembly of complex gene circuits which is useful particularly in synthetic biology research (Guye et al., 2013). The method fuses different transcriptional units (TU), each one containing a specific promoter and termination sequence. First, every TU is cloned into an appropriate destination vector called a "position vector" using Gateway recombination. Each TU in a position vector is flanked by two different 40 bp UNSs (unique nucleotide sequence) that are flanked in turn by two I-SceI restriction sites. I-SceI digestion of each position vector releases a fragment containing one TU with two different UNSs at its ends. These TU fragments are then assembled in a Gibson reaction where the UNS 1/2 fragment is joined to the UNS 2/3 fragment, which is then joined to the UNS 3/4 fragment. The final UNS (N-1)/N fragment is joined using an adapter vector containing UNS N/X to the carrier vector containing UNS 1/X and *Escherichia coli* propagation sequences. Using their method, they could assemble 11 TUs into a single vector and they could hierarchically assemble a 45 kb 7-TU module with 5 additional TUs to create a 63 kb 12-TU circuit (Guye et al., 2013).

5. Release from polyproteins

Polyproteins are long protein precursors spaced by proteolytic cleavage sites. This approach is inspired by certain viruses such as corona- and retroviruses that release their proteins from a large polyprotein precursor (Szymczak et al., 2004). Polyprotein approaches have proven to be particularly powerful for balancing the stoichiometry of coexpressed proteins. Nevertheless the degradation rate can still be different for each released protein leading to different protein ratios under steady state conditions. In addition, it is not possible to simultaneously express cytoplasmic and secreted proteins. For *in vivo* applications it is necessary to coexpress the protease, because endogenous proteases are not specific enough. Usually, proteases with a long recognition site such as TEV protease are used for this purpose to gain specificity (Chen et al., 2010). After cleavage, N- and C-termini are modified compared to the native protein, which might be a problem for certain applications.

The most elegant method involves the use of "self-cleaving" 2A peptide-based polyproteins, which cleave via a non-protease-based cleavage mechanism and therefore are independent of cofactors. The basis is rather a ribosomal skipping effect (Donnelly et al., 2001). The best characterized 2A peptides derive from Picornaviridae, as for example from foot-and-mouth disease virus. The recognition sequence, with an average length between 18 and 22 amino acids, encodes a highly conserved consensus motif (Asp-Val/Ile-Glu-X-Asn-Pro-Gly-⇓-Pro) that has to be placed between the two proteins. The 2A peptide appears to disrupt translation and impair normal peptide bond formation between the last glycine residue of the consensus sequence and the first proline residue of the second protein (marked by an arrow in the consensus sequence above). There are several advantages of using 2A peptide sequences, most notably their small size and their ability for efficient coexpression of genes that are placed between them. Furthermore, antibodies have been generated against the consensus 2A sequence, which thus serves as a useful 'tag' for identifying proteins in biochemical assays. On the other hand, a protein with modified ends is produced, which might be a problem for certain applications.

Szymczak, et al., showed that a single 2A peptide-linked retroviral vector can be used to generate all four CD3 proteins (CD3 epsilon, gamma, delta, zeta) and restore T cell development and function in CD3-deficient mice (Szymczak et al., 2004). Also, Fang, et al., described an antibody delivery system that allows continuous production of a full-length antibody at high concentrations *in vivo* after gene transfer. The antibody is expressed from a single open reading frame by linking the heavy and light chains with a 2A self-processing peptide derived from the foot-and-mouth disease virus. Using this expression system, they generated a recombinant adeno-associated virus vector

encoding a VEGFR2-neutralizing antibody which might be useful for the treatment of cancer or other diseases such as macular degeneration (Fang et al., 2005).

6. Cre recombination

The cre/loxP system can be used to fuse multiple plasmids bearing independent expression cassettes to produce a single plasmid. Cre recombinase is an enzyme derived from bacteriophage P1, and carries out site-specific recombination events. It catalyzes the recombination between two DNA recognition sites (loxP sites). A loxP site consists of a 34 base pair sequence composed of two 13 bp palindromic sequences flanking an asymmetric 8 bp spacer region. Cre recombinase can either excise the DNA between two loxP sites leading to an additional circular plasmid containing the excised DNA, or fuse two DNA fragments (Nagy, 2000). It is an equilibrium reaction that does not need additional cofactors, and it works *in vitro* and *in vivo* in a wide range of organisms. Therefore, the Cre/loxP system has found many applications in biotechnology, especially for tissue-specific conditional protein overexpression or gene knockout in mice.

MultiBac was the first system that used the cre/LoxP strategy for multiprotein expression. This system allows expression of up to 14 proteins in insect cells (Fitzgerald et al., 2006). The rationale of the system was to allow combinatorial assembly of the different components or mutants thereof for structural studies. The system was used to solve structures of multiprotein complexes, *e.g.*, for the structure determination of the core elements of TFIID (Bieniossek et al., 2013). The concept was in the meantime adapted for *E. coli* (Acembl, (Bieniossek et al., 2009)) and mammalian expression (MultiLabel; (Kriz et al., 2010)). MultiLabel consists of Acceptor and Donor vectors. All vectors contain a typical expression cassette (promoter – gene-of-interest – poly A signal), a loxP site, and different antibiotic resistance genes. Acceptor plasmids bear a ColE1 origin of replication whereas Donor plasmids have a conditional origin of replication (R6Kγ). One Acceptor plasmid is then fused with up to four Donor plasmids by an *in vitro* cre/loxP reaction. Correct fusion plasmids are then selected by appropriate combinations of antibiotics; unfused donor plasmids cannot propagate in standard cloning strains due to the conditional origin of replication. The resulting plasmid contains up to five independent expression cassettes that can be used for transient transfections, giving rise to homogenous cell populations (Kriz et al., 2010, 2011). Acceptor plasmids can contain additional features such as selection markers for the generation of stable cell lines or sites for homing endonuclease enabling linearization prior to transfection. The system was successfully used to generate stable cell lines with multiple fluorescent sensors or to monitor receptor trafficking (Ballmer-Hofer et al., 2011; Kriz et al., 2010).

7. Concluding remarks

Powerful technologies for the recombinant expression of multiple proteins in mammalian cells have been developed in recent years. Reliable protocols exist for all systems, and many examples show that they work. There are subtle differences in these systems that might be relevant for diverse applications. Future users should be aware if they need (1) transient expression, (2) stable expression, (3) a homogenous cell population, (4) modular cloning for screening, (5) a short construct, or (6) unmodified N- and C-termini. The system should be chosen based on these criteria.

Acknowledgments

We thank Drs. Kurt Ballmer-Hofer and Ned Mantei for critical reading of the manuscript. Our work is supported by the Swiss National Science Foundation (Grant 31003A_146975 to P.B.),

References

Aslanidis, C., de Jong, P.J., 1990. Ligation-independent cloning of PCR products (LIC-PCR). Nucleic Acids Res. 18, 6069–6074.
Ballmer-Hofer, K., Andersson, A.E., Ratcliffe, L.E., Berger, P., 2011. Neuropilin-1 promotes VEGFR-2 trafficking through Rab11 vesicles thereby specifying signal output. Blood 118, 816–826.
Bieniossek, C., Nie, Y., Frey, D., Olieric, N., Schaffitzel, C., Collinson, I., Romier, C., Berger, P., Richmond, T.J., Steinmetz, M.O., et al., 2009. Automated unrestricted multigene recombineering for multiprotein complex production. Nat. Methods 6, 447–450.
Bieniossek, C., Papai, G., Schaffitzel, C., Garzoni, F., Chaillet, M., Scheer, E., Papadopoulos, P., Tora, L., Schultz, P., Berger, I., 2013. The architecture of human general transcription factor TFIID core complex. Nature 493, 699–702.
Borroto-Escuela, D.O., Romero-Fernandez, W., Tarakanov, A.O., Marcellino, D., Ciruela, F., Agnati, L.F., Fuxe, K., 2010. Dopamine D2 and 5-hydroxytryptamine 5-HT((2)A) receptors assemble into functionally interacting heteromers. Biochem. Biophys. Res. Commun. 401, 605–610.
Chen, X., Pham, E., Truong, K., 2010. TEV protease-facilitated stoichiometric delivery of multiple genes using a single expression vector. Protein Sci. 19, 2379–2388.
Davenport, E.A., Nuthulaganti, P., Ames, R.S., 2009. BacMam: versatile gene delivery technology for GPCR assays. Methods Mol. Biol. 552, 199–211.
Donnelly, M.L., Luke, G., Mehrotra, A., Li, X., Hughes, L.E., Gani, D., Ryan, M.D., 2001. Analysis of the aphthovirus 2A/2B polyprotein 'cleavage' mechanism indicates not a proteolytic reaction, but a novel translational effect: a putative ribosomal 'skip'. J. Gen. Virol. 82, 1013–1025.
Dull, T., Zufferey, R., Kelly, M., Mandel, R.J., Nguyen, M., Trono, D., Naldini, L., 1998. A third-generation lentivirus vector with a conditional packaging system. J. Virol. 72, 8463–8471.
Fang, J., Qian, J.J., Yi, S., Harding, T.C., Tu, G.H., VanRoey, M., Jooss, K., 2005. Stable antibody expression at therapeutic levels using the 2A peptide. Nat. Biotechnol. 23, 584–590.
Fitzgerald, D.J., Berger, P., Schaffitzel, C., Yamada, K., Richmond, T.J., Berger, I., 2006. Protein complex expression by using multigene baculoviral vectors. Nat. Methods 3, 1021–1032.
French, T.J., Marshall, J.J., Roy, P., 1990. Assembly of double-shelled, virus like particles of bluetongue virus by the simultaneous expression of four structural proteins. J. Virol. 64, 5695–5700.
Fussenegger, M., Moser, S., Mazur, X., Bailey, J.E., 1997. Autoregulated multicistronic expression vectors provide one-step cloning of regulated product gene expression in mammalian cells. Biotechnol. Prog. 13, 733–740.
Geisse, S., Kocher, H.P., 1999. Protein expression in mammalian and insect cell systems. Methods Enzymol. 306, 19–42.
Ghattas, I.R., Sanes, J.R., Majors, J.E., 1991. The encephalomyocarditis virus internal ribosome entry site allows efficient coexpression of two genes from a recombinant provirus in cultured cells and in embryos. Mol. Cell. Biol. 11, 5848–5859.
Gheysen, D., Jacobs, E., de Foresta, F., Thiriart, C., Francotte, M., Thines, D., De Wilde, M., 1989. Assembly and release of HIV-1 precursor Pr55gag virus-like particles from recombinant baculovirus-infected insect cells. Cell 59, 103–112.

Gibson, D.G., Young, L., Chuang, R.Y., Venter, J.C., Hutchison 3rd, C.A., Smith, H.O., 2009. Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods 6, 343–345.
Gonzalez-Maeso, J., Ang, R.L., Yuen, T., Chan, P., Weisstaub, N.V., Lopez-Gimenez, J.F., Zhou, M., Okawa, Y., Callado, L.F., Milligan, G., et al., 2008. Identification of a serotonin/glutamate receptor complex implicated in psychosis. Nature 452, 93–97.
Guye, P., Li, Y., Wroblewska, L., Duportet, X., Weiss, R., 2013. Rapid, modular and reliable construction of complex mammalian gene circuits. Nucleic Acids Res. 41, e156.
Hennecke, M., Kwissa, M., Metzger, K., Oumard, A., Kroger, A., Schirmbeck, R., Reimann, J., Hauser, H., 2001. Composition and arrangement of genes define the strength of IRES-driven translation in bicistronic mRNAs. Nucleic Acids Res. 29, 3327–3334.
Hoffmann, E., Neumann, G., Kawaoka, Y., Hobom, G., Webster, R.G., 2000. A DNA transfection system for generation of influenza A virus from eight plasmids. Proc. Natl. Acad. Sci. USA 97, 6108–6113.
Holcik, M., Sonenberg, N., Korneluk, R.G., 2000. Internal ribosome initiation of translation and the control of cell death. Trends Genet. 16, 469–473.
Hotta, A., Kamihira, M., Itoh, K., Morshed, M., Kawabe, Y., Ono, K., Matsumoto, H., Nishijima, K., Iijima, S., 2004. Production of anti-CD2 chimeric antibody by recombinant animal cells. J. Biosci. Bioeng. 98, 298–303.
Jang, S.K., Krausslich, H.G., Nicklin, M.J., Duke, G.M., Palmenberg, A.C., Wimmer, E., 1988. A segment of the 5′ nontranslated region of encephalomyocarditis virus RNA directs internal entry of ribosomes during in vitro translation. J. Virol. 62, 2636–2643.
Kerppola, T.K., 2013. Simultaneous visualization of multiple protein interactions using multicolor bimolecular fluorescence complementation (BiFC) analysis. Cold Spring Harb. Protoc. 2013, 892–895.
Kost, T.A., Condreay, J.P., Ames, R.S., 2010. Baculovirus gene delivery: a flexible assay development tool. Curr. Gene Ther. 10, 168–173.
Kozak, M., 2005. A second look at cellular mRNA sequences said to function as internal ribosome entry sites. Nucleic Acids Res. 33, 6593–6602.
Kriz, A., Schmid, K., Baumgartner, N., Ziegler, U., Berger, I., Ballmer-Hofer, K., Berger, P., 2010. A plasmid-based multigene expression system for mammalian cells. Nat. Commun. 1, 120.
Kriz, A., Schmid, K., Ballmer-Hofer, K., Berger, P., 2011. Integration of multiple expression cassettes into mammalian genomes in a single step. Protoc. exch. http://dx.doi.org/10.1038/protex.2011.249.
Lartigue, C., Vashee, S., Algire, M.A., Chuang, R.Y., Benders, G.A., Ma, L., Noskov, V.N., Denisova, E.A., Gibson, D.G., Assad-Garcia, N., et al., 2009. Creating bacterial strains from genomes that have been cloned and engineered in yeast. Science 325, 1693–1696.
Li, M.Z., Elledge, S.J., 2007. Harnessing homologous recombination in vitro to generate recombinant DNA via SLIC. Nat. Methods 4, 251–256.
Massoud, T.F., Paulmurugan, R., De, A., Ray, P., Gambhir, S.S., 2007. Reporter gene imaging of protein-protein interactions in living subjects. Curr. Opin. Biotechnol. 18, 31–37.
Mokrejs, M., Vopalensky, V., Kolenaty, O., Masek, T., Feketova, Z., Sekyrova, P., Skaloudova, B., Kriz, V., Pospisek, M., 2006. IRESite: the database of experimentally verified IRES structures (www.iresite.org). Nucleic Acids Res. 34, D125–D130.
Mokrejs, M., Masek, T., Vopalensky, V., Hlubucek, P., Delbos, P., Pospisek, M., 2010. IRESite–a tool for the examination of viral and cellular internal ribosome entry sites. Nucleic Acids Res. 38, D131–D136.
Nagy, A., 2000. Cre recombinase: the universal reagent for genome tailoring. Genesis 26, 99–109.
Nakamura, Y., Patrushev, N., Inomata, H., Mehta, D., Urao, N., Kim, H.W., Razvi, M., Kini, V., Mahadev, K., Goldstein, B.J., et al., 2008. Role of protein tyrosine phosphatase 1B in vascular endothelial growth factor signaling and cell-cell adhesions in endothelial cells. Circ Res 102, 1182–1191.
Naldini, L., Blomer, U., Gallay, P., Ory, D., Mulligan, R., Gage, F.H., Verma, I.M., Trono, D., 1996. In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector. Science 272, 263–267.
Noad, R., Roy, P., 2003. Virus-like particles as immunogens. Trends Microbiol. 11, 438–444.
Pelletier, J., Sonenberg, N., 1988. Internal initiation of translation of eukaryotic mRNA directed by a sequence derived from poliovirus RNA. Nature 334, 320–325.
Perrakis, A., Romier, C., 2008. Assembly of protein complexes by coexpression in prokaryotic and eukaryotic hosts: an overview. Methods Mol. Biol. 426, 247–256.
Poyry, T.A., Hentze, M.W., Jackson, R.J., 2001. Construction of regulatable picornavirus IRESes as a test of current models of the mechanism of internal translation initiation. RNA 7, 647–660.
Roy, P., Noad, R., 2008. Virus-like particles as a vaccine delivery system: myths and facts. Hum. Vaccines 4, 5–12.
Sangiorgi, E., Capecchi, M.R., 2008. Bmi1 is expressed in vivo in intestinal stem cells. Nat. Genet. 40, 915–920.
Sun, Y., Day, R.N., Periasamy, A., 2011. Investigating protein–protein interactions in living cells using fluorescence lifetime imaging microscopy. Nat. Protoc. 6, 1324–1340.
Szymczak, A.L., Workman, C.J., Wang, Y., Vignali, K.M., Dilioglou, S., Vanin, E.F., Vignali, D.A., 2004. Correction of multi-gene deficiency in vivo using a single 'self-cleaving' 2A peptide-based retroviral vector. Nat. Biotechnol. 22, 589–594.
Takahashi, K., Yamanaka, S., 2006. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell 126, 663–676.
Vierbuchen, T., Ostermeier, A., Pang, Z.P., Kokubu, Y., Sudhof, T.C., Wernig, M., 2010. Direct conversion of fibroblasts to functional neurons by defined factors. Nature 463, 1035–1041.
Vijayachandran, L.S., Viola, C., Garzoni, F., Trowitzsch, S., Bieniossek, C., Chaillet, M., Schaffitzel, C., Busso, D., Romier, C., Poterszman, A., et al., 2011. Robots, pipelines, polyproteins: enabling multiprotein expression in prokaryotic and eukaryotic cells. J. Struct. Biol. 175, 198–208.
Weber, K., Mock, U., Petrowitz, B., Bartsch, U., Fehse, B., 2010. Lentiviral gene ontology (LeGO) vectors equipped with novel drug-selectable fluorescent proteins: new building blocks for cell marking and multi-gene analysis. Gene Ther. 17, 511–520.
Weber, K., Thomaschewski, M., Warlich, M., Volz, T., Cornils, K., Niebuhr, B., Tager, M., Lutgehetmann, M., Pollok, J.M., Stocking, C., et al., 2011. RGB marking facilitates multicolor clonal cell tracking. Nat. Med. 17, 504–509.
Zhu, J., Musco, M.L., Grace, M.J., 1999. Three-color flow cytometry analysis of tricistronic expression of eBFP, eGFP, and eYFP using EMCV-IRES linkages. Cytometry 37, 51–59.

Multigene delivery in Mammalian Cells: Advances and Recent Applications

Maysam Mansouri[1], Aurélien Rizk[1] and Philipp Berger[1]

[1] Paul Scherrer Institute, Biomolecular Research, Molecular Cell Biology, CH-5232 Villigen, Switzerland

[*] Corresponding author:
Dr. Philipp Berger
Paul Scherrer Institute
Molecular Cell Biology
CH-5232 Villigen
Phone: 0041 56 310 4728
FAX: 0041 56 310 5288
E-mail: Philipp.Berger@psi.ch

Abstract

Systems for multigene delivery in mammalian cells are essential tools in current biological research. Initially, they were based on RNA polymerase II promoters and were used for production of protein complexes and or for applications in cell biology like reprogramming of somatic cells to stem cells. Emerging new technologies like CRISPR/Cas9, enabling any alteration in the genomic level of an organism, need additional elements like U6-driven expression cassettes for RNA expression or homology constructs for targeted genome editing. For applications like this, systems with high DNA capacity, flexibility, and transfer rates are needed. In this article, we briefly give an update on some of recent strategies that facilitate multigene assembly and delivery into mammalian cells. Also, we review some of multigene delivery applications in a various fields of biology such as genome editing, cellular reprogramming, cell signaling and trafficking, and many others.

Key words: Multigene delivery system, Co-expression system, MultiLabel, MultiPrime.

Introduction

Multigene delivery systems gain increasing importance in biological research in both academia and industry. Many molecular process within mammalian cells such as DNA replication and repair, cell division, gene expression, protein sorting and trafficking rely on interaction between many protein or molecular machines, which were also termed as protein sociology in the cell (Robinson et al., 2007). Understanding the complex functions of these machines require the ability to manipulate their components in single cells by overexpression, repression or activation (Dominguez et al., 2015). In addition, manipulation of cells very often needs the introduction of heterologous DNA into a cell. Examples include labelling of living cells with various fluorescently-tagged sensors for monitoring changes in cellular architecture or metabolism (Ballmer-Hofer et al., 2011) or production of complex protein or therapeutic vectors (Assenberg et al., 2013a). Also, these systems allow to engineer mammalian cells on the genomic level (Hsu et al., 2014) or reprogram them through change of cell fate (Takahashi and Yamanaka, 2006). Construction of multigene circuits in synthetic biology (Guye et al., 2013), and gene therapy (Hotta and Yamanaka, 2015) are other examples that show central role of multigene delivery systems in biological research.
Nevertheless, delivery of multiple genes simultaneously to single cells is still challenging. Hallmarks of a good multigene delivery system are flexible design, easy to assembly, and efficient and reproducible delivery to a variety of cells including diving and non-dividing cells (Mansouri et al., 2016).

Multigene expression systems were initially developed to improve production of protein complexes in E. coli and insect cells. Although mammalian expression systems are of course also suitable for protein production, they are mainly used to study or to manipulate physiological process within cells (Mansouri and Berger, 2014). Multigene delivery systems can be used to introduce multiple expression cassettes into immortalized cell lines (Kriz et al., 2010). Established cell lines, harboring some alterations that lead to indefinitely division, are desired usually for generation of biotechnological products. Primary cells, which retain

many functions seen in vivo, since they express many protein at their in vivo level, are an attractive source in research. Furthermore, stem cells, of either embryonic or adult origin, as well as those derived from differentiated cells, are now an emerging and reliable source for investigations (Eglen and Reisine, 2011). For instance, induced pluripotent stem cells (iPSC) can be derived directly from patients and provide a valuable source of information (Dimos et al., 2008). The ability to delivery of multiple genetic expression cassettes into mammalian cells (e.g., established cell lines, primary cells and stem cells) make a great possibility for production, research and therapy. Here, we review available methods for delivery of multigene constructs into mammalian cells. Then, we describe some the most frequent and efficient strategies for assembling of multigene cassettes. Finally, we introduce some field of biological research that benefit from multigene delivery systems.

Methods for multigene delivery to mammalian cells

Three major methods are used for multigene delivery (as like single gene delivery) to mammalian cells, including physical methods, chemical methods and biological methods (Nelson and Gersbach, 2016). Physical methods, which are also known as carrier-free delivery systems, lead to a temporary disruption of cell the membrane allowing the desired plasmid to enter the cell (Knutson and Yee, 1987). Electroporation, magnetofection, microinjection and sonoportion are the most frequent physical methods used for delivery of plasmids into mammalian cells (Nelson and Gersbach, 2016). Physical methods are usually fast and simple but need to special equipment. In addition, reproducibility of achieved result by physical methods are divers and their damaging effect on cells is high. Chemical-mediated multigene delivery is the next method for gene delivery into mammalian cells. Liposome (e.g. DOTMA), cationic lipids (e.g. Lipofectamin) or non-liposomal lipids (e.g. Fusion HD) and polymers (e.g., polyethylenimine (PEI)) are the most common chemical reagents that have been pursued to deliver DNA to mammalian cells. These reagents are easy to work and fast and have moderate toxicity (Heiser, 2003). Biological methods include integrating viral

vectors (e.g., retrovirus and lentivirus), non-integrating viral vectors (e.g., Adenovirus, Adeno-associated virus, and Baculovirus) (Nelson and Gersbach, 2016; Smith, 1995). Biological methods deliver foreign genes with high efficiency to many cell types. On the other hand, they suffer from low cargo capacity for genetic material, and/or possibility to random integration into the host genome can cause mutations and tumorigenesis. Additionally, working with biological particles can raise biosafety issues. Table 1 compared some of the gene delivery features by these three methods.

An update on strategies for multigene expression in mammalian cells

Many strategies for assembly of multigene cassettes and for delivery to mammalian cells exist. Here, we briefly review some these strategies as well as an update in the topic. The most convenient way to co express several genes in mammalian cells is either co-transfection of plasmids using chemical or physical methods or co-infection with a pool of virus each bearing single cassette. This strategy is straightforward, fast and easy to do but leading to generation of a heterologous cell population or imbalance level of expression in every single cell (Kriz et al., 2010). Internal ribosomal entry sites (IRES) is a good choice for co-expression of reporter/selection genes (e.g., GFP or antibiotic resistance gene) along with protein of interest. This system ensures expression of a second gene that is located downstream of IRES sequence. In this strategy, genes located up and downstream of IRES sequence are transcribed as a single polycistronic mRNA. Unfortunately, ratio of expressed genes in the up/down stream of IRES sequence are imbalanced and not predictable (Zhu et al., 1999). Polyprotein approaches have been proven to be particularly powerful for balancing the stoichiometry of co-expressed proteins. Usually, proteases with a long recognition sites such as TEV protease are used for this purpose to gain specificity (Chen et al., 2010). After cleavage, N- and C-termini are modified compared to the native protein, which might be a problem for certain applications. The most elegant method involves the use of ‘‘self-cleaving’’ 2A

peptide-based polyproteins, which produce multiple proteins via ribosomal skipping. (de Felipe, 2004).

Many strategies for direct cloning of multiple expression cassettes exist. Examples are overlap extension polymerase chain reaction, circular polymerase extension cloning, sequence and ligase independent cloning (Wang et al., 2012; Ye et al., 2014). Another approach for simultaneous DNA fragment assembly is provided by Golden Gate and Golden Gate-related systems (Engler et al., 2008, 2009; Werner et al., 2012). A powerful assembly technique that can be used for joining multiple DNA fragments simultaneously in a single tube is Gibson assembly (Gibson et al., 2009). Gibson assembly allows joining the DNA via overlapping ends. One flexible approach to assemble multigene expression cassettes to mammalian cells is based on in vitro cre/loxP recombination systems. We have developed two systems called MultiLabel (Kriz et al., 2010) and MultiPrime (Mansouri et al., 2016) which are based on plasmid transfection or baculovirus transduction. MultiLabel and MultiPrime consist of Acceptor and Donor vectors. Acceptor vectors for MultiLabel contain selection genes for the generation of stable cell lines whereas MultiPrime Acceptor contain elements for the generation of baculoviruses. Baculoviruses have the potential to transduce many cell types including primary cells. All vectors have their own individual expression cassette (promoter – gene-of-interest – poly A signal), a loxP site, and different antibiotic resistance genes. Acceptor plasmid can be fused with up to four Donor plasmids by an in vitro cre/loxP recombination reaction. (Kriz et al., 2010). Then, recombinant baculovirus genome can be applied for production of baculovirus particles within insect cells. (Mansouri et al., 2016).

Therefore, choosing a rapid, modular and reliable strategy for assembly of multigene construct is an important step for efficient cargo delivery. In the table 2 we have summarized some features of each strategy.

Application of multigene delivery in genome editing

Genome editing refers to the process of making targeted any modifications (mutation, deletion or insertion) to the genome, its contexts (e.g., epigenetic

marks), or its outputs (e.g., transcripts) (Hsu et al., 2014). Genome editing is powerful technique for gene targeting in biological researches in order to study of involved genes in disease (Nelson and Gersbach, 2016) and development of novel therapeutic medicine (Sander and Joung, 2014). Genome editing employs two key players: programmable nuclease that introduce DNA double-strand breaks (DSBs) into user-defined sites in the genome and the DSB repair mechanism in host cells (Sakuma and Yamamoto, 2015; Yamamoto, 2015). DSBs are mainly repaired through two pathways, either non- homologous end joining (NHEJ) that is inherently error-prone or homology-directed repair (HDR) which is precise compared with NHEJ but requires a repair template. In the absence of a donor repair template, the error-prone non-homologous end-joining (NHEJ) pathway fill the DSB gaps with small insertions or deletions (Indels) that can be used to disrupt the coding region of a gene of interest by alteration of open reading frame (ORF) or induction a premature stop codon (knock out) (Hsu et al., 2014; Jiang and Marraffini, 2015b). Additionally, co-introduction of targetable nucleases and exogenous repair donors (or targeting vectors) increases the HDR- mediated targeted transgene insertion (knock in) (Sakuma and Yamamoto, 2015). Among four main classes of targeted DBP-based genome editing, Clustered regularly interspaced short palindromic repeat (CRISPR)/Cas9 is more popular due to its simplicity, high efficiency and versatility in genome editing (Hsu et al., 2014; Nelson and Gersbach, 2016). CRISPR/Cas9 system, including the Cas9-based type II CRISPR system, is a part of adaptive immune systems in bacteria and archaea against invading phage (Jiang and Marraffini, 2015b). CRISPR/Cas9 system is composed of two major components: an endonuclease enzyme called Cas9 and at least one guide RNA (gRNA). Hybridization between gRNA and DNA in the specific sequence of genome leading to direction of Cas9 to the target place and cleave the DNA (Hsu et al., 2014; Jiang and Marraffini, 2015a; Nelson and Gersbach, 2016; Sander and Joung, 2014).

Three types of Cas9 nucleases are available: Cas9 and two mutated forms of it including Cas9n and dCas9. Cas9 induce double stands break in DNA whereas

Cas9n (harboring D10A mutation) generate a single stand break (nick) in DNA. dCas9 in addition to D10A bearing H840A mutation that lead to inactivation of cleavage activity, but do not prevent DNA binding (Qi et al., 2013).

The most significant challenge for the clinical translation of genome editing is the development of a safe and effective delivery vehicles (Nelson and Gersbach, 2016). In the following part, we briefly describe some of the recent application of delivery methods, their potential issues for carrying Cas9, gRNA(s), HDR and other related genes into mammalian cells.

The largest component of the CRISPR/Cas9 system is Cas9 nuclease (Nelson and Gersbach, 2016). Cas9, originally from Streptococcus pyogenes, is encode by a 4.2 kb DNA. Although gRNA itself is almost 100 bp but along with a RNA polymerase III promoter (usually U6 promoter) would be around 330 bp (Mali et al., 2013). Therefore, carrying Cas9 and gRNA on the same vector need to a robust system with large cargo capacity. Although, delivery of Cas9 by transfection can be quite efficient in many cell types (Cong et al., 2013) but in hard-to-transfect cells (e.g., primary cell types) transduction with a viral vector provides an alternative way (Mansouri et al., 2016). Major problems for delivery of CRISPR/Cas9 components by virus are either capacity of viral vector for packaging of transgenes or integration of virus into the host genome (Smith, 1995). AAV and adenovirus are a non-integrating virus and generally consider as safe and effective delivery viral vehicles (Seto et al., 2012). These viruses have a small packaging size which is almost 4.7 kb. Using smaller variant of Cas9 (e.g., staphylococcus aureus cas9 or SaCas9 encoded by a 3.2 kb DNA) or driving gRNA expression with smaller promoter, utility of AAV and adenoviruses for genome editing is increasable (Nelson and Gersbach, 2016). Integrating viral vectors like retrovirus and lentiviruses are the other options for CRISPR/Cas9 components delivery. Although retrovirus is limited to dividing cells but lentivirus is able to transduce both dividing and non-dividing cells (Smith, 1995). Transgene packaging capacity of these two viruses is almost 10 Kb (Nelson and Gersbach, 2016). This capacity is practically good for knocking out of one or multiple genes. For example, Kabadi et al (Kabadi et al., 2014) have used

multiplex knock out using single lentiviral virus. Anyway, knock in based genome editing usually need to at least one more construct including gene of interest and homology arms (Platt et al., 2014a). It seems that loading of lenti- and retroviruses with CRISPR/Cas9 is more than their capability for packaging. Additionally, major problem of lenti- and retroviruses is their random intergradation to the host genome which may lead to insertional mutagenesis and cause cancer (Nelson and Gersbach, 2016; Shao and Wu, 2010). A non-integrating virus that has a large transgene capacity is baculovirus (Kost et al., 2010). Baculovirus normally infect insect cells but by using promoters which are active in mammalian cells (e.g., CMV, CAG and etc.) they can be harnessed to deliver large constructs to mammalian cells (Mansouri et al., 2016). Recently using baculovirus, we fused GFP exactly before stop codon in the last exon of HMGA1 (a histone protein in human cells) gene (Mansouri et al., 2016). We have reported successful knock in GFP both in established cell line (HEK293) and primary cell (HUVEC). One possibility to overcome to inability of virus for large cargo delivery is co-infection of cell in a sequential manner. A good example for this case is loss-of-function screening for certain gene knock out in a high throughput assay. Here, cells are first transduced with a lentivirus expressing the Cas9 nuclease and then select to be ensure that they are expressing a high level of Cas9. Cells are then co-transduced with a lentivirus expressing a pooled gRNA library, at a multiplicity of infection (MOI) such that each cell expresses no more than one gRNA. Such libraries will aid in the study of gene functions via functional genomics in human cells (Gilbert et al., 2013, 2014). One drawback of this system is isolation of stable cell that express Cas9, expansion and again transduction is time consuming and tedious.

In the first, CRISPR/Cas9 system has developed for insertion (KI), deletion and mutation (KO) in the targeted place in the genome. KO and KI have many applications in molecular cell biology. KO can be used for study of gene functionality even in large scale for genome wide screening (Gilbert et al., 2014). Also, KI can be used for correction of a mutation that led to disease (Schwank et al., 2013), make reporter cell lines (Wu et al., 2016), creation of modeling animal

in cancer biology harboring mutation(s) in different genes (Sánchez-Rivera and Jacks, 2015), and etc.

CRISPR/Cas9 system is not only a tool for gene editing, but also it can be recapitulated for gene regulation (Sander and Joung, 2014). dCas9 protein, which is inactivated form of Cas9, retains the ability to target specific sequences through the gRNA. dCas9 can be used for down regulation of genes in mammalian cells through interfere with transcription elongation by blocking RNA polymerase II (Pol II) or the binding of important transcription factors (Dominguez et al., 2015). This genetic perturbation though CRISPR system refer as CRSPRi (interference). To increase repression by CRISPRi, the KRAB (Krüppel-associated box) or four concatenated SID4X (mSin3 interaction domains) were fused to dCas9 (dCas9 chimera). Together with a target-specific gRNA, the dCas9–KRAB or dCas9–SID4X fusion proteins can efficiently repress endogenous genes in mammalian cells (Gilbert et al., 2013, 2014). In this approach, K562 cells were stably expressed dCas9 chimera and then infected with lentivirus bearing gRNAs. Conversely, co-expression of dCas9 with recruit transcription activators that mediate gene activation is called CRISPRa (activation). Co-expression of dCas9 along with VP64 or of the p65 activation domain (p65AD) in mammalian cells lead to activation of endogenous genes, with a single gRNA (Farzadfard et al., 2013; Perez-Pinera et al., 2013). A good example for CRISPRa is SunTag system which is a co-expressed array of scFV-GFP-VP64 fused to dCas9. Sun Tag improved genomic imaging by amplifying the fluorescent signal (Tanenbaum et al., 2014). Sun Tag transfected to HEK and U2OS cells by PEI-based transfection procedure. Also, Sun Tag is based on constrictively activation of gene expression whereas using inducible promoter enable temporal gene regulation which is more interested.

Also, dCas9 can be used for regulation of epigenetic events in mammalian cells (Dominguez et al., 2015). For example, co-expression of histone demethylase LSD1with Neisseria meningitidis Cas9 (Nm dCas9) and specific gRNA removes the histone 3 Lys4 dimethylation (H3K4me2) mark from targeted distal enhancers and leading to transcription repression (Kearns et al., 2015). They used Nm

dCas9 which is an orthogonal and substantially smaller Cas9 variant allowing lentivirus packaging. Also, Nm dCas9 can works on alternative PAM (Protospacer Adjacent Motif) sequences.

dCasp can be also fused to a domain to activate gene expression. For example, Co-expression of dCas9 with histone acetyltransferase p300 and specific gRNA can activate transcription through acetylation of H3K27 (H3K27ac) (Hilton et al., 2015).

Thanks to multigene expression systems, CRISPR/dCas9 can be also used to simultaneous repression and activation of multiple genes in single mammalian cells using scaffold RNAs (Dominguez et al., 2015). Scaffold RNAs (scRNA) are extended form of gRNA with protein binding RNA aptamers. Zalatan et al (Zalatan et al., 2015) shown that co-expression of a scRNA against β-1,4-N-acetyl-galactosaminyl transferase 1 (B4GALNT1) gene and dCas9 can recruit expressed aptamer binding protein that called Com. Com, itself is fused to KRAB domain and can repress B4GALNT1 gene. Simultaneously, co-expression of scRNA, dCas9, and aptamer binding protein MSC fused to VP64 activated CXCR4 gene in mammalian cells.

Application of multigene delivery systems in cellular reprogramming

The field of stem cell biology has undergone rapid development in recent years. Some features make stem cells very attractive in medicine and industry. First is self-renewal property of stem cells that enables them to proliferate unlimited times while maintain their undifferentiated state. Second, their remarkable potential to develop into many different cell types in the body during development (Martello and Smith, 2014). Three methods and their delivery systems that enable to change the cell fate are discussed here. These methods are included reprogramming of somatic cells to stem cells (iPSC), and direct conversion of somatic cells to somatic cells (trans differentiation).

Induced- Pluripotent Stem Cell (iPSCs) are useful technique in disease modeling and drug discovery, and it promises to provide a new generation of cell-based therapeutics (Schlaeger et al., 2014). Although generation of iPS cells from

somatic cells by protein transduction or in combination with small molecules as virus-free procedure and safe is possible, but reprogramming efficiency is very low (Nelson and Terzic, 2009). The most of the techniques for reprogramming and trans-differentiation utilize plasmids or related viral vehicles that carrying one or multiple genes. These multigene delivery systems overexpress or repress certain transcription factors or pathways (Shao and Wu, 2010). In 2006, Shinya Yamanaka made a groundbreaking discovery that enable to 'reprogramme' adult, specialized somatic cells to turn them into stem cells state (Takahashi and Yamanaka, 2006). Yamanaka induced pluripotency of differentiated cells in vitro to generate induced pluripotent stem cells (iPSCs) through retroviral co-transduction of four transcription factors (TFs) include OCT3/4, SOX2, KLF4, MYC (Known as OSKM or Yamanaka's factors) (Takahashi and Yamanaka, 2006). However, cMyc may be reactivated in the iPS cells during differentiation and then induce tumor formation (32 and 33 lijing). Exclude c-Myc from OSKM lead to generation of fully functional iPS from mouse fibroblast but with more less efficiency (Nakagawa et al., 2007). Another problem of this reprogramming system is that retroviruses are not able to transduce non dividing cells (Smith, 1995). Also, they integrate to the genome and increase risk of tumorigenesis afterwards. Lentivirus encoding Oct4, Sox2, Nanog and Lin-28 has been used to reprogram human fibroblast into iPS cells (Yu et al., 2007). To increase efficiency of delivery, lentivirus was generated that four transcription factors were on the single open reading frame (ORF) and they were linked together by self-cleavage 2A (Sommer et al., 2009). Conversely, Carey et al, reported significantly lower efficiency of this reprogramming system than single vector co-transduction (Carey et al., 2009). One interpretation is co-transduction of several individual vectors allows to express factors with different (and optimal) stoichiometry whereas 2A system support near equimolar protein expression. Lentiviral system may also lead to rearrangement and instability of chromosome specially when iPS cells are differentiated into various lineage leading to tumor formation. Some groups tried to tackle this problem with developing of doxycycline-controlled lentiviral vectors (Hockemeyer et al., 2008; Maherali et al., 2008). Another way to

delivery of reprogramming factors with viral vectors is using adenovirus as vehicle. Adenovirus enable overexpression of TFs while integration into host genome is very rare. Stadtfeld et al. generated iPS cells from mouse hepatocytes using adenovirus that expressed oct4, c-Myc, Klf4 and Sox2 (Stadtfeld et al., 2008). Anyway, efficacy of this method to generate iPS cells is low (Shao and Wu, 2010). One of the safe method with lower chance for integration into the genome is transient transfection of TFs in somatic cells. Yamanaka and Gonzalez generated iPS cells from MEF though transient co-transfection (Okita et al., 2008) or a single polycistronic (Gonzalez et al., 2009) vector encoding four TFs, respectively. Problem of transient transfection is that primary cells usually are hard to transfect and also gene expression duration is usually short. An alternative for plasmid transfection is episomal vectors. This type of vectors, such as OriP/Epstein-Barr nuclear antigene-1 (EBNA1), can replicate autonomously without integration into host genome (Yates et al., 1985). Yu (Yu et al., 2009) cloned seven TFs into three different EBNA1 vectors and generated iPS cells from human fibroblast cells. Another way to generate relatively clean iPS cells is to deliver TFs to somatic cells using piggyBac system. Piggyback system is a class II mobile genetic elements that can excise themselves from their original place inside the genome and insert themselves to the new locus within genome. Problem of episomal vectors and piggyback systems is that they are systems that rely on plasmids transfection whereas transfection is hard in some cells (like primary cells) and sometimes it can be toxic for living cells. Baculocirus can be an option to reprogramming of somatic cells to iPSC. Takata et al reprogrammed MEF to iPSC by baculovirus encoding OKSM TFs in the single ORF separated by 2A self-cleavage sequence (Takata et al., 2011). Baculovirs has great capacity to deliver large DNA constructs but preparation of virus harboring all expected constructs, due to rearrangement of virus genome during production, is difficult (Fitzgerald et al., 2006).

In some cases, it is even possible to directly convert fully differentiated mature somatic cells into a variety of other somatic cell types while bypassing an intermediate pluripotent state via trans-differentiation (Jopling et al., 2011).

Marius developed an inducible lentiviral based system for trans differentiation of MEF directly to induced neuron (iN) through co expression of Ascl1, Myt1L and Brn2 (Lujan and Wernig, 2013; Vierbuchen et al., 2010; Wapinski et al., 2013). Although these viruses integrated to the genome of host cells but their expression were inducible. Therefore, it reduces chance of reactivation of TFs after trans-differentiation but problem of random integration to host genome still exist. We made a single baculovirus harboring Ascl1, Myt1L and Brn2 that can express each TF independently. Then we trans differentiated MEF to neuron (Mansouri et al., 2016). Advantage of this system is that MEF cells transduce with a high efficiency and also virus does not integrate to the genome. Trans differentiation has been also used to direct conversion of fibroblast to dopaminergic neuron, motor neuron, hepatocyte through force expression of different transcription factor based on lenti- or retroviral systems (Caiazzo et al., 2015; Jopling et al., 2011).

Although many systems have developed so far but a perfect system for reprogramming, trans differentiation and differentiation would be a single non-integrating system that is able to deliver multiple TFs or inhibitory factors with a high efficacy into dividing or non-dividing mammalian cells.

Application of multigene delivery in synthetic biology

Synthetic biology aims to engineer cells by introduction of genetic circuits in order to rewiring endogenous gene regulatory systems, rescue cellular metabolism, and regulate cellular behaviors (Wang et al., 2013). Synthetic biology is also leading to creation of innovative tools for disease modeling, drug discovery and production, and development of novel therapeutic approaches (Abil et al., 2015). Building blocks in synthetic biology is assembly of multiple gene circuits and successfully delivery of them to mammalian cells (Heng et al., 2014; Wang et al., 2013). First step in synthetic biology is to establish some methods to facilitate assembly and subsequence deliver genetic circuits faster, simpler and more efficiently (Guye et al., 2013). One of most useful applications of synthetic biology in molecular biology is construction of transcription regulators.

Transcription regulator tool enable scientist to repress or activate one or a cluster of genes at the certain time (Way et al., 2014). As motioned earlier, fusion of KRAB and VP64 to Cas9 endonuclease protein and co expression with specific gRNA(s) can act as repressor or activator, respectively (Dominguez et al., 2015; Maeder et al., 2013). A circuit of transcription factors can also be engineered to interact with endogenous signaling network (Lienert et al., 2014). Xie and et al (Xie et al., 2011) described a sensor platform consist of transcriptional/posttranscriptional synthetic regulatory circuit which were able to detect a pattern of up to six endogenous miRNAs. This sensor platform transduced via lentivirus to HEK293 (as normal) and HeLa cell (as cancer cell). Sensor platforms have used to classify cancer cells from normal cells in a mixture HEK293/Hela culture cells using unique miRNA signature. Here, miRNAs are usually small constructs and multiple of them can be convey by lentivirus. Also, lentivirus can transduce a broad range of cells enabling broad application of this sensor platform. Furthermore, production of virus is relatively fast and do not need to special equipment. One of the growing subject in synthetic biology is genetic switch tools. For example, these tools can be used for control of gene expression in mammalian cells. Deans et al have developed a tunable genetic switch to regulate gene expression (Deans et al., 2007). They showed that coupling repressor proteins and an RNAi repressor could effectively turn off expression of GFP, Cre recombinase and diphtheria toxin genes in different tested mammalian cells. They have used Lipofectamin 2000 for delivery of repressor and RNAi array to cells which is often toxic for cells. Also, primary cells are usually resistant to this type of transfection. One of the attractive applications of synthetic biology which rely on multigene delivery systems is generation of new signaling pathway and networks through pathway engineering in mammalian cells (Lienert et al., 2014; Wang et al., 2013). Recently, Morsut et al have developed a synthetic Notch (synNotch) pathways (Morsut et al., 2016). synNotch is a chimeric forms of Notch, in which both the extracellular and the intracellular of Notch receptor are replaced with heterologous protein domains. This system can drive user-defined functional responses in diverse mammalian

cell types. For inastance, they engineered a double synNotch receiver cells expressing both an anti-CD19 receptor and an anti-GFP receptor. These two receptors were each linked to a different intracellular transcription activation domain (Gal4-VP64 and tTA respectively) that in turn drives a distinct reporter fluorescent protein (mCherry and tagBFP, respectively). Stimulation of receiver cells by a sender cells (K562) expressing CD19 and GFP on the surface, lead to expression of both mCherry and BFP in the receiver cells. In this approach, chimeric Notch receptors built by fusion of different domain though 2A system and delivered to mammalian cells by co-lentiviral transduction. In this case, Lentivirus is a good vehicle because it can transduce non-dividing (e.g., neuron) and dividing cells with a high efficiency. Another application of synthetic biology that require efficient delivery of a complex gene circuits is cell engineering (Lienert et al., 2014; Way et al., 2014). One important example in this case is recapitulating of T-cell receptor (TCR) signaling pathway in non-immune cells using heterologously expression of more than 10 proteins (James and Vale, 2012). they have used a combination of both lentiviral transduction and transient transfection of cells for this T-cell engineering. Briefly, they made a cell line that constitutively expressed TCR and then co-transduced or co-transfect cells using lentivirus or GeneJuice-transfection reagent, respectively. Although they believed that for instance co transfection with four individual plasmids leading to expression of four genes simultaneously in the single cells, but in our hand it is very difficult. Usually transfection with multiple plasmid generate a heterogeneous cell population which make it difficult to analysis. Design all interested genes in one vector (e.g., MultiLabel system) can facilitate generation of homogenous cell population that all cells express all the expected genes in a favorite level. One attractive example of synthetic biology usage in clinical therapy is chimeric antigen receptor (CAR) therapy (Lienert et al., 2014). In this approach, T-cell of patient are engineered to express CAR, which is specialized against a certain antigen, and then transfer back to the patients (Brentjens et al., 2011; Gross et al., 1989). CAR based T-cell therapy highly depended on efficient and safe multigene delivery systems. For example, a synthetic signaling cascade

mediated by CAR-based T-cells has reported for treatment chronic lymphocytic leukemia (Porter et al., 2011; Ye and Fussenegger, 2014). In this example, CAR was formed from fusion of four different domains: an anti-CD19 single chain, a CD8α-derived transmembrane domain, a human CD137-derived co-stimulatory domain, and a CD3ζ signaling domain. Patient-derived T cells were transduced with CAR-encoding lentiviral particles and then were injected to patient. CD19-positive cancer cells were recognized by CAR-transgenic T cells which resulted in expansion of the engineered T-cell population and killing of the tumor cells. Although delivery of CAR constructs to the T-Cells by lentivirus is an efficient way but due to some concerns about lentivirus, its application for ex vivo experiments is restricted. Lentivirus can randomly integrate to the genome of T-cells. This integration may lead to genetic rearrangements and instability of chromosome or even it can induce mutation and cancer formation (Smith, 1995). A good application of multigene expression in applied synthetic biology has shown in cellular studies using optogenetic toolkits (Tischer and Weiner, 2014). Optogenetics is a method that uses light to modulate molecular events in a targeted manner in living cells or organisms (Müller and Weber, 2013). It relies on the use of genetically-encoded light-sensitive proteins that change conformation in the presence of light to alter cell behavior, for example, by activating signaling pathways or changing the membrane voltage potential of excitable cells (Tischer and Weiner, 2014). The most frequently used light sensors in the design of optogenetic tools for eukaryotic systems are PHYB (Phytochrom B), CRY2 (Cryptochrom 2), LOV (Light–oxygen–voltage) domains, and Dronpa (Müller and Weber, 2013; Tischer and Weiner, 2014). One of the optogenetic systems that highly required to simultaneous expression of different constructs is recently described by Bergeijk et al (van Bergeijk et al., 2015). They have established an optogenetic tool to recruit specific cytoskeletal motor proteins (kinesin, dynein or myosin) to selected cargoes within cells. This system enabled them to control intracellular transport through rapid organelle repositioning by light. For example, two plasmids harboring individual expression cassettes (PEX3-REF-LOV) and (ePDZ-GFP-Kinsin) were co transfected into

COS-7 cells to demonstrate that the motility of peroxisomes can be locally and repeatedly induced or stopped by light. They have transfected COS-7 and neuron cells by Fugene6 transfection reagent and Lipofectamin 2000, respectively. These chemical reagents are easy to work but viability and homogeneity of transfected cells are usually lower than transduction with viral vecthors.

Application of multigene delivery in developmental biology

Developmental biology aims to understand principles, mechanisms and dynamics of fundamental biological events during both normal development and its deviations in disease ([CSL STYLE ERROR: reference with no printed form.]). In many cases, it is required to analysis expression of some markers simultaneously. Although co-immunostaining of different markers using antibodies from different species is possible but it cannot extend to living cells or in vivo studies. For example, analysis of novel cell reporter lines using real-time imaging techniques allows visualization of the activity and dynamics of signaling pathways within cells. As a sample, Dietrich et al (Dietrich et al., 2015) established a fluorescence gene trap system for quantitative description of gene expression dynamics at a single-cell resolution in living mouse embryos. The gene trap cassette was consisting of a splice acceptor (SA) followed by IRES site and Venus as fluorescent marker and transduced into a Denuded 2 cell stage embryos by a self-inactivating lentivirus. This system allowed them to build a comprehensive lineage map of mouse pre-implantation development.

Multigene delivery system also can help to study role and mechanism of cell differentiation during development by over expressions or down regulations of certain multiple genes for specific duration time. For example, to understand mechanism of regulation in neuronal differentiation, Packard et al (Packard et al.) used retroviral-vector transduction to over-express Pax6 and Sox2 individually and together during post-lesion recovery. They found out that Pax6 accelerates neuronal production, while Sox2 retards it and expands the pool of neuronal progenitors. Lineage tracing is another method that highly depending to

multigene expression and delivery systems. Although established techniques (like brainbow) are based on transgenic animals but usually first step of this techniques need to delivery of multigene constructs to embryonic stem cells though viral vectors or physical methods (Weissman and Pan, 2015).

Application of multigene delivery for production of complex proteins

Expression of heterologous proteins in E. coli host system often remains a convenient choice for some reasons; low cost, fast growth and high production yield (Assenberg et al., 2013b). However, production of heterologous protein in E. coli can lead to insoluble or non-functional target proteins due to lack of post-translational modifications like mammalian cells (Tolia and Joshua-Tor, 2006). Therefore, co expression of multiple genes in E. coli, like nuclear receptor partners (Li et al., 1997) or a GroESL chaperonin (Martínez-Alonso et al., 2010; Yan et al., 2012), along with desired protein can partially overcome this shortage. Ability to production of larger and glycosylated recombinant protein with a higher yield in comparison with E. coli for structural biology purposes promoted multigene expression systems in insect cells (Assenberg et al., 2013b; Berger and Mayr, 2013). One of the most relevant systems for multiprotein expression in insect cells is MultiBac (Barford et al., 2013; Bieniossek et al., 2012). The MultiBac allows for simultaneous expression of multiple proteins in a single cell, which can be used to produce protein complexes and to recapitulate metabolic pathways in insect cells. Although glycosylation pattern in insect cells is similar to mammalian cells but still it is not completely the same (Assenberg et al., 2013b). Many attempts have been performing to improve glycosylation pattern in insect cells. Glycoengineering of either the baculovirus or its host insect cells through co expression of mammalian glycosyltransferases is one of the strategies that improve this drawback (Aumiller et al., 2012; Harrison and Jarvis, 2006; Palmberger et al., 2012). In spite of all the achieved progress, production of highly glycosylated mammalian proteins (such as secreted proteins or membrane proteins) in insect cells should be consider carefully. Conversely, mammalian expression systems are the systems of choice for production of therapeutically

important human proteins in a functional form (Jenkins et al., 2008; Walsh and Jefferis, 2006). The mammalian environment, chaperone systems and specific modifiers (i.e., kinases, phosphatases, glycosylases, and others) allowing for native post-translational modifications of the target protein are the most privilege of production of complex proteins in mammalian cells (Trowitzsch et al., 2011). Depending on the experiments, transient gene expression (easier and faster) or generation of stable cell line (more tedious but cheaper and more relevant for large-scale production) can be applied to deliver multiple-gene constructs to mammalian cells. Backliwal et al. developed an optimized protocol that though polyethyleneimine (PEI), transiently co-transfected light and heavy chains of antibody in HEK293E. They achieved an antibody expression yield of 1 g/L (Backliwal et al., 2008). This yield is in the rate of those that achieved from stable cell lines. Anyway, transient transfection is easier and process can be done in shorter time than stable cell line generation. It seems that heavy and light chains that are produced in the same cells are more functional. Luke et al reviewed that expressing the heavy chain in one cell and the light chain in another cell would not produce a functional product (Luke and Ryan, 2013). Some strategies help to force cells to express light and heavy chains together. For example, Fang et al (Fang et al., 2005) described a monoclonal antibody production system based on 2A self-processing peptide. This system, allowed continuous production of a full-length antibody and transiently transfected to HEK cells. They showed that expression of the same antibody using internal ribosomal entry sites (IRES) leads to substantially lower expression of IRES- based antibody in HEK 293 cell (0.1 μg/ml) in comparison with 2A-based Ab (1.6 μg/ml). In agree with results, Ho et al (Ho et al., 2013) compared quality and expression level of antibody produced by IRES- or furin 2A- based systems. They have shown that although level of expressed heavy chain and light chains by 2A system is higher than IRES- mediated expression but almost 40% of products generated by F2A were aggregated showing a poor post translational process after production with furin 2A system. Also, they showed that ordering of heavy and light chains can effect on expression level. They have obtained higher rate of expression when light

chain (and not heavy chain) were in the first cistron. Therefore, a system that ensure co expression of both light and heavy chain in the same ratio and independent from ordering of genes is more desirable. In agreement, we used baculovirus to deliver individual heavy and light chains, harboring on the same vector, in mammalian cells and insect cells. We produced recombinant fully functional IgG from scFV in HEK293 and primary HUVEC cells. In the same way, Goehring have used BacMam (baculovirus-mediated transduction of mammalian) technology for production and screening of recombinant membrane proteins in mammalian cells (Goehring et al., 2014). It seems that BacMam system is a good choice when transient transfection of large protein with high rate of delivery is needed. Recently Coleman et al (Coleman et al., 2016) has produced human serotonin transporter (SERT) into HEK293S GnTI− cells using BacMam system for structural biology research. Additionally, they reported X-ray crystallographic structures of human SERT, which was co expressed with GFP, at 3.15 Å resolution. Mammalian cells can also be used for large-scale production of Adeno-associated virus (AAV) requiring for gene therapy (Ayuso, 2016). Production of AAV viruses for gene therapy usually can be performed either by co transfection of three plasmids (pAAV-transgene, pAAVrep2/cap8 and pHelper) in HEK cells (Grimm et al., 1998) or two plasmids (Bac-AAV transgene and Bac-rep2/cap8) in baculovirus (Kotin, 2011).

In the all mentioned applications, a robust multi gene expression strategy as well as a reliable delivery system for production of desired complex protein/component is needed.

Application of multigene delivery in cell signaling and trafficking

Monitoring and analysis of different parameters within living cells may need to co-expression of multiple fluorescently tagged biosensors (Ballmer-Hofer et al., 2011), co-overexpression of impaired proteins (Berger et al., 2011), or up/down regulation of multiple genes (Greber and Fussenegger, 2007). We have developed a series of different intracellular biosensor systems, such as Rab GTPases regulating vesicular membrane traffic in cells, phosphoinositide binding

signaling proteins or fluorescently-labelled cytoskeletal markers (Kriz et al., 2010; Mansouri et al., 2016). These biosensors were simultaneously delivered and expressed in mammalian cells allowing to quantitatively monitor a large variety of intracellular parameters. For example, using Rab biosensors, we studied cellular trafficking of EGF- receptor after activation with EGF (Mansouri et al., 2016). In this case, we transduced COS-7 cells with a baculovirus harboring three different labeled Rab GTPases biosensors which were able to localize in different cell compartments . Co-localization of receptor-ligand complex with Rabs markers evaluated by Squassh (Rizk et al., 2014) or SquasshAnalyst (Rizk et al., 2015) softwares over the time. This method is a fair way to complete delivery of all interested biosensors simultaneously in many cells (including primary cells). Anyway, generation of baculovirus in comparison with the other viruses (like Lentivirus, adenoviruses) is more time consuming.

Other important techniques for in protein-protein interaction studies which require multi-gene delivery systems are BiFC (Bimolecular fluorescence complementation), BRET (bioluminescence resonance energy transfer), FRET (fluorescence resonance energy transfer) and split luciferase complementation assays (Kaláb and Soderholm, 2010; Trowitzsch et al., 2011). The BiFC assay is based on reconstitution of an intact fluorescent protein when two complementary non-fluorescent fragments are brought together by a pair of interacting proteins (Miller et al., 2015). FRET relies on the distance-dependent transfer of energy from a donor molecule to an acceptor molecule. FRET is the method of choice for analyzing short-term interactions of proteins in living cells (Kaláb and Soderholm, 2010). Parsons et al. (Parsons et al., 2016) have used BiFC to measure initiator caspases (caspase-2, -8, and-9) activation. They have fused non-fluorescent fragments of the fluorescent protein Venus to initiator caspase monomers. After transient co-transfection, the initiator caspase-BiFC fusion proteins remain as monomers in unstimulated cells. Following a caspase activating stimulus, the caspase fusion protein monomers interact to each other so that they refold and reconstitute the full length Venus protein, leading to fluorescence. Also, Lucker et al (Luker and Luker, 2016) described a dual-color

click beetle luciferase complementation assay for cell-based analysis interactions of two different chemokine receptors, CXCR4 and ACKR3, with the intracellular scaffolding protein β-arrestin 2. This assay provided a real-time quantification of receptor activation and signaling in response to chemokine CXCL12. They used either IRES or 2A system for co expression of different either fluorescent proteins or selectable markers along with proteins of interest to distinguish cells that received both receptors and β-arrestin 2. Hida et al (Hida et al., 2009) dissected luciferases from firefly (FLuc), click beetle in green (ELuc), and click beetle in red (CBR) to the N- and C-terminal parts. Then, the N-terminal and C-terminal fragments were fused to FK506-binding protein (FKBP) and FKBP-binding domain (FRB) and co-expressed in mammalian cells and the luminescence intensities were examined using a luminometer. This system is an imaging method for real-time analysis of protein-protein interactions using multicolor luciferases with different spectral characteristics.

In all the mentioned examples, using a reliable multigene delivery system allow co-expression of full or halves of fluorescent proteins and/or luciferase fragments which are fused to interested proteins. Although co-expression of induvial plasmids is easy but find the cells expressing both constructs are difficult. Also, cleaved proteins of interest can be expressed along with different fluorescent proteins or selectable markers using IRES- or 2A- based systems. These procedures allow to select cells that express all expected fragments by FACS or combination of antibiotics selection. Some systems like MultiLabel and MultiPrime can provide split fragments as well as fluorescent protein or selectable marker in a single vector.

Application of multigene delivery in immunology and gene therapy

Gene therapy is the introduction of new genetic material into the cells of an individual for therapeutic purposes. Therapeutic proteins are usually expensive to produce, have a short half-lives of stability, and need to administration frequently (Chernajovsky et al., 2004). Therefore, gene therapy can be an alternative method to overcome to these drawbacks. In the other hand, some therapeutic

proteins are either relatively large and complex or need to be express in combination with other proteins for a better effect. Here, a multigene delivery system can fairly facilitate delivery of all desired therapeutic proteins. For example, Fang et al generated a rAAV encoding single ORF VEGFR2-neutralizing mAb DC101 (rAAV8-DC101) using a 2A self-processing system. rAAV is a preferred vector system when long-term gene expression is desired. rAAV vectors can stably transduce host cells and are capable of expressing therapeutic proteins at high constant levels in vivo following a single vector administration (Fang et al., 2005). Recently, Gianni-Barrera et al have made a retrovirus encoding single bicistronic construct co-expressing VEGF-A and PDGF-BB spaced apart by IRES sequence (VEGF-IRES-PDGFB). Co-delivery of VEGF-A and PDGF-BB promoted long-lasting and safe angiogenesis in a schemia mouse model. In this case, expression of VEGF and PDGFB were imbalanced due to using IRES sequence (Gianni-Barrera et al., 2016). Using a single 2A peptide–linked retroviral vector, Szymczak et al could restore T-cell development and function in CD3-deficient mice (Szymczak et al., 2004). This retroviral vector was containing of all four CD3 proteins (CD3 ε, γ, δ, ζ) of T-cell receptor:CD3 complex. Although 2A systems ensure equivalent level of all the domain but post-translational modifications (N- and C-terminal alternation) can effect on the functionality of the complex proteins.

Multiprotein expression system can be harnessed for vaccine development. Hoffmann et al developed an eight-plasmid DNA transfection system for the rescue of infectious influenza A virus (H6N1) from cloned cDNA (Hoffmann et al., 2000). In this plasmid-based co-expression system, eight plasmids harboring single cDNA transfected into co-cultured MDCK and 293T cells. This strategy profits from the fact that a replication competent virus is produced. It is therefore sufficient that only a small portion of the cells take up all eight plasmids. Production of virus like particles (VLPs) is another example than can show role of multigene delivery systems in vaccine development research. VLPs consisting of components of a virus without its genetic material can be used to produce safe vaccines (Roy and Noad). VLPs were shown to stimulate the immune response

and could replace attenuated viruses that are currently used for vaccination. Many of the VLPs synthesized for vaccine development have already entered the preclinical stage, or phase I, or were licensed as vaccines (Noad and Roy, 2003). Tang et al produced VLP though transduction of mammalian cells by baculovirus (Tang et al., 2011). In this BacMam system, three CMV-promoters drive the hemagglutinin, neuraminidase, and matrix of influenza virus was constructed. After transduction of HEK 293 cells with BacMam virus, influenza VLP secreted to the medium.

Targeted immunotherapy is another attractive subject that clearly depicting importance of multigene delivery in a potent therapy (Maude et al., 2015). In a chimeric antigen receptor engineering approach, T cell of patient are isolated and retarget toward cells with the targeted surface antigens (Gross et al., 1989). Three generation of CAR are available which in addition to extracellular domain co-express different intercellular domains: CD3ζ intracellular signaling domain of the T-cell receptor either alone (first generation) or in combination with 1 (second generation) or 2 (third generation) costimulatory domains (Maude et al., 2015).

For gene therapy, safety of multigene delivery vehicle has first priority. Some physical approach like in eutro gene delivery can be used to transfer genes to some specific tissues (e.g., brain) but it is limited to animal. Chemical- based multigene delivery usually suffer from toxicity although some progress using nanoparticles has reported. Using viral vectors due to possibility of virus integration to the host genome and beyond problem like induction of mutation causing cancer, is restricted. Non-integrating viral vectors that are able to deliver multigene constructs can be an option but more research in this area is needed.

Conclusion

Multigene delivery and subsequent cellular expression is a key technology for a wide range of applications in current biology. Depending on the mammalian cell type (established cell lines, primary cells or stem cells) many strategies and methods are available for delivery of multigene constructs albeit each of them

with their own merits. Depending to the interested experiments, generation of stable cell line or transient gene expression (TGE) system can be chosen. Although for production of recombinant proteins in large scale generation of stable cell line can be a good choice but it is anyway a tedious task and takes long time (weeks). TGE systems are easy to work and desired result can be prepared very fast. Anyway, some cells (like primary cell and stem cells) are hard to transfect. A variety of chemical and physical methods for transferring multigene cassettes to mammalian cells are available. Although these methods are easy to use but they have often toxic effect on the cells or have some limitations for in vivo applications. Viral- based methods, conversely, are more natural but production of virus usually is time consuming and raise some biosafety issues. Viral systems usually lead to high infection rates (up to 100%), but their capacity as carriers of foreign DNA is limited and the expression is usually transient.

To choose a suitable strategy for multigene assembly some parameters like, duration time for gene expression, availability of numerous antibiotics, modification of N- and C-terminal of proteins, stoichiometry ratio of genes of interest, homogeneity or heterogeneity of cell population and biosafety should be carefully considered. We have developed two multigene expression toolkits, MultiLabel and MultiPrime, to facilitate multigene expression for easy and hard cells to transfection. Both systems leading to homogenous cell population, no modification in N- and C-terminal of interested proteins, expression in desired levels and individually.

We have outlined in this review a broad range of applications that could potentially improve both research and production. Multigene delivery and subsequent expression could be instrumental for the production of important multiprotein assemblies such as antibodies or VLPs. Genome editing specially for in vivo application demand to a reliable system that enable delivery of Cas9 and other necessary components in a safe and amenable manner to the target cells. Synthetic biology would need to easier, or automated able multigene assembly techniques to accelerate creation of new devices for manipulation,

construction and re-organization of cell behaviors. Moreover, we anticipate that many researches in contemporary research and drug discovery applications, benefit from multigene delivery and expression in mammalian cells.

References:

Abil, Z., Xiong, X., and Zhao, H. (2015). Synthetic Biology for Therapeutic Applications. Mol. Pharm. *12*, 322–331.

Assenberg, R., Wan, P.T., Geisse, S., and Mayr, L.M. (2013a). Advances in recombinant protein expression for use in pharmaceutical research. Curr. Opin. Struct. Biol. *23*, 393–402.

Assenberg, R., Wan, P.T., Geisse, S., and Mayr, L.M. (2013b). Advances in recombinant protein expression for use in pharmaceutical research. Curr. Opin. Struct. Biol. *23*, 393–402.

Aumiller, J.J., Mabashi-Asazuma, H., Hillar, A., Shi, X., and Jarvis, D.L. (2012). A new glycoengineered insect cell line with an inducibly mammalianized protein N-glycosylation pathway. Glycobiology *22*, 417–428.

Ayuso, E. (2016). Manufacturing of recombinant adeno-associated viral vectors: new technologies are welcome. Mol. Ther. Methods Clin. Dev. *3*, 15049.

Backliwal, G., Hildinger, M., Chenuet, S., Wulhfard, S., De Jesus, M., and Wurm, F.M. (2008). Rational vector design and multi-pathway modulation of HEK 293E cells yield recombinant antibody titers exceeding 1 g/l by transient transfection under serum-free conditions. Nucleic Acids Res. *36*, e96.

Balboa, D., Weltner, J., Eurola, S., Trokovic, R., Wartiovaara, K., Otonkoski, T., Arda, H.E., Benitez, C.M., Kim, S.K., Baron, U., et al. (2015). Conditionally Stabilized dCas9 Activator for Controlling Gene Expression in Human Cell Reprogramming and Differentiation. Stem Cell Reports *5*, 448–459.

Ballmer-Hofer, K., Andersson, A.E., Ratcliffe, L.E., and Berger, P. (2011). Neuropilin-1 promotes VEGFR-2 trafficking through Rab11 vesicles thereby specifying signal output. Blood *118*, 816–826.

Barford, D., Takagi, Y., Schultz, P., and Berger, I. (2013). Baculovirus expression: tackling the complexity challenge. Curr. Opin. Struct. Biol. *23*, 357–364.

van Bergeijk, P., Adrian, M., Hoogenraad, C.C., and Kapitein, L.C. (2015). Optogenetic control of organelle transport and positioning. Nature *518*, 111–114.

Berger, I., and Mayr, L.M. (2013). Protein production for structural biology: new solutions to new challenges. Curr. Opin. Struct. Biol. *23*, 317–318.

Berger, P., Tersar, K., Ballmer-Hofer, K., and Suter, U. (2011). The CMT4B disease-causing proteins MTMR2 and MTMR13/SBF2 regulate AKT signalling. J. Cell. Mol. Med. *15*, 307–315.

Bieniossek, C., Imasaki, T., Takagi, Y., and Berger, I. (2012). MultiBac: expanding the research toolbox for multiprotein complexes. Trends Biochem. Sci. *37*, 49–57.

Brentjens, R.J., Rivière, I., Park, J.H., Davila, M.L., Wang, X., Stefanski, J., Taylor, C., Yeh, R., Bartido, S., Borquez-Ojeda, O., et al. (2011). Safety and persistence of adoptively transferred autologous CD19-targeted T cells in patients with relapsed or chemotherapy refractory B-cell leukemias. Blood *118*, 4817–4828.

Caiazzo, M., Giannelli, S., Valente, P., Lignani, G., Carissimo, A., Sessa, A., Colasante, G., Bartolomeo, R., Massimino, L., Ferroni, S., et al. (2015). Direct Conversion of Fibroblasts into Functional Astrocytes by Defined Transcription Factors. Stem Cell Reports *4*, 25–36.

Carey, B.W., Markoulaki, S., Hanna, J., Saha, K., Gao, Q., Mitalipova, M., and Jaenisch, R. (2009). Reprogramming of murine and human somatic cells using a single polycistronic vector. Proc. Natl. Acad. Sci. U. S. A. *106*, 157–162.

Chakraborty, S., Ji, H., Kabadi, A.M., Gersbach, C.A., Christoforou, N., Leong, K.W., Bikard, D., Jiang, W., Samai, P., Hochschild, A., et al. (2014). A CRISPR/Cas9-Based System for Reprogramming Cell Lineage Specification. Stem Cell Reports *3*, 940–947.

Chavez, A., Scheiman, J., Vora, S., Pruitt, B.W., Tuttle, M., P R Iyer, E., Lin, S., Kiani, S., Guzman, C.D., Wiegand, D.J., et al. (2015). Highly efficient Cas9-mediated transcriptional programming. Nat. Methods *12*, 326–328.

Chen, X., Pham, E., and Truong, K. (2010). TEV protease-facilitated stoichiometric delivery of multiple genes using a single expression vector. Protein Sci. *19*, 2379–2388.

Chen, Y., Cao, J., Xiong, M., Petersen, A.J., Dong, Y., Tao, Y., Huang, C.T.-L., Du, Z., Zhang, S.-C., Acampora, D., et al. (2015). Engineering Human Stem Cell Lines with Inducible Gene Knockout using CRISPR/Cas9. Cell Stem Cell *17*, 233–244.

Chernajovsky, Y., Gould, D.J., and Podhajcer, O.L. (2004). Gene therapy for autoimmune diseases: quo vadis? Nat. Rev. Immunol. *4*, 800–811.

Coleman, J.A., Green, E.M., and Gouaux, E. (2016). X-ray structures and mechanism of the human serotonin transporter. Nature *532*, 334–339.

Cong, L., Ran, F.A., Cox, D., Lin, S., Barretto, R., Habib, N., Hsu, P.D., Wu, X., Jiang, W., Marraffini, L.A., et al. (2013). Multiplex genome engineering using CRISPR/Cas systems. Science *339*, 819–823.

Deans, T.L., Cantor, C.R., and Collins, J.J. (2007). A Tunable Genetic Switch Based on RNAi and Repressor Proteins for Regulating Gene Expression in Mammalian Cells. Cell *130*, 363–372.

Dietrich, J.-E., Panavaite, L., Gunther, S., Wennekamp, S., Groner, A.C., Pigge, A., Salvenmoser, S., Trono, D., Hufnagel, L., and Hiiragi, T. (2015). Venus trap in the mouse embryo reveals distinct molecular dynamics underlying specification of first embryonic lineages. EMBO Rep. *16*, 1005–1021.

Dimos, J.T., Rodolfa, K.T., Niakan, K.K., Weisenthal, L.M., Mitsumoto, H., Chung, W., Croft, G.F., Saphier, G., Leibel, R., Goland, R., et al. (2008). Induced pluripotent stem cells generated from patients with ALS can be differentiated into motor neurons. Science *321*, 1218–1221.

Dominguez, A.A., Lim, W.A., and Qi, L.S. (2015). Beyond editing: repurposing CRISPR–Cas9 for precision genome regulation and interrogation. Nat. Rev. Mol. Cell Biol. *17*, 5–15.

Eglen, R., and Reisine, T. (2011). Primary Cells and Stem Cells in Drug Discovery: Emerging Tools for High-Throughput Screening. Assay Drug Dev. Technol. *9*, 108–124.

Engler, C., Kandzia, R., and Marillonnet, S. (2008). A One Pot, One Step,

Precision Cloning Method with High Throughput Capability. PLoS One *3*, e3647.

Engler, C., Gruetzner, R., Kandzia, R., and Marillonnet, S. (2009). Golden Gate Shuffling: A One-Pot DNA Shuffling Method Based on Type IIs Restriction Enzymes. PLoS One *4*, e5553.

Fang, J., Qian, J.-J., Yi, S., Harding, T.C., Tu, G.H., VanRoey, M., and Jooss, K. (2005). Stable antibody expression at therapeutic levels using the 2A peptide. Nat. Biotechnol. *23*, 584–590.

Farzadfard, F., Perli, S.D., and Lu, T.K. (2013). Tunable and Multifunctional Eukaryotic Transcription Factors Based on CRISPR/Cas. ACS Synth. Biol. *2*, 604–613.

de Felipe, P. (2004). Skipping the co-expression problem: the new 2A "CHYSEL" technology. Genet. Vaccines Ther. *2*, 13.

Fitzgerald, D.J., Berger, P., Schaffitzel, C., Yamada, K., Richmond, T.J., and Berger, I. (2006). Protein complex expression by using multigene baculoviral vectors. Nat. Methods *3*, 1021–1032.

Gerl, M.J., Bittl, V., Kirchner, S., Sachsenheimer, T., Brunner, H.L., Lüchtenborg, C., Özbalci, C., Wiedemann, H., Wegehingel, S., Nickel, W., et al. (2016). Sphingosine-1-Phosphate Lyase Deficient Cells as a Tool to Study Protein Lipid Interactions. PLoS One *11*, e0153009.

Gianni-Barrera, R., Burger, M., Wolff, T., Heberer, M., Schaefer, D.J., Gürke, L., Mujagic, E., Banfi, A., Mozaffarian, D., Potente, M., et al. (2016). Long-term safety and stability of angiogenesis induced by balanced single-vector co-expression of PDGF-BB and VEGF164 in skeletal muscle. Sci. Rep. *6*, 21546.

Gibson, D.G., Young, L., Chuang, R.-Y., Venter, J.C., Hutchison, C.A., and Smith, H.O. (2009). Enzymatic assembly of DNA molecules up to several hundred kilobases. Nat. Methods *6*, 343–345.

Gilbert, L.A., Larson, M.H., Morsut, L., Liu, Z., Brar, G.A., Torres, S.E., Stern-Ginossar, N., Brandman, O., Whitehead, E.H., Doudna, J.A., et al. (2013). CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes. Cell *154*, 442–451.

Gilbert, L.A., Horlbeck, M.A., Adamson, B., Villalta, J.E., Chen, Y., Whitehead, E.H., Guimaraes, C., Panning, B., Ploegh, H.L., Bassik, M.C., et al. (2014). Genome-Scale CRISPR-Mediated Control of Gene Repression and Activation. Cell *159*, 647–661.

Goehring, A., Lee, C.-H., Wang, K.H., Michel, J.C., Claxton, D.P., Baconguis, I., Althoff, T., Fischer, S., Garcia, K.C., and Gouaux, E. (2014). Screening and large-scale expression of membrane proteins in mammalian cells for structural studies. Nat. Protoc. *9*, 2574–2585.

Gonzalez, F., Barragan Monasterio, M., Tiscornia, G., Montserrat Pulido, N., Vassena, R., Batlle Morera, L., Rodriguez Piza, I., and Izpisua Belmonte, J.C. (2009). Generation of mouse-induced pluripotent stem cells by transient expression of a single nonviral polycistronic vector. Proc. Natl. Acad. Sci. U. S. A. *106*, 8918–8922.

Greber, D., and Fussenegger, M. (2007). Multi-gene engineering: Simultaneous expression and knockdown of six genes off a single platform. Biotechnol. Bioeng. *96*, 821–834.

Grimm, D., Kern, A., Rittner, K., and Kleinschmidt, J.A. (1998). Novel tools for production and purification of recombinant adenoassociated virus vectors. Hum. Gene Ther. *9*, 2745–2760.

Gross, G., Waks, T., and Eshhar, Z. (1989). Expression of immunoglobulin-T-cell receptor chimeric molecules as functional receptors with antibody-type specificity (chimeric genes/antibody variable region). Immunology *86*, 10024–10028.

Guye, P., Li, Y., Wroblewska, L., Duportet, X., and Weiss, R. (2013). Rapid, modular and reliable construction of complex mammalian gene circuits. Nucleic Acids Res. *41*, e156.

Harrison, R.L., and Jarvis, D.L. (2006). Protein N-glycosylation in the baculovirus-insect cell expression system and engineering of insect cells to produce "mammalianized" recombinant glycoproteins. Adv. Virus Res. *68*, 159–191.

Heiser, W.C. (2003). Gene Delivery to Mammalian Cells (New Jersey: Humana Press).

Heng, B.C., Aubel, D., and Fussenegger, M. (2014). G protein-coupled receptors revisited: therapeutic applications inspired by synthetic biology. Annu. Rev. Pharmacol. Toxicol. *54*, 227–249.

Hida, N., Awais, M., Takeuchi, M., Ueno, N., Tashiro, M., Takagi, C., Singh, T., Hayashi, M., Ohmiya, Y., and Ozawa, T. (2009). High-sensitivity real-time imaging of dual protein-protein interactions in living subjects using multicolor luciferases. PLoS One *4*, e5868.

Hilton, I.B., D'Ippolito, A.M., Vockley, C.M., Thakore, P.I., Crawford, G.E., Reddy, T.E., and Gersbach, C.A. (2015). Epigenome editing by a CRISPR-Cas9-based acetyltransferase activates genes from promoters and enhancers. Nat. Biotechnol. *33*, 510–517.

Ho, S.C.L., Bardor, M., Li, B., Lee, J.J., Song, Z., Tong, Y.W., Goh, L.-T., Yang, Y., Aggarwal, S., Nelson, A., et al. (2013). Comparison of Internal Ribosome Entry Site (IRES) and Furin-2A (F2A) for Monoclonal Antibody Expression Level and Quality in CHO Cells. PLoS One *8*, e63247.

Hockemeyer, D., Soldner, F., Cook, E.G., Gao, Q., Mitalipova, M., and Jaenisch, R. (2008). A drug-inducible system for direct reprogramming of human somatic cells to pluripotency. Cell Stem Cell *3*, 346–353.

Hoffmann, E., Neumann, G., Kawaoka, Y., Hobom, G., and Webster, R.G. (2000). A DNA transfection system for generation of influenza A virus from eight plasmids. Proc. Natl. Acad. Sci. U. S. A. *97*, 6108–6113.

Hotta, A., and Yamanaka, S. (2015). From Genomics to Gene Therapy : Induced Pluripotent Stem Cells Meet Genome Editing. 1–24.

Hsu, P.D., Lander, E.S., Zhang, F., Barrangou, R., Oost, J. van der, Barrangou, R., Fremaux, C., Deveau, H., Richards, M., Boyaval, P., et al. (2014). Development and Applications of CRISPR-Cas9 for Genome Engineering. Cell *157*, 1262–1278.

James, J.R., and Vale, R.D. (2012). Biophysical mechanism of T-cell receptor triggering in a reconstituted system. Nature *487*, 64.

Jenkins, N., Murphy, L., and Tyther, R. (2008). Post-translational modifications of recombinant proteins: significance for biopharmaceuticals. Mol. Biotechnol. *39*, 113–118.

Jiang, W., and Marraffini, L. a. (2015a). CRISPR-Cas: New Tools for Genetic Manipulations from Bacterial Immunity Systems. Annu. Rev. Microbiol. *69*, 150724172101001.

Jiang, W., and Marraffini, L.A. (2015b). CRISPR-Cas: New Tools for Genetic Manipulations from Bacterial Immunity Systems. Annu. Rev. Microbiol. *69*, 209–228.

Jopling, C., Boue, S., and Izpisua Belmonte, J.C. (2011). Dedifferentiation, transdifferentiation and reprogramming: three routes to regeneration. Nat. Rev. Mol. Cell Biol. *12*, 79–89.

Kabadi, A.M., Ousterout, D.G., Hilton, I.B., and Gersbach, C.A. (2014). Multiplex CRISPR/Cas9-based genome engineering from a single lentiviral vector. Nucleic Acids Res. *42*, e147.

Kaláb, P., and Soderholm, J. (2010). The design of Förster (fluorescence) resonance energy transfer (FRET)-based molecular sensors for Ran GTPase. Methods *51*, 220–232.

Kearns, N.A., Pham, H., Tabak, B., Genga, R.M., Silverstein, N.J., Garber, M., and Maehr, R. (2015). Functional annotation of native enhancers with a Cas9–histone demethylase fusion. Nat. Methods *12*, 401–403.

Knutson, J.C., and Yee, D. (1987). Electroporation: parameters affecting transfer of DNA into mammalian cells. Anal. Biochem. *164*, 44–52.

Kost, T.A., Condreay, J.P., and Ames, R.S. (2010). Baculovirus gene delivery: a flexible assay development tool. Curr. Gene Ther. *10*, 168–173.

Kotin, R.M. (2011). Large-scale recombinant adeno-associated virus production. Hum. Mol. Genet. *20*, R2–R6.

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. Nat. Commun. *1*, 120.

Li, C., Schwabe, J.W., Banayo, E., and Evans, R.M. (1997). Coexpression of nuclear receptor partners increases their solubility and biological activities. Proc. Natl. Acad. Sci. U. S. A. *94*, 2278–2283.

Lienert, F., Lohmueller, J.J., Garg, A., and Silver, P.A. (2014). Synthetic biology in mammalian cells: next generation research tools and therapeutics. Nat. Rev. Mol. Cell Biol. *15*, 95–107.

Lujan, E., and Wernig, M. (2013). An indirect approach to generating specific human cell types. Nat. Methods *10*, 44–45.

Luke, G.A., and Ryan, M.D. (2013). The protein coexpression problem in biotechnology and biomedicine: virus 2A and 2A-like sequences provide a solution. Future Virol. *8*, 983–996.

Luker, K.E., and Luker, G.D. (2016). Dual-Color Luciferase Complementation for Chemokine Receptor Signaling. Methods Enzymol. *570*, 119–129.

Maeder, M.L., Linder, S.J., Cascio, V.M., Fu, Y., Ho, Q.H., and Joung, J.K. (2013). CRISPR RNA–guided activation of endogenous human genes. Nat. Methods *10*, 977–979.

Maherali, N., Ahfeldt, T., Rigamonti, A., Utikal, J., Cowan, C., and Hochedlinger, K. (2008). A high-efficiency system for the generation and study of human induced pluripotent stem cells. Cell Stem Cell *3*, 340–345.

Mali, P., Yang, L., Esvelt, K.M., Aach, J., Guell, M., DiCarlo, J.E., Norville, J.E., and Church, G.M. (2013). RNA-guided human genome engineering via Cas9. Science *339*, 823–826.

Mansouri, M., and Berger, P. (2014). Strategies for multigene expression in eukaryotic cells. Plasmid *75*, 12–17.

Mansouri, M., Bellon-Echeverria, I., Rizk, A., Ehsaei, Z., Cianciolo Cosentino, C., Silva, C.S., Xie, Y., Boyce, F.M., Davis, M.W., Neuhauss, S.C.F., et al. (2016). Highly efficient baculovirus-mediated multigene delivery in primary cells. Nat. Commun. *7*, 11529.

Martello, G., and Smith, A. (2014). The Nature of Embryonic Stem Cells. Annu. Rev. Cell Dev. Biol. *30*, 647–675.

Martínez-Alonso, M., García-Fruitós, E., Ferrer-Miralles, N., Rinas, U., and Villaverde, A. (2010). Side effects of chaperone gene co-expression in recombinant protein production. Microb. Cell Fact. *9*, 64.

Maude, S.L., Teachey, D.T., Porter, D.L., and Grupp, S.A. (2015). CD19-targeted chimeric antigen receptor T-cell therapy for acute lymphoblastic leukemia. Blood *125*.

Miller, K.E., Kim, Y., Huh, W.-K., and Park, H.-O. (2015). Bimolecular Fluorescence Complementation (BiFC) Analysis: Advances and Recent Applications for Genome-Wide Interaction Studies. J. Mol. Biol. *427*, 2039–2055.

Morsut, L., Roybal, K.T., Xiong, X., Gordley, R.M., Coyle, S.M., Thomson, M., Lim, W.A., Artavanis-Tsakonas, S., Rand, M.D., Lake, R.J., et al. (2016). Engineering Customized Cell Sensing and Response Behaviors Using Synthetic Notch Receptors. Cell *164*, 780–791.

Müller, K., and Weber, W. (2013). Optogenetic tools for mammalian systems. Mol. BioSyst. Mol. BioSyst *9*, 596–608.

Nakagawa, M., Koyanagi, M., Tanabe, K., Takahashi, K., Ichisaka, T., Aoi, T., Okita, K., Mochiduki, Y., Takizawa, N., and Yamanaka, S. (2007). Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts. Nat. Biotechnol. *26*, 101–106.

Nelson, C.E., and Gersbach, C.A. (2016). Engineering Delivery Vehicles for Genome Editing. Annu. Rev. Chem. Biomol. Eng. *7*, annurev – chembioeng – 080615–034711.

Nelson, T.J., and Terzic, A. (2009). Induced pluripotent stem cells: reprogrammed without a trace. Regen. Med. *4*, 333–335.

Nihongaki, Y., Yamamoto, S., Kawano, F., Suzuki, H., Sato, M., Brieke, C., Rohrbach, F., Gottschalk, A., Mayer, G., Heckel, A., et al. (2015). CRISPR-Cas9-based Photoactivatable Transcription System. Chem. Biol. *22*, 169–174.

Noad, R., and Roy, P. (2003). Virus-like particles as immunogens. Trends Microbiol. *11*, 438–444.

Okita, K., Nakagawa, M., Hyenjong, H., Ichisaka, T., and Yamanaka, S. (2008). Generation of mouse induced pluripotent stem cells without viral vectors. Science *322*, 949–953.

Packard, A.I., Lin, B., and Schwob, J.E. Sox2 and Pax6 Play Counteracting Roles in Regulating Neurogenesis within the Murine Olfactory Epithelium. PLoS One *11*, e0155167.

Palmberger, D., Wilson, I.B.H., Berger, I., Grabherr, R., and Rendic, D. (2012). SweetBac: a new approach for the production of mammalianised glycoproteins in insect cells. PLoS One *7*, e34226.

Parsons, M.J., Fassio, S.R., and Bouchier-Hayes, L. (2016). Detection of Initiator Caspase Induced Proximity in Single Cells by Caspase Bimolecular Fluorescence Complementation. Methods Mol. Biol. *1419*, 41–56.

Perez-Pinera, P., Kocak, D.D., Vockley, C.M., Adler, A.F., Kabadi, A.M., Polstein, L.R., Thakore, P.I., Glass, K.A., Ousterout, D.G., Leong, K.W., et al. (2013). RNA-guided gene activation by CRISPR-Cas9–based transcription factors. Nat. Methods *10*, 973–976.

Platt, R.J., Chen, S., Zhou, Y., Yim, M.J., Swiech, L., Kempton, H.R., Dahlman, J.E., Parnas, O., Eisenhaure, T.M., Jovanovic, M., et al. (2014a). CRISPR-Cas9 knockin mice for genome editing and cancer modeling. Cell *159*, 440–455.

Platt, R.J., Chen, S., Zhou, Y., Yim, M.J., Swiech, L., Kempton, H.R., Dahlman, J.E., Parnas, O., Eisenhaure, T.M., Jovanovic, M., et al. (2014b). CRISPR-Cas9 Knockin Mice for Genome Editing and Cancer Modeling. Cell *159*, 440–455.

Polstein, L.R., and Gersbach, C.A. (2015). A light-inducible CRISPR-Cas9 system for control of endogenous gene activation. Nat. Chem. Biol. *11*, 198–200.

Porter, D.L., Levine, B.L., Kalos, M., Bagg, A., and June, C.H. (2011). Chimeric Antigen Receptor–Modified T Cells in Chronic Lymphoid Leukemia. N. Engl. J. Med. *365*, 725–733.

Qi, L.S., Larson, M.H., Gilbert, L.A., Doudna, J.A., Weissman, J.S., Arkin, A.P., and Lim, W.A. (2013). Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression. Cell *152*, 1173–1183.

Rizk, A., Paul, G., Incardona, P., Bugarski, M., Mansouri, M., Niemann, A., Ziegler, U., Berger, P., and Sbalzarini, I.F. (2014). Segmentation and quantification of subcellular structures in fluorescence microscopy images using Squassh. Nat. Protoc. *9*, 586–596.

Rizk, A., Mansouri, M., Ballmer-Hofer, K., and Berger, P. (2015). Subcellular object quantification with Squassh3C and SquasshAnalyst. Biotechniques.

Robinson, C. V., Sali, A., and Baumeister, W. (2007). The molecular sociology of the cell. Nature *450*, 973–982.

Roy, P., and Noad, R. Virus-like particles as a vaccine delivery system: myths and facts. Hum. Vaccin. *4*, 5–12.

Sakuma, T., and Yamamoto, T. (2015). CRISPR/Cas9: The Leading Edge of Genome Editing Technology. In Targeted Genome Editing Using Site-Specific Nucleases, (Tokyo: Springer Japan), pp. 25–41.

Sánchez-Rivera, F.J., and Jacks, T. (2015). Applications of the CRISPR–Cas9 system in cancer biology. Nat. Rev. Cancer *15*, 387–395.

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

Schlaeger, T.M., Daheron, L., Brickler, T.R., Entwisle, S., Chan, K., Cianci, A., DeVine, A., Ettenger, A., Fitzgerald, K., Godfrey, M., et al. (2014). A comparison of non-integrating reprogramming methods. Nat. Biotechnol. *33*, 58–63.

Schwank, G., Koo, B.-K., Sasselli, V., Dekkers, J.F., Heo, I., Demircan, T., Sasaki, N., Boymans, S., Cuppen, E., van der Ent, C.K., et al. (2013). Functional Repair of CFTR by CRISPR/Cas9 in Intestinal Stem Cell Organoids of Cystic Fibrosis Patients. Cell Stem Cell *13*, 653–658.

Seto, J.T., Ramos, J.N., Muir, L., Chamberlain, J.S., and Odom, G.L. (2012). Gene replacement therapies for duchenne muscular dystrophy using adeno-associated viral vectors. Curr. Gene Ther. *12*, 139–151.

Shao, L., and Wu, W.-S. (2010). Gene-delivery systems for iPS cell generation. Expert Opin. Biol. Ther. *10*, 231–242.

Smith, A.E. (1995). Viral Vectors in Gene Therapy. Annu. Rev. Microbiol. *49*, 807–838.

Sommer, C.A., Stadtfeld, M., Murphy, G.J., Hochedlinger, K., Kotton, D.N., and Mostoslavsky, G. (2009). Induced pluripotent stem cell generation using a single lentiviral stem cell cassette. Stem Cells *27*, 543–549.

Stadtfeld, M., Nagaya, M., Utikal, J., Weir, G., and Hochedlinger, K. (2008). Induced pluripotent stem cells generated without viral integration. Science *322*, 945–949.

Szymczak, A.L., Workman, C.J., Wang, Y., Vignali, K.M., Dilioglou, S., Vanin, E.F., and Vignali, D.A.A. (2004). Correction of multi-gene deficiency in vivo using a single “self-cleaving” 2A peptide-based retroviral vector. Nat. Biotechnol. *22*, 589–594.

Takahashi, K., and Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. Cell *126*, 663–676.

Takata, Y., Kishine, H., Sone, T., Andoh, T., Nozaki, M., Poderycki, M., Chesnut, J.D., and Imamoto, F. (2011). Generation of iPS Cells Using a BacMam Multigene Expression System. Cell Struct. Funct. *36*, 209–222.

Tanenbaum, M.E., Gilbert, L.A., Qi, L.S., Weissman, J.S., and Vale, R.D. (2014). A protein-tagging system for signal amplification in gene expression and fluorescence imaging. Cell *159*, 635–646.

Tang, X.-C., Lu, H.-R., and Ross, T.M. (2011). Baculovirus-produced influenza virus-like particles in mammalian cells protect mice from lethal influenza challenge. Viral Immunol. *24*, 311–319.

Tischer, D., and Weiner, O.D. (2014). Illuminating cell signalling with optogenetic tools. Nat. Rev. Mol. Cell Biol. *15*, 551–558.

Tolia, N.H., and Joshua-Tor, L. (2006). Strategies for protein coexpression in Escherichia coli. Nat. Methods *3*, 55–64.

Trowitzsch, S., Klumpp, M., Thoma, R., Carralot, J.-P., and Berger, I. (2011). Light it up: highly efficient multigene delivery in mammalian cells. Bioessays *33*, 946–955.

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.

Walsh, G., and Jefferis, R. (2006). Post-translational modifications in the context of therapeutic proteins. Nat. Biotechnol. *24*, 1241–1252.

Wang, T., Ma, X., Zhu, H., Li, A., Du, G., and Chen, J. (2012). Available methods for assembling expression cassettes for synthetic biology. Appl. Microbiol. Biotechnol. *93*, 1853–1863.

Wang, Y.-H., Wei, K.Y., and Smolke, C.D. (2013). Synthetic biology: advancing the design of diverse genetic systems. Annu. Rev. Chem. Biomol. Eng. *4*, 69–102.

Wapinski, O.L., Vierbuchen, T., Qu, K., Lee, Q.Y., Chanda, S., Fuentes, D.R., Giresi, P.G., Ng, Y.H., Marro, S., Neff, N.F., et al. (2013). XHierarchical mechanisms for direct reprogramming of fibroblasts to neurons. Cell *155*, 621–635.

Way, J.C., Collins, J.J., Keasling, J.D., and Silver, P.A. (2014). Integrating biological redesign: where synthetic biology came from and where it needs to go. Cell *157*, 151–161.

Weissman, T.A., and Pan, Y.A. (2015). Brainbow: new resources and emerging biological applications for multicolor genetic labeling and analysis. Genetics *199*, 293–306.

Werner, S., Engler, C., Weber, E., Gruetzner, R., and Marillonnet, S. (2012). Fast track assembly of multigene constructs using golden gate cloning and the MoClo system. Bioeng. Bugs *3*, 38–43.

Wu, J., Hunt, S.D., Xue, H., Liu, Y., and Darabi, R. (2016). Generation and validation of PAX7 reporter lines from human iPS cells using CRISPR/Cas9 technology. Stem Cell Res. *16*, 220–228.

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 (80-.). *333*.

Yamamoto, T. (2015). Targeted genome editing using site-specific nucleases: ZFNs, TALENs, and the CRISPR/Cas9 system.

Yan, X., Hu, S., Guan, Y.-X., and Yao, S.-J. (2012). Coexpression of chaperonin GroEL/GroES markedly enhanced soluble and functional expression of recombinant human interferon-gamma in Escherichia coli. Appl. Microbiol. Biotechnol. *93*, 1065–1074.

Yates, J.L., Warren, N., and Sugden, B. (1985). Stable replication of plasmids derived from Epstein–Barr virus in various mammalian cells. Nature *313*, 812–815.

Ye, H., and Fussenegger, M. (2014). Synthetic therapeutic gene circuits in mammalian cells. FEBS Lett. *588*, 2537–2544.

Ye, L., Wang, J., Beyer, A.I., Teque, F., Cradick, T.J., Qi, Z., Chang, J.C., Bao, G., Muench, M.O., Yu, J., et al. (2014). Seamless modification of wild-type induced pluripotent stem cells to the natural CCR5Δ32 mutation confers resistance to HIV infection. Proc. Natl. Acad. Sci. U. S. A. *111*, 9591–9596.

Yu, J., Vodyanik, M.A., Smuga-Otto, K., Antosiewicz-Bourget, J., Frane, J.L., Tian, S., Nie, J., Jonsdottir, G.A., Ruotti, V., Stewart, R., et al. (2007). Induced pluripotent stem cell lines derived from human somatic cells. Science *318*, 1917–1920.

Yu, J., Hu, K., Smuga-Otto, K., Tian, S., Stewart, R., Slukvin, I.I., and Thomson, J.A. (2009). Human induced pluripotent stem cells free of vector and transgene sequences. Science *324*, 797–801.

Zalatan, J.G., Lee, M.E., Almeida, R., Gilbert, L.A., Whitehead, E.H., La Russa, M., Tsai, J.C., Weissman, J.S., Dueber, J.E., Qi, L.S., et al. (2015). Engineering complex synthetic transcriptional programs with CRISPR RNA scaffolds. Cell *160*, 339–350.

Zhu, J., Musco, M.L., and Grace, M.J. (1999). Three-color flow cytometry analysis of tricistronic expression of eBFP, eGFP, and eYFP using EMCV-IRES linkages. Cytometry *37*, 51–59.

An Introduction to Molecular Developmental Biology.

Table 1. comparison between viral gene delivery systems

	Baculovirus	lentivirus	Adenovirus	Adeno-associated virus (AAV)	Retrovirus
Infection of dividing cells	Yes	Yes	Yes	Yes	Yes
Infection of non-dividing cells	Yes	Yes	Yes	Yes	No
Genome integration	No	Yes	No	No	Yes
Size of the wild type genome	134 kb	9.7 kb	37.7	4.7 kb	8.3 kb
Biosafety level	BSL1	BSL2	BSL2	BSL1	BSL2
Cargo capacity	Up to 50 kb	2.5-5.0 Kb	3.0-8.0 Kb	2.5 kb	2.5-5.0 Kb
Packaging cells	Insect cells/27 ^{o}C	HEK293T/37^{o}C	HEK293T/37^{o}C	HEK293T/37^{o}C	HEK293T/37^{o}C

Table 2. Advantages and disadvantages of different strategies for multigene delivery to mammalian cells

strategy	technique	advantages	disadvantages
Co-transfection/ co-infection	*Several plasmids or virus bearing individual gene	* Fast and straightforward	* Generation of heterogonous cell population * Different level of gene expression * Difficult to analysis the results
IRES	*Single transcript with multiple sites for translation initiation	* Co expression of all genes * Ideal for co expression of markers or antibiotic resistance genes	* Imbalanced level of expression * Different activity in different cells
polyprotein	* Protease/proteolytic processing * Reinitiation * 2A and ribosomal skipping	*Ensured co-expression *Ideal for multigene expression *A variety of sequences is available	* N- or C-terminal modification of proteins of interest
Direct assembly	* Multiple transcription units with their own promoters	* Flexible	* Consider many items for cloning design * Restricted to certain enzymes * Decrease efficiency by increasing number of genes
MultiLabel	* Multiple expression cassettes in one single plasmid	*Every gene has its own promoter *ideal for multigene expression	* Large size * Restricted to transfection approach * Low efficiency in primary and stem cells
MultiPrime	* Multiple expression cassettes in one single baculovirus	*Every gene has its own promoter *Ideal for multigene expression *High efficiency in primary and stem cells	* Rearrangement of virus during over amplification * Need to insect cell culture facilities

CHAPTER 2

MULTIPRIME UNLOCKS MULTIGENE DELIVERY TO PRIMARY MAMMALIAN CELLS

For certain cell types including widely used HEK293 and HeLa cells, multigene delivery can be achieved by plasmid-based transfection. However, a large number of cell lines and particularly primary cells are recalcitrant to plasmid transfection, thus requiring a different approach. Infection by viral vectors as an efficient way can be an alternative. However, most of the common viral vectors (Lentivirus, Retroviruses, Adenovirus, AAV) suffer from one or more of the following disadvantages: inability to infect non-dividing cells, small transgene packaging capacity, and risks associated with insertional mutagenesis. Baculovirus, which normally infects insect cells, can be used to overcome these bottlenecks. Therefore, I developed the *MultiPrime* system, which is a baculovirus-mediated multigene delivery system for mammalian cells (Mansouri et al., 2016).

This chapter contains my original paper which was recently published. In the introduction part, I explain the importance, concept and workflow of the MultiPrime system. In the result section, I demonstrate possibilities of MultiPrime for simultaneous delivery of multiple genes to a variety of cells including cell lines, primary cells and stem cells. Also, I show different applications of the MultiPrime system for example for reprogramming fibroblasts into neurons, for producing functional antibodies and for CRISPR/Cas9-mediated genome engineering.

nature COMMUNICATIONS

ARTICLE

Received 17 Feb 2016 | Accepted 4 Apr 2016 | Published 4 May 2016

DOI: 10.1038/ncomms11529 OPEN

Highly efficient baculovirus-mediated multigene delivery in primary cells

Maysam Mansouri[1], Itxaso Bellon-Echeverria[2], Aurélien Rizk[1], Zahra Ehsaei[3], Chiara Cianciolo Cosentino[4], Catarina S. Silva[2], Ye Xie[1], Frederick M. Boyce[5], M. Wayne Davis[6], Stephan C.F. Neuhauss[4], Verdon Taylor[3], Kurt Ballmer-Hofer[1], Imre Berger[2,7] & Philipp Berger[1]

Multigene delivery and subsequent cellular expression is emerging as a key technology required in diverse research fields including, synthetic and structural biology, cellular reprogramming and functional pharmaceutical screening. Current viral delivery systems such as retro- and adenoviruses suffer from limited DNA cargo capacity, thus impeding unrestricted multigene expression. We developed MultiPrime, a modular, non-cytotoxic, non-integrating, baculovirus-based vector system expediting highly efficient transient multigene expression from a variety of promoters. MultiPrime viruses efficiently transduce a wide range of cell types, including non-dividing primary neurons and induced-pluripotent stem cells (iPS). We show that MultiPrime can be used for reprogramming, and for genome editing and engineering by CRISPR/Cas9. Moreover, we implemented dual-host-specific cassettes enabling multiprotein expression in insect and mammalian cells using a single reagent. Our experiments establish MultiPrime as a powerful and highly efficient tool, to deliver multiple genes for a wide range of applications in primary and established mammalian cells.

[1] Biomolecular Research, Molecular Cell Biology, Paul Scherrer Institute, CH-5232 Villigen, Switzerland. [2] European Molecular Biology Laboratory (EMBL), Grenoble Outstation, B.P. 181, 38042 Grenoble Cedex 9, France. [3] Department of Biomedicine, University of Basel, CH-4058 Basel, Switzerland. [4] Institute of Molecular Life Sciences, University of Zürich, CH-8057 Zürich, Switzerland. [5] Department of Neurology, Massachusetts General Hospital, Cambridge, Massachusetts 02139, USA. [6] Department of Biology and Howard Hughes Medical Institute, University of Utah, Salt Lake City, Utah 84112-0840, USA. [7] School of Biochemistry, University of Bristol, Bristol BS8 1TD, UK. Correspondence and requests for materials should be addressed to I.B. (email: imre.berger@bristol.ac.uk) or to P.B. (email: Philipp.Berger@psi.ch).

Multigene delivery into cultured cells or tissues is emerging as an indispensable tool for many applications in biological research and development. Examples include simultaneous labelling of living cells with various fluorescently-tagged sensors for monitoring changes in cellular architecture or metabolism, lineage tracing during morphogenesis to follow regenerative tissue processes, visualization of multi-component molecular pathways for high-content screening in pharmacological applications or the construction of recombinant adeno-associated viruses for gene therapy[1–5]. Multigene delivery systems also allow reprogramming of somatic cells to stem cells[6] or to specifically differentiated cell lines[7]. The construction of complex multigene circuits in mammalian cells is a core concept in synthetic biology requiring the flexible generation of modular multigene expression systems[8,9]. Moreover, structural and biophysical characterization of multiprotein complexes relies on co-expression of an ensemble of genes that may include ancillary factors, such as chaperones or protein modifying enzymes[10]. All applications share in common that they require versatile tool-kits to flexibly engineer and to simultaneously, efficiently and reproducibly deliver multiple genes into target host cells.

Several strategies for multigene expression in mammalian cells exist, each with its own merits[11]. All of these applications require specific boundary conditions. For instance, it is essential that all transfected cells in a population express all heterologous genes at the same defined level, on an equal time frame. Other applications require that the proteins of interest retain native N- or C termini. Furthermore, long-term stable expression versus transient expression is a crucial parameter to be considered. Ideally, an efficient multigene-delivery system would provide the means to afford many or all of these requirements.

We have developed systems for the delivery of multigene constructs in prokaryotic and eukaryotic hosts[12–14]. A central feature of these technologies is the assembly of multiple gene expression cassettes by recombineering[15], from custom designed plasmids encoding specific genes, into a single multicomponent DNA construct for gene delivery. This approach was shown to overcome the limitations hampering classical co-transfection or co-infection techniques, which for statistical reasons, are inherently unbalanced[16,17]. More recently, we introduced MultiLabel[14] and demonstrated that homogenous mammalian cell populations could be achieved by transient introduction of single recombineering-based multigene expression plasmids by classical transfection methods. This method performs well with cell lines that are readily transfected, such as HEK293 or HeLa cells. However, a large number of cell lines and particularly primary cells are markedly recalcitrant to plasmid transfection, thus requiring a different approach. Primary cells are a central focus of contemporary biological research efforts, and efficient multigene delivery in primary cells is thus highly desirable.

Infection by viral vectors emerged as the dominant method of choice to deliver genes into primary cells[18]. An ideal viral vector for multigene delivery should have virtually unlimited foreign DNA cargo capacity allowing for integration of a multitude of independent expression cassettes, functionalities and regulatory elements. Moreover, such an optimal viral vector should exhibit low cytotoxicity in mammalian cells and should enable transduction of dividing and non-dividing mammalian cells alike. Currently used lenti- and other retroviruses, as well as adeno- and adeno-associated viruses have a limitation on DNA cargo size due to spatial constraints imposed by the tight geometry of their capsids.

Baculoviral vectors, in contrast, can accommodate very large DNA cargo insertions[19]. The *Autographa californica* multiple nuclear polyhedrovirus (AcMNPV), is a baculovirus with a large (134 kb) double-stranded circular DNA genome that normally infects specific moth larvae[19]. Transgene capacity of AcMNPV is very large, extending probably beyond 100 kbp. Replication of AcMNPV is highly insect-cell specific; however, AcMNPV is capable of efficiently transducing not only insect but also mammalian cells. Transduction is usually transient without DNA integration into the target cell genome and such viruses are replication deficient[20,21]. In baculoviruses used for mammalian cell transduction (BacMam)[22,23] heterologous genes are placed under the control of mammalian promoters and inserted into the baculoviral genome, and viral stocks are produced in insect cells. Once the baculovirus enters mammalian cells, these genes are actively transcribed within 9 h and the cells produce the heterologous gene product. In the last decade, baculovirus has emerged as a useful and safe technology to deliver heterologous genetic material to mammalian cell types both *in vitro* and *in vivo*[24].

Here we introduce MultiPrime, a novel tool-kit specifically designed for efficient delivery of large multigene constructs into primary and established mammalian cells. MultiPrime enables simultaneous expression of multiple, independent cassettes in mammalian cells. This system combines the ease and flexibility of our recombination-mediated multigene DNA assembly technology with the superior performance of baculovirus as a viral vector for transducing mammalian cells. We transduced with MultiPrime a large variety of cell types including difficult to transfect stem cells and primary neurons. Moreover, we utilized MultiPrime for reprogramming mouse embryo fibroblasts (MEFs) into neurons. Further, we show that our system is not limited to mammalian transduction but can also be used to transduce zebrafish embryos. We applied our system to create synthetic multicomponent intracellular biosensor systems, such as Rab GTPases regulating vesicular membrane traffic in cells, phosphoinositide binding signalling proteins or fluorescently-labelled cytoskeletal markers. These biosensors were simultaneously delivered and expressed in mammalian cells allowing to quantitatively monitor a large variety of intracellular parameters. A wide range of promoters are available in MultiPrime, to regulate and fine-tune individual heterologous target gene expression.

With the objective to provide a means to concomitantly exploit with one single expression system the advantages of high-level protein production in baculovirus-infected insect cells and protein expression in a 'native' mammalian environment, we incorporated dual-host-specific promoters in MultiPrime that are functional in both insect and mammalian cells.

Our MultiPrime system is not limited to the assembly of RNA polymerase II-based expression cassettes. It can likewise harbour U6-driven expression cassettes and homology constructs that are required for CRISPR/Cas9-mediated genome engineering. We demonstrate the aptitude of our system by applying MultiPrime-CRISPR/Cas9 to express a HMGA1-EGFP fusion protein in primary cells from the native genomic HMGA locus.

Results

MultiPrime system design. We developed MultiPrime specifically to overcome the limitations of transfecting mammalian cells for multigene transfer applications. We redesigned our previous pFL plasmid[25], which contains the Tn7R and Tn7L DNA elements required for Tn7-transposase-mediated integration into a baculovirus genome containing a Tn7-attachment site[26]. Expression cassettes with promoters active in mammalian cells, or, alternatively, with activity in both mammalian and insect cells, were inserted into this MultiPrime acceptor plasmid. This acceptor plasmid is poised to receive further multigene expression cassettes by incorporating donor plasmids (Supplementary

Figs 1–3). All donors from our previous plasmid-based MultiLabel system can be used for incorporation into this new acceptor to generate acceptor-donor fusions by recombineering. Moreover, expression cassettes can be freely exchanged between donors and acceptors due to the modular design[14].

Acceptors or acceptor–donor fusions containing multigene expression cassettes are integrated into baculoviral genomes by means of Tn7 transposition[25]. We utilized two different baculoviral genomes in our experiments. In addition to our EMBacY baculoviral genome[25] we generated in this study a new baculoviral genome, MultiBacMam, which expresses a vesicular stomatitis virus glycoprotein (VSV-G) and the fluorescent protein mCherry in insect cells during virus production. Both genomes were generated from the original MultiBac virus by integrating genes encoding EYFP (EMBacY) or mCherry and VSV-G (MultiBacMam) by Cre recombination into the LoxP site present on the MultiBac virus[12,25]. The expression of mCherry results in a characteristic purple colour of MultiBacMam infected cell cultures (Supplementary Fig. 1), thereby simplifying the tracking of virus amplification by eye. Moreover, the MultiBacMam virus gives rise to baculovirions displaying VSV-G on their surface. The presence of VSV-G in the baculoviral envelope has been shown to increase the efficacy of mammalian cell transduction[27]. All composite baculoviruses prepared in this study were produced in Sf21 cells. Virus was obtained with comparable efficiency and resulting in similar viral titers to what we had observed in previous multigene expressions in insect cells with recombinant MultiBac or EMBacY viruses[25]. Our multigene viruses are characterized by multiple use of regulatory elements such as the CMV promoter, which in theory could introduce genomic instability during repeated rounds of viral amplicfication[14,25]. We had developed previously efficient protocols to amplify baculoviral constructs containing multiple copies of late viral promoters (polh and p10) by stringently applying a low multiplicity of infection (MOI) regimen and few, ideally not more than two amplification rounds thus preventing accumulation of non-productive viruses containing genomic deletions[28]. Strictly adhering to this strategy for amplifying baculoviruses containing multiple copies of CMV promoter-driven expression cassettes again resulted in stable virus producing all proteins of choice in each transduced mammalian cell in homogenous cultures, while failure to adhere to the protocol resulted in heterogeneous cell populations where expression of individual heterologous genes had been lost, presumably due to accumulation of deletion virus species (Supplementary Fig. 4). Occasionally, we observed cell fusion in tissue culture plates of insect cells infected with MultiBacMam viruses, which, however, was found not to be detrimental to transduction experiments with the viral supernatant. In the following, we use the EMBacY baculoviral genome to prepare composite MultiPrime baculoviruses, unless indicated otherwise.

Transduction and viability. As proof-of-concept, we generated a composite MultiPrime baculovirus expressing five fluorescently-tagged proteins (EBFP2-nucleus, mTFP1-FYVE (PI-3-P-binding, early endosomes), EYFP-tubulin, Mito-dsRED (mitochondria) and Plum-PLCδ-PH (PI-4,5-P_2-binding, plasma membrane) localizing to different subcellular compartments. We used this baculovirus to test a variety of different cell types for their propensity to be transduced efficiently. We initially transduced well-established cell lines including HEK293, PAE, COS7, HeLa, SK-MEL-28, CCL39 and Swiss 3T3. All cell lines that were tested proved to be highly transduceable by MultiPrime and efficiently expressed all heterologous proteins (Supplementary Fig. 5 and data not shown). Typically, transduction efficiencies ranged between 20 and 100%. Transduction frequencies up to 100% were observed even in cell lines, such as PAE that are traditionally considered to be difficult to transfect.

We then asked whether we can use MultiPrime to transfect primary cells. Primary cells are an intense focus of contemporary research efforts for many reasons, and it is well documented that they are typically extremely difficult to transfect. For this experiment, we utilized human umbilical vein endothelial cells (HUVEC), rat embryo fibroblasts (REF), rat cortical neurons and human-induced pluripotent stem cells (iPS). With our MultiPrime virus, all of these primary cell types could be transduced efficiently and again expressed all heterologous genes of interest, compellingly underscoring the utility of our approach (Fig. 1 and Supplementary Fig. 5).

To analyze viability and functionality, we proceeded to express multiple intracellular sensors simultaneously from a MultiPrime baculovirus. COS7 cells expressing epidermal growth factor receptor (EGFR) endogenously were used to monitor trafficking of EGFR. Genes encoding fluorescently-tagged markers for early (RAB5A), recycling (RAB11A) and late (RAB7A) endosomes were expressed from a single MultiPrime virus and the cells were stimulated 40 h post-transduction with Cy5-labelled epidermal growth factor (EGF). As expected, EGF was found in early (RAB5A+) and late (RAB7A+) endosomes, but not in recycling (RAB11A+) endosomes after 30 min (Fig. 2a). A time-resolved quantitative analysis confirmed that EGF was transferred to late endosomes but not to RAB11-positive, recycling vesicles (Supplementary Movie 1). Next, we used PAE cells stably expressing VEGFR2 and neuropilin-1 (NRP1). A composite MultiPrime baculovirus was used to express fluorescently-tagged Rab4, Rab5 and Rab7 in these cells. Cells were then stimulated with Nt647-labelled VEGF-A165a for 3 h. VEGF was found in Rab5 and Rab7 vesicles clearly indicating that it was properly internalized, in accordance with previous reports[3]. Using the same cell line transduced with the RAB5A-RAB11A-RAB7A virus in a time-resolved study, we could show that VEGFR2 behaves differently from EGFR. VEGFR2 is, in contrast to EGFR, partly recycled through the RAB11 compartment (Supplementary Movie 2).

Next, we tested the functionality of baculovirus-transduced HUVEC in migration and angiogenesis assays. HUVEC transduced with a MultiPrime baculovirus encoding mTFP1-actin, EYFP-tubulin and Mito-DsRed were plated on matrigel and incubated for 14 h. HUVEC are known to establish a characteristic vascular network within this time frame. Transduced cells efficiently integrated in this network, clearly demonstrating that MultiPrime-transduced HUVEC show identical behaviour compared to untransduced cells (Fig. 2c). Moreover, in a migration assay, we could show that transduced cells migrate with similar efficiency as untransduced cells (Fig. 2d).

We quantitatively compared baculovirus-mediated transduction with the classical plasmid transfection approach. In addition to EMBacY, we used our MultiBacMam baculovirus displaying VSV-G on its surface in this experiment. As expected, we observed increased transduction rates with MultiBacMam compared to EMBacY, albeit the gain in efficacy in our hands turned out to be modest in many cases. Clearly, transduction with both EMBacY and MultiBacMam outperformed the classical transfection approach (Fig. 3a). We routinely obtained transduction efficiencies higher than 50% when using the MultiPrime baculovirus compared to transfection efficiencies well below 20% with the corresponding plasmid (13,305 bp) that had been used to generate the composite baculovirus.

Baculovirus displaying VSV-G on its surface was superior to virus lacking VSV-G at all tested MOI. Saturation was usually obtained at a MOI of 500 (Fig. 3b). The relative expression levels

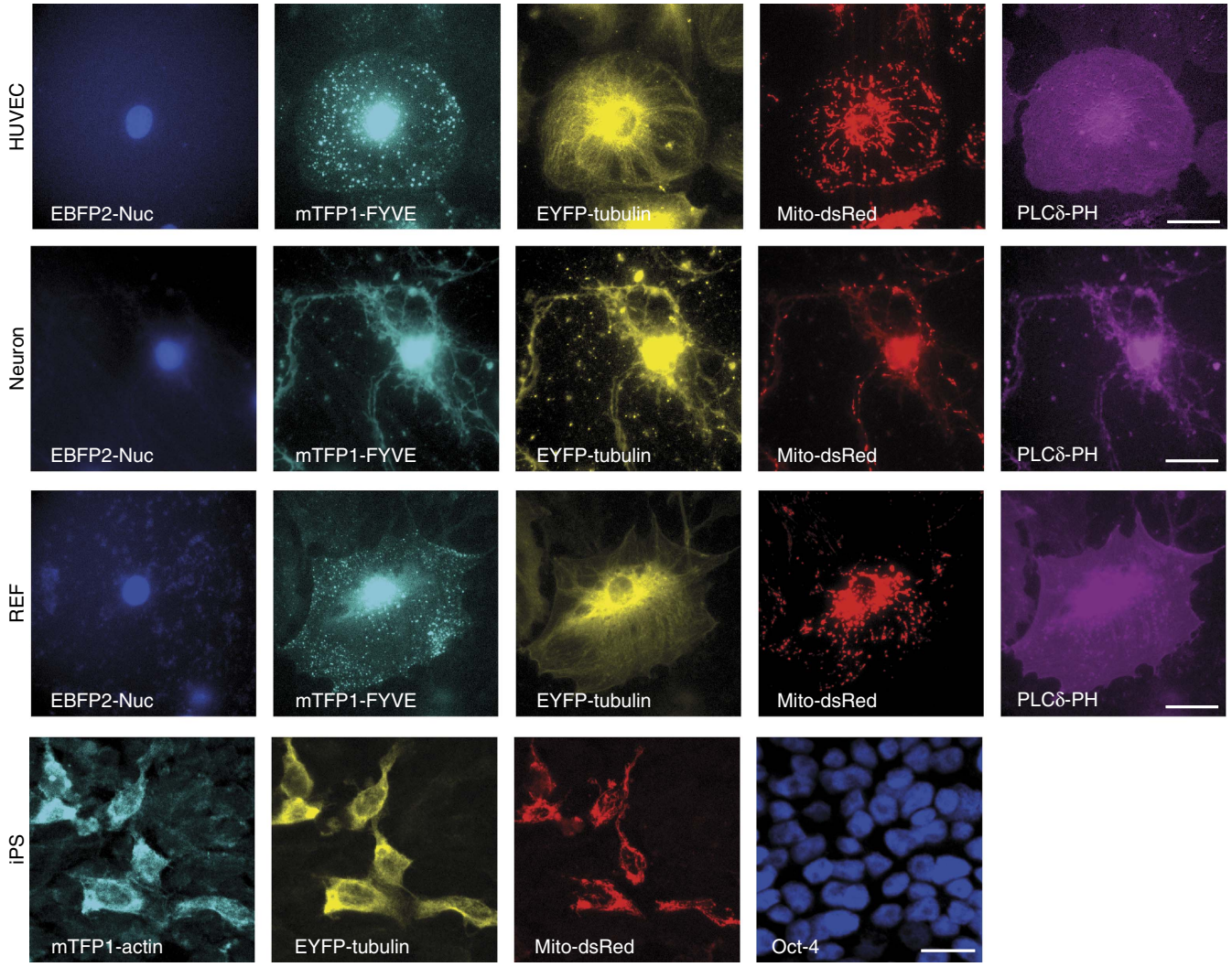


Figure 1 | Multigene expression in primary cells by MultiPrime. HUVEC, REF cells and rat cortical neurons were infected with a MultiPrime baculovirus encoding EBFP2-Nuc (labelling the nucleus), mTFP1-FYVE (PI-3-P containing endosomes), tubulin-EYFP (cytoskeleton), Mito-dsRed (mitochondria) and PLCδ-PH (PI-4,5-P2; plasma membrane). iPS cells were transduced with a virus encoding mTFP1-actin, EYFP-tubulin and Mito-dsRed. Oct4 was used as a marker for pluripotent stem cells. All infected cells express all heterologous proteins. Scale bar, 20 μm.

between cells appear to be similar (Fig. 3c). This is in contrast to transfected cells that typically show a wide variety of expression levels (data not shown). Since we use relatively high MOI, the toxicity of the virus could conceivably be an issue. We tested baculovirus toxicity at MOI 500 compared to plasmid transfection with Fugene HD, which is considered to be a mild transfection reagent. Both EMBacY- and MultiBacMam-derived viruses exhibit negligible toxicity similar to plasmid-based transfection (Fig. 3d).

Baculovirus transduction of mammalian cells is transient in nature as the foreign DNA does not integrate into the host genome. We therefore tested the persistence of recombinant expression following transduction with a MultiPrime baculovirus by immunofluorescence and western blotting. In our experiments, the percentage of positive cells decreased to ~20% after 20 days and 5% after 30 days (Fig. 3e).

Modulation of expression levels in mammalian cells. The hCMV-IE1 promoter we used in our experiments is considered to be the strongest promoter available for heterologous expression in most mammalian cells. It may be desirable to have alternative promoters that are characterized by lower levels of expression. We expanded our tool-box by incorporating the SV40, PGK and UBC promoters in alternative expression cassettes in our MultiPrime system (Supplementary Fig. 6). We determined expression levels from these alternative promoters by expressing EYFP-tubulin, and simultaneously expressing citrine from a CMV promoter as a bench-mark to normalize expression levels. All three alternative promoters show distinctly lower expression levels in HEK293 and PAE cells as well as in primary REF compared to CMV promoter-driven expression (Fig. 4a,b). Furthermore, we included a tetracycline-inducible promoter in our system (Supplementary Fig. 6). Tetracycline-inducible promoters are dependent on a transactivator, for example tTA, to initiate expression[29]. We observed approximately four times higher expression levels in the absence of doxycycline in HeLa cells stably producing tTA, which were transduced with a MultiPrime baculovirus containing a tetracycline-inducible expression cassette, in good agreement with reports involving tetracycline-inducible promoters on plasmids (Fig. 4c,d).

Bifunctional dual-host promoters. Expression plasmids that could be used for heterologous protein production in insect as well as in mammalian cells have not found wide-spread

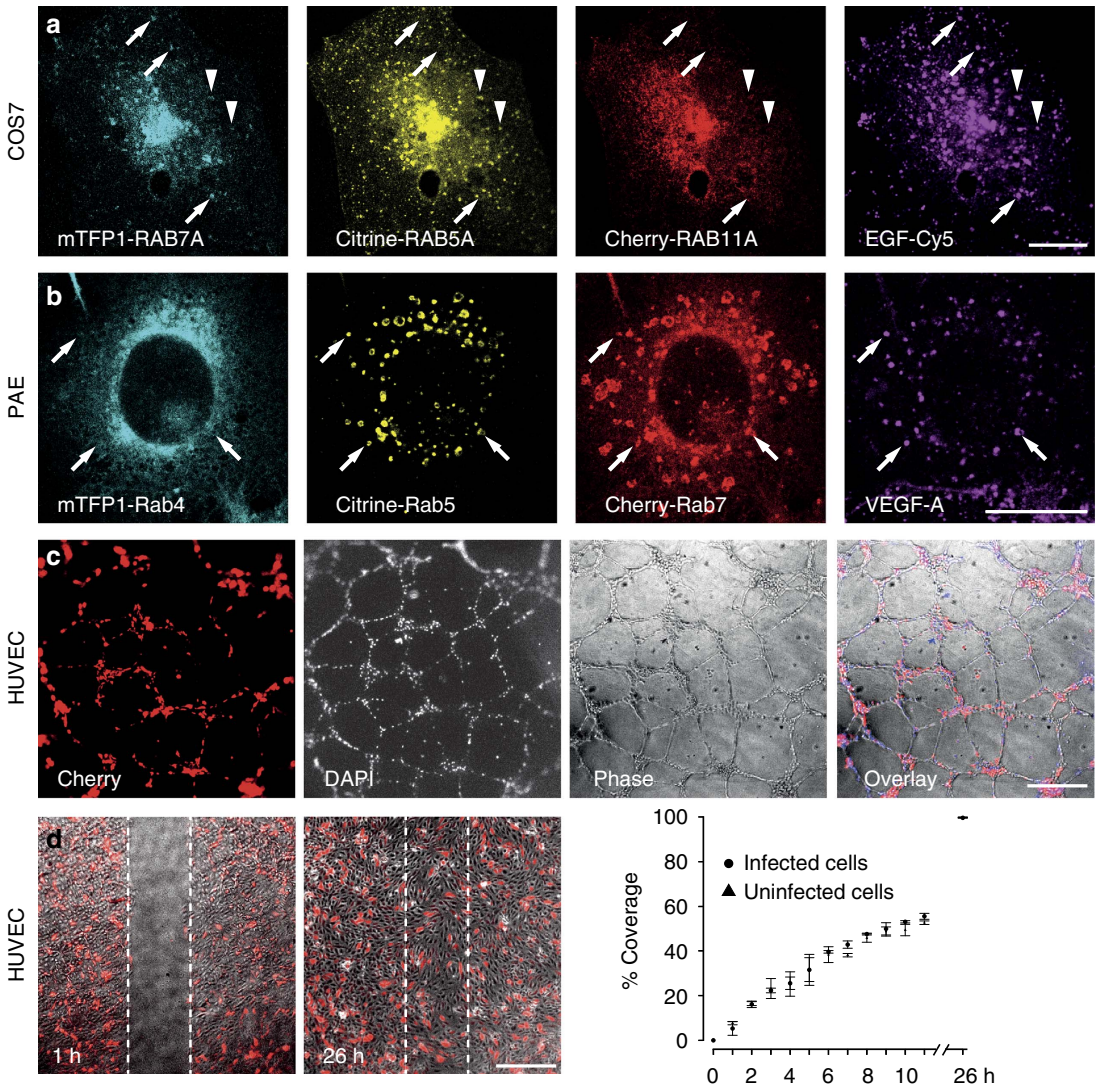


Figure 2 | Infected cells retain functionality. (**a**) COS7 cells were infected with MultiPrime baculoviruses expressing the indicated fluorescently-tagged RAB GTPases. Cells were stimulated for 3 h with Cy5-labelled EGF. As expected, EGF was found in RAB7A vesicles (arrows) and RAB5A vesicles (arrowheads) but not in RAB11 vesicles. A quantitative time-resolved analysis is provided (Supplementary Movie 1). (**b**) PAE cells stably expressing VEGFR2 and neuropilin-1 were infected with baculoviruses expressing the indicated fluorescently-tagged RAB GTPases. Cells were then stimulated with Nt647-labelled VEGF. VEGF can be found in RAB5 and Rab7 vesicles (arrows). (**c**) Tube formation: HUVEC were infected with a baculovirus expressing mTFP1-actin, EYFP-tubulin and Mito-dsRed (only red channel is shown). DAPI was used to counterstain nuclei of all cells. Approximately 30% of cells were infected. Infected and uninfected cells contribute to tubes. (**d**) Migration: HUVEC were infected with a baculovirus expressing mTFP1-actin, EYFP-tubulin and Mito-dsRed (only red channel is shown). Approximately 30% of cells were infected. Infected and uninfected cells migrate at same rates. Scale bar, 20 μm (**a**,**b**); 500 μm (**c**,**d**).

application so far, possibly because comprehensive comparative data which would have encouraged their use is currently lacking. We addressed this issue by creating, validating and incorporating dual-host promoters as a choice in our MultiPrime system. Our objective was to provide a single expression reagent, which is the composite MultiPrime baculovirus containing the genes of choice controlled by this validated dual-host promoter, for example to produce a protein or protein complex of choice efficiently in insect cells for structural studies and in mammalian cell lines for functional validation. We used two promoters, the first one (denoted CMVP10) is a fusion of the CMV promoter and the baculoviral very late promoter p10, the second (denoted CMVintP10) contains the p10 promoter in an intron of the CMV transcription unit (Fig. 5a and Supplementary Fig. 7). These two dual-host promoters were validated in mammalian cells by expressing EYFP-tubulin from a MultiPrime baculovirus, which also expressed citrine driven by a CMV promoter for normalization purposes. In HEK293 and PAE cells, the dual function promoters expressed at comparable levels to the original mammalian-only CMV promoter. In REF cells, the intron-less CMVP10 promoter resulted in lower expression (Fig. 5b,c). We quantified expression from these MultiPrime baculoviruses in insect cells and found them entirely satisfactory (Fig. 5d,e). Furthermore, we tested MultiPrime constructs expressing human transcription factors, which we had produced before for structural studies in insect cells with our MultiBac insect-cell expression system (Supplementary Fig. 8)[30,31]. We observed virtually indistinguishable levels of expression for complexes formed by these human TATA-box associated factors (TAFs) from dual function promoters as compared to the MultiBac expressed complexes. Transduction of HeLa cells with the TAF producing MultiPrime baculoviruses resulted

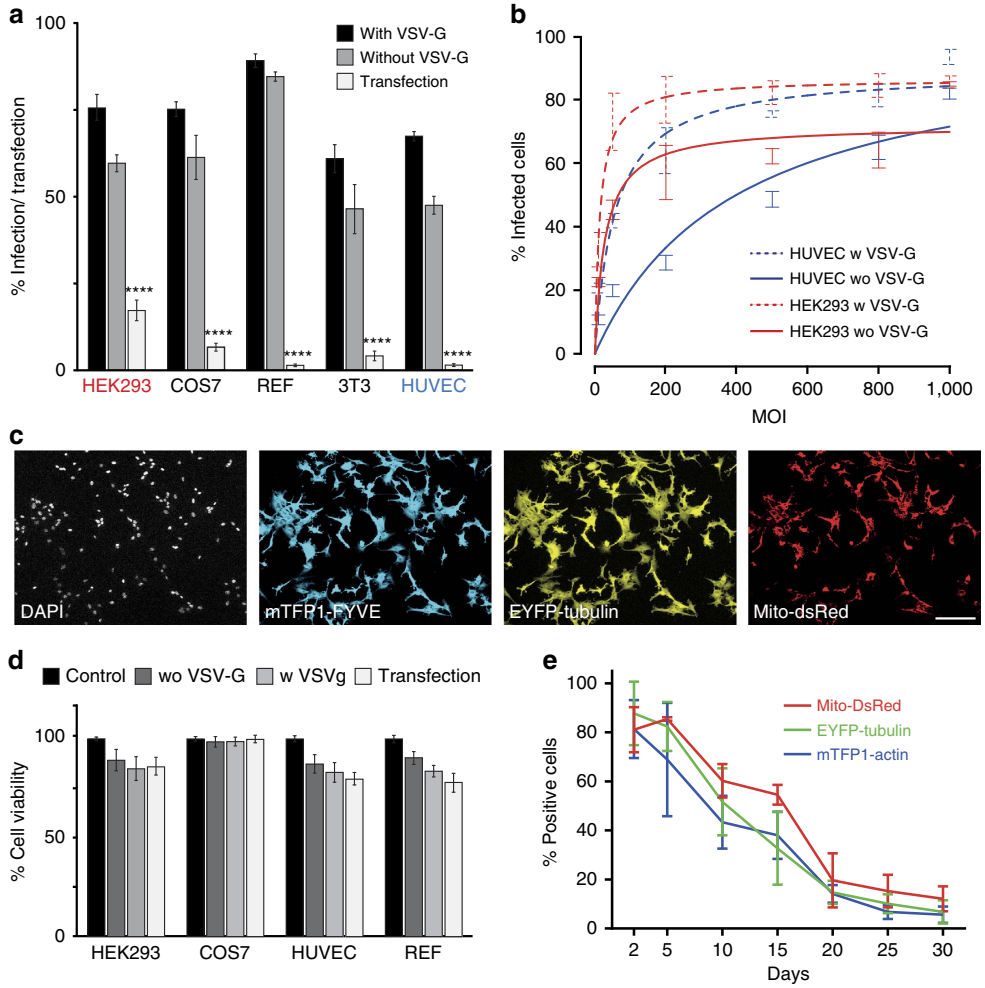


Figure 3 | Transduction efficacy. (**a**) Transduction by MultiPrime baculovirus was compared to plasmid-based transfection. MultiBacMam bacoluvirus expressing VSV-G and EMBacY baculovirus devoid of VSV-G were used. For plasmid-based transfection, the plasmid (13 kb) used originally inserted into the recombinant baculoviruses was utilized. MultiPrime-mediated transduction is markedly superior in all cell types tested. Data shows mean value ± s.d.; $n = 3$; ****$P < 0.0001$ determined by comparing transfection with transduction with or without VSV-G using one way analysis of variance followed by the Dunnet's *post hoc* test. (**b**) Effects of the MOI are shown. (**c**) PAE cells were infected with a MultiPrime baculovirus expressing three proteins at a MOI of 500. DAPI was used to counterstain nuclei of all cells. Virtually all cells are infected and express all heterologous proteins. Scale bar, 100 µm. (**d**) Toxicity of transduction was measured by means of a MTT assay. MultiPrime transduction exhibits comparable, low toxicity as plasmid transfection. (**e**) The persistence of heterologous expression was quantified by fluorescence in REF cells. The percentage of positive cells for each individual protein is shown over time.

in close to complete transduction rates (Supplementary Fig. 8).

Genome engineering by CRISPR/Cas9. CRISPR/Cas9-mediated genome engineering requires the expression of Cas9, the concomitant expression of a U6-driven guide RNA (gRNA) and the provision of a DNA construct for homologous recombination. Currently used viral systems can harbour Cas9 and the gRNA but are unable to include a homology construct due to limited cargo capacity[32]. We assembled DNAs for the expression of a HMGA1-EGFP fusion protein from its endogenous locus in a MultiPrime virus[33] (see Supplementary Fig. 9 for details). Transduction of HEK293 and HUVEC led to expression of HMGA1-EGFP in the nucleus in ~1% of the cells. Successful homologous integration of the DNA construct was verified by PCR (Fig. 6a and Supplementary Fig. 9).

Reprogramming by MultiPrime. We next investigated whether MultiPrime viruses are suitable for reprogramming of cells. Currently, this is mainly carried out with lentivirus, which is a retrovirus that stably integrates into the genome of cells. We assembled a MultiPrime virus expressing the transcription factors Asc1, Brn2 and Myt1L, which were shown to convert MEFs into neurons[7]. Transduction of MEFs with this MultiPrime virus resulted in cells with neuron-like morphology, which expressed the neuronal markers MAP2 and β-tubulin III 20 days after transduction, indistinguishable from co-infection with three lentiviruses each expressing one of the transcription factors (Fig. 6b). Our results provide compelling evidence that reprogramming can be successfully achieved with a transient expression system such as MultiPrime.

Functional antibody production. Our MultiPrime approach can not only induce morphological changes in cells but also potentially interfere with it. We addressed this by using MultiPrime to express functional antibodies in primary cells. Our previously described single-chain antibodies targeting VEGF (SZH9) and VEGFR2 (ADH9), and a control single-chain

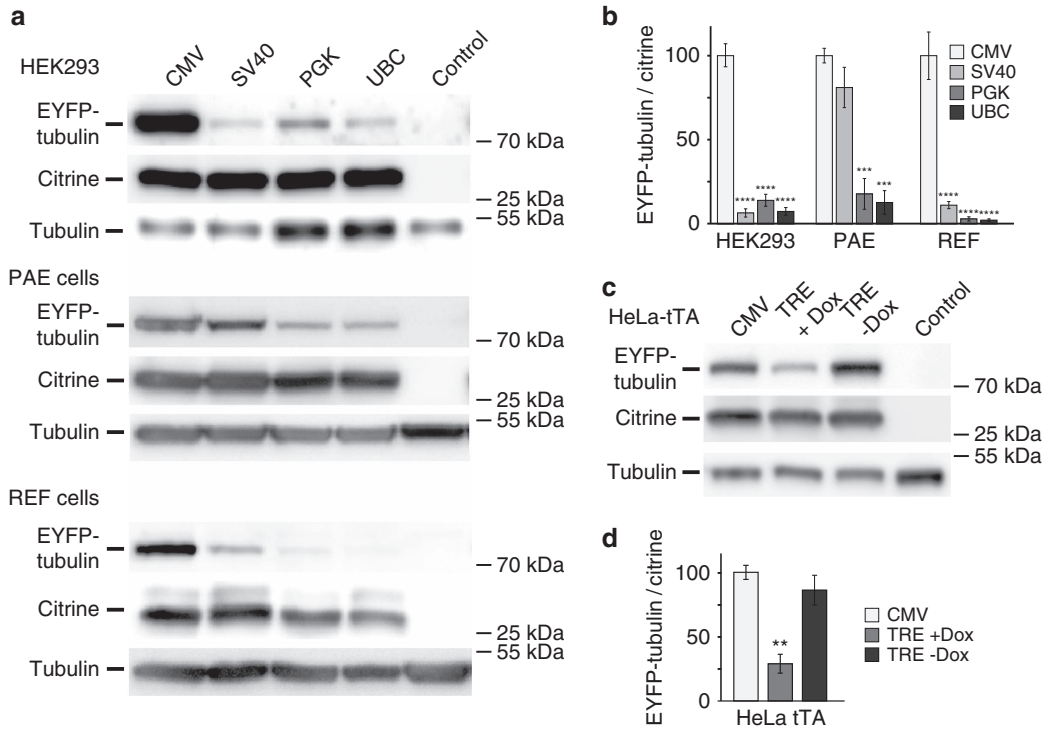


Figure 4 | Modulation of expression levels with alternative promoters. (**a**) Transduction of HEK, PAE and REF cells with baculoviruses expressing EYFP-tubulin under the control of the indicated promoters and citrine under the control of the CMV promoter is shown. (**b**) Quantification of the blots. EYFP-tubulin/citrine ratio was used as a measure for promoter strength. Endogenous tubulin was used as loading control. Data shows mean value ± s.d.; $n=3$; ****$P<0.0001$, ***$P<0.001$ determined by comparing CMV promoter with alternative promoters using one-way analysis of variance followed by the Dunnet's *post hoc* test. The CMV promoter is the strongest promoter in all tested cell lines. (**c**) HeLa-tTA cells were infected with a baculovirus expressing EYFP-tubulin under the control of a tetracyclin inducible element and citrine under the control of the CMV promoter. (**d**) Quantification of the blots shown above. Data shows mean value ± s.d.; $n=3$; **$P<0.01$ determined by comparing induced versus non-induced tet promoter using one-way analysis of variance followed by the Tukey *post hoc* test. No significant difference between CMV promoter and induced tet promoter.

antibody (A1) were converted into a full length IgG consisting of light and heavy chains[34,35]. The dual-host promoter CMV-CMVintP10 was utilized to drive recombinant IgG expression. The resulting MultiPrime baculovirus was successfully tested for expression in HEK293 and insect cells (Fig. 6c). All IgG antibodies tested were expressed at comparable levels. The same virus was then used to transduce HUVEC that were then placed into a tube formation assay. Only the function-blocking anti-VEGF antibody SZH9 was able to interfere with tube formation. All other antibodies, including the VEGFR2-binding but not function-blocking antibody ADH9, did not interfere with tube formation (Fig. 6c).

Zebrafish transduction. It was previously shown that mammalian promoters can be used for heterologous expression in zebrafish embryos[36]. We set out to establish whether MultiPrime viruses are restricted to mammalian and insect cells, or whether they can also be used to transduce zebrafish. A MultiPrime virus encoding mTFP1-actin, EYFP-tubulin and Mito-dsRed under control of mammalian CMV promoters was injected into intercellular spaces in the brain region of zebrafish embryos at 24 h post fertilization. Injection of this virus showed heterologous expression of all genes in zebrafish embryos. Expression was restricted to the site of injection and could be detected for at least 5 days (Fig. 6d).

Discussion

In the three decades since their inception, baculovirus-based expression systems have become well-established and widely used for recombinant protein production in insect cells. Later, it was discovered that baculoviruses not only infect insect cells but can also drive heterologous protein expression in mammalian cells if appropriate mammalian regulatory elements are provided in the recombinant baculovirus genome[22,23,37]. This so-called 'BacMam' method has been applied to produce heterologous proteins in academic and industrial research and development, notably for pharmacological screening[37]. Today, it is becoming increasingly evident that most physiological activities are mediated by multiple proteins forming complex assemblies. Therefore, a powerful tool that exploits recombinant baculovirus to deliver multiple genes simultaneously and reproducibly into a range of mammalian cell types and notably primary cells is highly desirable to study health and disease states, and to analyze molecular mechanisms of cell fate. Notwithstanding, such a tool has been lacking so far. Therefore, we developed MultiPrime, a versatile, flexible and fully modular system for efficient multigene delivery and expression in any mammalian cell type, primary and established. MultiPrime relies on a set of customized DNA plasmid modules, called acceptors and donors that provide the means to combine a theoretically unlimited number of genes of interest with different promoters, terminators and other control elements in multiple expression cassettes to generate multigene-delivery constructs, which are then inserted into engineered baculoviral genomes. Moreover, they can comprise all the elements necessary for genome engineering including editing functions and the sequences required for homologous recombination. We provide and validate a range of mammalian promoters that can be introduced into our MultiPrime system in this way. In addition,

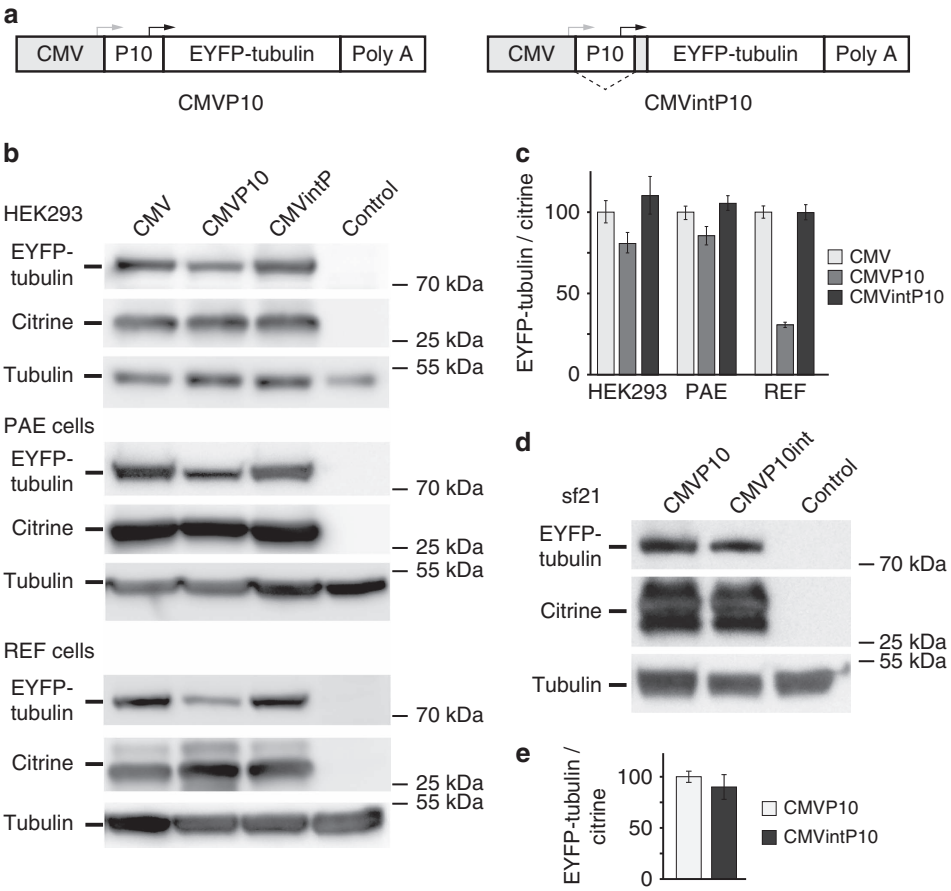


Figure 5 | Promoters active in mammalian and insect cells. (**a**) Structure of the tested dual promoters is shown schematically. In CMVP10, the baculoviral very late promoter p10 was inserted downstream of the CMV promoter. In CMVintP10, the p10 promoter was placed within an intron and is spliced out from the transcript of the CMV promoter. Grey arrow: transcription initiation of the CMV promoter; black arrow: transcription initiation of the p10 promoter. (**b**) HEK293, PAE and REF cells were infected with a MultiPrime baculovirus expressing EYFP-tubulin under the control of the above promoters and citrine under the control of the CMV promoter. The lysates of the cells were analysed by western blotting. (**c**) Quantification of blots shown in **b**. EYFP-tubulin/citrine ratio was used as a measure for promoter strength. Endogenous tubulin was used as loading control. The CMVintP10 promoter expresses at a similar level as the original CMV promoter, while the CMVP10 promoter expresses at a lower level in mammalian cells. (**d**,**e**) The same baculoviruses were used to infect insect (Sf21) cells. Citrine was expressed by a P10 driven expression cassette in the backbone of the baculovirus. Data shows mean value ± s.d.; $n = 3$; there is no significant difference ($P > 0.05$) by comparing the two CMVP10 promoter variants using the Student's *t*-test.

we provide dual-host promoters to drive heterologous multiprotein production in both insect and mammalian cells. This highly versatile and flexible tool-box allows users to conveniently introduce many different proteins simultaneously into mammalian and insect cells. Corroborating previous observations, we found negligible toxicity and sustained viability when infecting a range of mammalian cells with recombinant MultiPrime reagents. Importantly, we demonstrate here that MultiPrime infected cells are competent to divide and migrate normally and are capable of adequately responding to external stimuli as, for example, growth factors.

In this study, we utilized two baculovirus types, EMBacY and MultiBacMam. These engineered baculoviral genomes are characterized by reduced proteolysis and delayed cell lysis during virus amplification in insect cells, resulting in high quality, high titre virus[25]. The EMBacY and MultiBacMam viruses express either EYFP or mCherry fluorescent marker genes, to signal late replication cycle entry. Expression of fluorescent marker proteins during virus production is a convenient tool to simplify and standardize monitoring the production of baculovirus, in particular for non-specialist users. MultiBacMam virus generated in this study expresses also VSV-G during virus production in insect cells. Decorating baculovirus with VSV-G has been shown to improve mammalian transduction efficiencies. Consequently, we observed superior transduction efficiencies with our MultiBacMam-derived viruses that display VSV-G on their capsids. We note here that, at least in Switzerland where these experiments were performed, MultiBacMam-derived reagents expressing VSV-G have to be handled at biosafety level 2, which requires specific laboratory infrastructure. To circumvent this complication the EMBacY virus variant can be utilized, which is devoid of VSV-G but still resulted in satisfactory transduction rates in our experiments. Nonetheless, for experiments which may rely on maximum transduction efficiencies, the VSV-G containing MultiBacMam virus is recommended.

All viral genomes we utilized contain a site-specific integration site in the backbone distal from the Tn7-attachment site. This LoxP site allows introduction of additional genes by Cre-LoxP mediated fusion *in vivo*[12]. This enables a range of options to modify and tailor the baculovirus genomes for specific applications. For example, a baculovirus called SweetBac was developed to achieve mammalian-type glycosylation of recombinant secreted proteins such as antibodies[38,39].

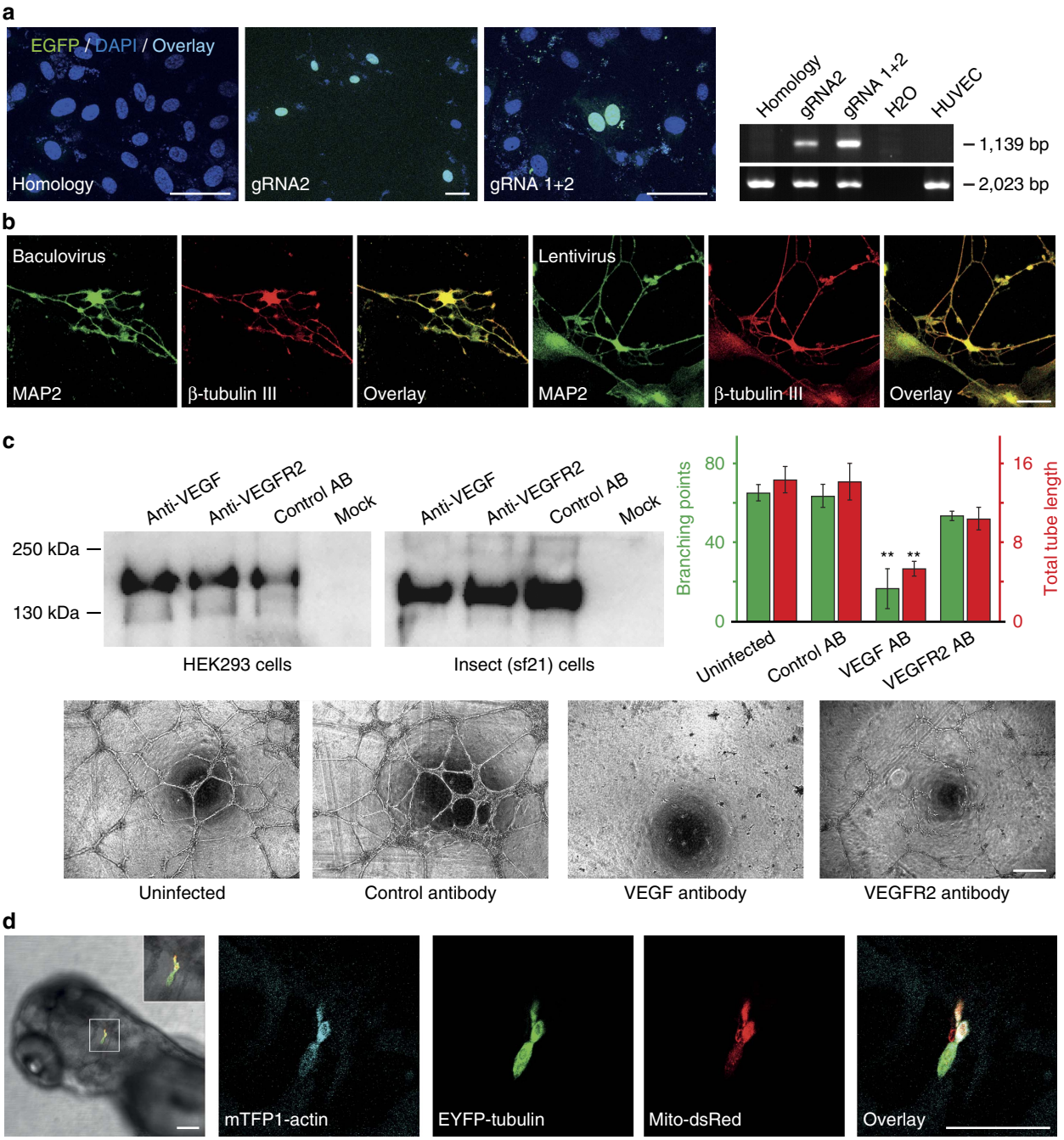


Figure 6 | MultiPrime applications. (**a**) Genome engineering. Infection of HUVEC with a baculovirus containing a HMGA1-EGFP homology construct did not lead to cells with HMGA1-EGFP expression (left). Co-expression with Cas9 and HMGA1-gRNA1 (2nd panel) or co-expression with Cas9, HMGA1-gRNA1 and HMGA1-gRNA2 (3rd panel) led to cells with HMGA1-EGFP expression in the nucleus. Correct integration of the homology construct was verified with PCR. The wild-type allele yielded a fragment of 2,023 bp, whereas the mutant allele results in a fragment with 1,139 bp (right). Scale bar, 50 μm. (**b**) Reprogramming of cells. MEF cells were infected with a Multiprime virus expressing Ascl1, Brn2 and Myt1L or co-infected with three lentiviruses individually expressing the same transcription factors. Both strategies led to cells with neuron-like morphology that express the neuronal markers MAP2 and β-tubulin III. Scale bar, 50 μm. (**c**) Functional antibody expression. MultiPrime viruses, encoding light and heavy chains of three different IgGs (anti-VEGF, anti-VEGFR2, and unspecific) were used to express antibodies in HEK293 (left) and also in insect cells (middle). HUVEC cells were infected with these viruses and the cells were used in a Matrigel-based angiogenesis assay. As expected, only the anti-VEGF antibody is capable of blocking tube formation Scale bar, 500 μm. Data shows mean value ± s.d.; $n = 4$; $^{**}P < 0.01$ when comparing VEGF function-blocking antibody versus control antibody. There is no significant difference ($P > 0.05$) when comparing the non-function-blocking VEGFR2 antibody with the control antibody. Both P values are determined using one way analysis of variance followed by the Dunnet's *post hoc* test. (**d**) Baculovirus-mediated gene expression in zebrafish. Dorsal view of the head of a 3-day-old zebrafish larva after injection of MultiPrime baculoviruses expressing mTFP1-actin, EYFP-tubulin and Mito-DsRed into the hindbrain region at 24 h post fertilization. All infected cells express all heterologous proteins. Scale bar, 100 μm.

Text

Currently, these functionalizations are limited to applications in insect cells. We anticipate that a wide range of functions to modify, enhance and regulate multiprotein production in mammalian cells will be exploited by modifying the baculoviral genome accordingly, providing appropriate expression cassettes active in mammalian cells in the LoxP locus of these vectors.

Multigene expression systems are rapidly gaining prominence for producing protein complexes for structural and functional studies. Often, several expression systems must be tested to obtain functional complexes in sufficient quantity and quality. This typically requires recloning of genes into different sets of expression plasmids given that the regulatory elements in each system, here mammalian and insect cells, are optimized for a particular host, and are typically not compatible between the different species. The incorporation of dual-host promoters into MultiPrime allows simultaneous testing of expression constructs in insect and mammalian cells by using the same reagent. This feature can be conveniently exploited if high-level production of a complex protein of interest is carried out in insect cells, while functional analysis of the same complex is performed in mammalian cells, which is increasingly the case in current structural biology. The possibility to use the same reagent for both host systems will also benefit analysis of structure–function correlations requiring multiple mutational analysis. MultiPrime affords the means to carry out such elaborate studies, notably also of complexes controlling cell fate, which can be mechanistically dissected by infecting primary cells.

Baculovirus constitutes an attractive tool for gene therapy for a number of reasons. Due to its flexible envelope structure, very large heterologous DNA cargo can be incorporated into the baculoviral genome. Moreover, baculovirus is replication incompetent in mammalian cells, and virtually no viral protein expression occurs on transfection in a mammalian host. Initial *in vivo* experiments had limited success since injected baculoviruses are rapidly inactivated and cleared by the immune complement system. Strategies were developed to overcome this impediment and many successful *in vivo* applications were published since then (reviewed in ref. 24). For example, expression of VEGF-D-induced vascularization in rabbit skeletal muscle suggesting that baculovirus-driven VEGF-D expression might be an option to cure lymphatic disorders[40]. Nevertheless as a non-integrative virus it is *a priori* limited to transient expression, which can be an advantage or a disadvantage depending on the application. Transient expression may be desirable, for example, for vaccination or to promote changes in cell fate. Of note, altering cell fate is a particularly interesting application for multigene expression systems, as it relies on the simultaneous and temporally restricted expression of several transcription factors. Induced-pluripotent stem cells have been generated before with a BacMam virus *in vitro* using a fusion protein construct[41]. Four transcription factors were expressed as a fusion protein from a single open reading frame (ORF) via self-cleaving 2A peptides. Here, we converted MEFs into neurons using independent expression cassettes, which offer advantages especially when different protein combinations need to be tested in a combinatorial fashion.

MultiPrime is not restricted to the delivery of expression constructs. We anticipate that genome engineering will be an important future application, owed to the very large cargo capacity of baculoviruses. Other viruses such as lentiviruses or adeno-associated viruses cannot accommodate all DNA elements needed to produce Cas9, a gRNA and a construct for homologous recombination. With MultiPrime, we were able to modify the HMGA1 locus of HUVEC, which are human primary cells that show restricted replication potential.

Our results compellingly validate MultiPrime as a powerful vehicle for multigene delivery, protein expression and genome engineering, relevant for a large number of applications, *in vitro* and *in vivo*, and underscore the enormous potential of our baculoviral system to deliver large multigene DNA constructs into a wide range of mammalian cells, notably including primary cells. A multitude of genes and regulatory elements can be delivered due to the very large heterologous DNA cargo capacity of the system, offering novel exciting possibilities for biological research. Entire signalling cascades, gene regulatory systems or metabolic pathways and multiple mutants thereof, can be efficiently engineered with MultiPrime. We anticipate that many applications will benefit from MultiPrime, notably when efficient transfer of multiple genes or efficient engineering of genomes is required.

Methods

Molecular biology. DNA construction in MultiPrime follows the high-throughput compatible logic of our ACEMBL concept to prepare multicomponent DNA constructs from acceptor and donor plasmid DNA modules that are conjoined by the Cre-LoxP fusion reaction[13,14]. Plasmid pSI-AGR10 is the common acceptor in MultiPrime, and has been developed from our previous pFL plasmid[25]. The ampicillin resistance gene and an internal SapI site were removed and the expression cassettes for insect-cell expression replaced by the CMV-based expression cassette from plasmid pSI-AKR1 by standard cloning methods (Supplementary Fig. 2a)[14]. In addition, pSI-AGR10 contains a gentamycin resistance marker, a LoxP site, and the DNA elements (Tn7R, Tn7L) required for transposition into the baculovirus genome by Tn7 transposase.

All donor plasmids of the original MultiLabel system are compatible with this acceptor[14]. Donors are fused to pSI-AGR10 by Cre-LoxP recombination concomitantly or in a sequential manner. Acceptor–donor assembly was performed as described and electrocompetent DH10β or $CaCl_2$ competent XL1-blue cells were used for transformation[42]. Sequences were assembled *in silico* using the 'Multi-Cre Recombination Tool' in the plasmid editor software Ape (http://biologylabs.utah.edu/jorgensen/wayned/ape/) or, alternatively, with software Cre-ACEMBLER[43]. Integrity of all fusion plasmids was confirmed by restriction mapping. Alternative mammalian promoters and dual-host promoters active in both mammalian and insect cells were synthesized by Genewiz (South Plainfield, USA) on the basis of sequences provided in the Supplementary Materials and inserted as AscI–HindIII fragments into parent Acceptor plasmid pSI-AGR10.

Recombinant baculoviral genomes. Two baculoviral genomes were used in this study, our previously described EMBacY genome and the novel MultiBacMam genome, which we constructed in this study. Both baculoviral genomes are present as bacterial artificial chromosomes (BAC) in *E. coli* cells (DH10EMBacY and DH10MultiBacMam, respectively). EMBacY produces yellow fluorescent protein (YFP) as a marker in infected insect cells as a means to track virus amplification and performance by monitoring the fluorescence signal[25].

Display of a VSV-G on the baculovirion was reported to enhance mammalian transduction efficiency by baculovirus[27]. We therefore constructed a novel MultiBacMam baculovirus by modifying our original MultiBac baculoviral genome, retaining its advantageous features including reduced proteolysis and delayed cell lysis[12]. A synthetic gene (Genscript, Psicataway, NJ) encoding for VSV-G was inserted into a modified pUCDM donor plasmid[25] by using BamHI and XbaI restriction sites to yield plasmid pLox-VSV-G. Subsequently, a second cassette containing a synthetic gene for mCherry (Genscript) was inserted by using the multiplication module as described[13]. The resulting pLox-VSV-G-mCherry Donor plasmid was incorporated into the MultiBac virus by transforming DH10MultiBac^{Cre} cells harbouring the MultiBac baculoviral genome as a BAC and Cre recombinase expressed from a pBADZHisCre helper plasmid on arabinose induction[12]. Positive integrands were selected by antibiotic screening. Successful Cre-mediated integration was further verified by PCR analysis as described[44]. Competent DH10MultiBacMam cells were prepared following standard protocols and contain in addition to the MultiBacMam baculovirus also a helper plasmid expressing Tn7 transposase on induction with isopropyl β-D-1-thiogalactopyranoside. Expression of mCherry from this baculovirus in infected insect cells during virion production results in the cell culture adopting a characteristic purple color, allowing tracking of successful viral infection and production easily by eye.

Generation of composite MultiPrime baculovirus. MultiPrime acceptors or acceptor–donor fusions were transformed into electrocompetent DH10EMBacY or DH10MultiBacMam cells, respectively. Composite baculovirus generation occurred by Tn7 tranposition mediated by Tn7 transposase expressed from a helper plasmid. Transformants were selected and composite baculoviral genomes prepared as described[44] Sf21 insect cells were transfected with Cellfectin II (Life Technologies) at a density of 0.5×10^6 cells ml^{-1} according to manufacturer's recommendations.

We took particular care during virus amplification to prevent accumulation of defective virus, which would not express all heterologous genes. We applied a protocol we had developed previously for successful amplification of composite baculovirus containing multiple copies of viral late promoters (polh, p10), preserving the integrity of the viral genome[28]. Briefly, primary baculovirus stock (V0, 2 ml) was harvested 50–60 h after transfection and 0.5 ml was used to infect 4 ml new Sf21 cells for 75 h yielding V1 stock. Overall 3 ml of this V1 baculovirus stock was then used for a further round of virus amplification for 60 h (V2, 100 ml). The amplification of the virus was followed in this phase by monitoring EYFP (EMBacY) or mCherry (MultibacMam) expression from the viral backbone (Supplementary Fig. 1). Less than 1% of cells were positive when harvesting V0. When harvesting V1, 20–30% of cells were positive and after V2, 80–90% of cells were positive. Incubation times must not be extended during amplification, otherwise over-amplification of the virus can occur, resulting in loss of heterologous insert (Supplementary Fig. 4). The V2 virus stocks were stored either at 4 °C or after addition of 5% FBS at − 80 °C. For sensitive cells (for example, iPS) or zebrafish, virus was concentrated by ultracentrifugation. For this purpose, virus supernatant was placed on a sucrose cushion (25% sucrose/ 5 mM NaCl/ 10 mM EDTA) and then centrifuged for 90 min at 80,000*g*. The pellet containing the virus was resuspended in PBS pH 6.2 (ref. 45). The titre of baculovirus stocks was determined using end-point dilution assay[46]. Viruses displaying VSV-G were handled as biosafety level 2 agents in Switzerland.

Cell culture. Insect cells (Sf21, Sf9; Life Technologies) were cultured in SF-4 BaculoExpress ICM medium (Amimed) containing 1% FBS at 27 °C. Mammalian cells were incubated at 37 °C in a humidified atmosphere containing 5% CO_2. HEK293, COS7, REF, Swiss 3T3 and HeLa cells were cultured in DMEM (Amimed) containing 10% FBS (Life Technologies) and 100 units ml^{-1} penicillin and 100 µg ml^{-1} Streptomycin (Life Technologies). PAE and HUVEC (Life Technologies) cells were maintained in Ham-F12 media supplemented with 10% FBS and penicillin/ streptomycin and M-200 medium (Life Technologies), respectively. Primary rat cortex neurons (Life Technologies) were cultured in Neurobasal medium supplemented with 1% B27 and 1% Glutamax (Life Technologies). Human iPS cells (NAS2) were obtained from Tilo Kunth (University of Edinburgh) and cultured in feeder-free maintenance medium for human ES/ iPS cells (mTESR1 medium; Stem Cell Technology)[47]. iPS cells were immunostained with an Oct4 antibody (Santa Cruz Biotechnology, sc-5279, 1:500 in 10% NDS/ 0.2% Triton X100/ 1% BST/ PBS), transfections were performed with FusionHD (Promega) according to manufacturer's recommendations and cells were analyzed 42 h after transfection. For endothelial tube formation assays, MultiPrime-transduced HUVEC cells were seeded on a ibiTreat µ-slide angiogenesis plate (ibidi GmbH, Germany) at a density of 5,000 cells/slide in EGM-2 medium (Life Technologies) and analyzed 16 h later. Migration assays were performed by seeding 21000 MultiPrime-transduced HUVEC cells in ibidi Culture-Insert plate (ibidi GmbH, Germany). The culture-insert was removed after 14 h and migration into the gap was monitored every 2 h for 24 h.

Transduction of mammalian cells. Mammalian cells were plated at a density of 2.5×10^5 cells per well in six-well plates 1 day before transduction. Baculovirus was added at a MOI between 100 and 500 in 80% insect medium/ 20% DMEM without any FBS or antibiotics. Transduced cells were incubated at 37 °C for 8 h, and the medium was then replaced with fresh mammalian cell culture medium. Plates were cultured for one or two additional days. Cytotoxicity of baculovirus transduction was monitored by MTT assay in various mammalian cells lines. Overall 10^4 cells were plated in a 96-well plate and transduced with baculovirus at a MOI 500. After 24 h, the medium was replaced with culture medium containing 20 µM resazurin and the cells were incubated for 2–4 h. The number of viable cells was obtained by monitoring resazurin fluorescence with a microplate spectrofluorometer (Tecan Ltd).

For CRISPR/Cas9-mediated genome engineering, HEK293 cells and HUVEC were transduced with a Multiprime virus that expresses CMV-driven Cas9, U6-driven gRNAs and a homology construct as described in Supplementary Fig. 9. Immunostaining for Cas9 after 40 h revealed 80% transduction. Cells were fixed after 4 or 6 days and analyzed for nuclear EGFP expression by microscopy or DNA was extracted with QIAamp DNA mini kit (Qiagen). Correct integration was verified by PCR using primers described in Supplementary Fig. 9.

MEF cells used for reprogramming to neurons were obtained from Amsbio, and were used at passage 3. Cells were transduced with concentrated bacuolovirus expressing Ascl1, Brn2 and Myt1L in MEF medium for 8 h. In parallel, MEFs were infected with lentiviruses containing expression constructs Tet-o-FUW-Ascl1, Tet-o-FUW-Brn2 and Tet-o-FUW-Myt1L (all from Addgene) in presence of polybrene (8 µg ml^{-1}). Cells were cultured in N3 medium (DMEM/F12, B27, N2 (all (Life Technologies), 25 µg ml^{-1} Insulin (Sigma-Aldrich)). Doxycycline (2 µg ml^{-1}) was added to lentivirus-transduced cells[7,48]. Cells were fixed after 6, 12 and 20 days. Immunostaining was performed with chicken anti MAP2 (Neuromics, CH22103, 1:5,000 in 10% NDS/ 0.2% Triton X100/ 1% BST/ PBS) and mouse anti β-tubulin III (Sigma, T8578, 1:600 in 10% NDS/ 0.2% Triton X100/ 1% BST/ PBS) antibodies.

Microscopy. Cells for microscopic analysis were plated on glass coverslips. Untreated coverslips were used for COS7, REF, Swiss 3T3 and PAE cells. Poly-L-lysine (Sigma P4707) treated coverslips were used for HEK293 cells and 0.1% gelatin (Sigma G1393) treated coverslips for HUVEC. Poly-D-lysine hydrobromide coated coverslips (Sigma P7280) were used for primary rat cortex neurons and iPS were plated on hESC qualified Matrigel (BD Bioscence). Analysis of cells was performed 27, 42 and 48 h after transduction. Cells were fixed with 4% formaldehyde in PBS and mounted with Gelvatol. Imaging was performed on a Leica SP5 laser scanning confocal microscope or on an Olympus IX81 equipped with an Andor iXonEM camera. On Leica SP5, EBFP2 was excited with the 405 nm laser line and the emission was collected from 430 to 450 nm (405/ 430-450). The other fluorescent proteins were analyzed as follows: mTFP1 (458/ 485-510), mCitrine (514/ 525-545), mCherry (543/ 585-620) and mPLUM (633/ 640-800). In addition, the spectral mode (xyλ) of the microscope was used to verify the presence of all fluorescent proteins (data not shown). Standard excitation and emission filters were used on the Olympus IX81. Quantification was performed with Squassh[49].

Western blotting. Mammalian cells were lysed 42 h after transduction with lysis buffer (0.5% Triton X100, 50 mM Tris–HCl, 100 mM NaCl, pH 7.5). The supernatant was used for western blotting after sonification and centrifugation. Rabbit anti-GFP (Abcam ab137827; diluted 1:2500 in 3% BSA/TBST) and mouse anti-tubulin (Sigma T5168; diluted 1:2,500 in 3% BSA/TBST) were used as primary antibodies. As secondary antibodies, alkaline phosphatase-coupled goat anti-rabbit and anti-mouse as well as donkey anti-human IgGs (Southern Biotech, diluted 1:10,000 in TBST) were used, followed by chemiluminescence detection. Quantification was performed with ImageJ. Original western blots are shown in Supplementary Fig. 10.

Transduction of zebrafish embryos. All experiments were performed in accordance with the animal welfare guidelines of the Federal Veterinary Office of Switzerland. Zebrafish (*Danio rerio*) were maintained as described[50]. Embryos of the wild-type strain WIK were raised at 28°C in E3 medium (5 mM NaCl, 0.17 mM KCl, 0.33 mM $CaCl_2$ and 0.33 mM $MgSO_4$), and pigment development was inhibited by phenylthiourea (1-phenyl-2-thiourea; Sigma-Aldrich) as described in Westerfield[50]. For injections, individual dechorionated embryos at 24 h post fertilization were anesthetized in 200 mg ml^{-1} 3-aminobenzoic acid methyl ester (MESAB, Sigma-Aldrich) and 4.6 mM $NaHCO_3$ and embedded in 1% low melting agarose (Lonza). Concentrated viruses were injected in intercellular spaces using borosilicate glass microcapillary injection needles (1 mm outside diameter × 0.78 mm inside diameter, Science Products GmbH, Hofheim, Germany) and a PV820 Pneumatic PicoPump (World Precision Instruments, Sarasota, Florida, USA). After injection, the infected embryos were returned to E3 medium. For in vivo imaging, the larvae were anesthetized and embedded in 1% low melting agarose in a 35 mm imaging dish with a glass bottom (Ibidi) and imaged using a CLSM SP5 Mid UV–vis Leica inverted microscope.

Bifunctional CMV/ P10 dual-host promoters. Subcomplexes of human general transcription factor TFIID were produced in insect cells from the CMV-CMVintP10 promoter (Supplementary Fig. 7). A complex formed by human TBP associated factor (TAF) 8 and TAF10, and a complex formed by TAF5, TAF6 and TAF9, were expressed. Production levels of these complexes on the basis of the dual-host promoter was compared to previously established production levels on the basis of the baculoviral polyhedrin (polh) promoter. A polyprotein strategy was utilized[51], which we had developed for high-level expression of complexes[30,31]. Briefly, the genes encoding for the TAFs were placed in a single ORF flanked by genes encoding for tobacco etch virus NIa (TEV) protease at 5′ and a cyan fluorescent protein at the 3′-end[51]. The ORFs give rise to self-processing polyproteins, which are cleaved by TEV protease at high specific TEV protease cleavage sites in between the constituent proteins. The polyproteins are shown schematically in Supplementary Fig. 8.

Polyproteins were expressed from the CMVintP10 and the polh promoter, respectively, by using EMBacY, and purified as described[30,31]. Production levels of the polyproteins in insect cells were indistinguishable notwithstanding the promoter used (Supplementary Fig. 8). Next, the MultiBacMam virus (see above) was used in conjunction with the CMVintP10 promoter, again leading to indistinguishable expression levels (Supplementary Fig. 8). HeLa cell cultures were transfected with MultiBacMam-derived baculoviruses that had been used to express the proteins in insect cells. Complete DMEM containing 10% foetal calf serum and 8 mM L-glutamine was utilized. Best results were achieved by supplementing the media with 3 mM sodium butyrate for cell recovery after aspirating the virus. Nearly all HeLa cells were transduced as revealed by measuring the specific fluorescence of the cyan fluorescent protein marker encoded by the polyprotein (Supplementary Fig. 8).

References

1. Kretzschmar, K. & Watt, F. M. Lineage tracing. *Cell* **148,** 33–45 (2012).
2. Livet, J. *et al.* Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature* **450,** 56–62 (2007).

3. Ballmer-Hofer, K., Andersson, A. E., Ratcliffe, L. E. & Berger, P. Neuropilin-1 promotes VEGFR-2 trafficking through Rab11 vesicles thereby specifying signal output. *Blood* **118,** 816–826 (2011).
4. Trowitzsch, S., Klumpp, M., Thoma, R., Carralot, J. P. & Berger, I. Light it up: highly efficient multigene delivery in mammalian cells. *Bioessays* **33,** 946–955 (2011).
5. Marsic, D. *et al.* Vector design Tour de Force: integrating combinatorial and rational approaches to derive novel adeno-associated virus variants. *Mol. Ther* **22,** 1900–1909 (2014).
6. Takahashi, K. & Yamanaka, S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell* **126,** 663–676 (2006).
7. Vierbuchen, T. *et al.* Direct conversion of fibroblasts to functional neurons by defined factors. *Nature* **463,** 1035–1041 (2010).
8. Church, G. M., Elowitz, M. B., Smolke, C. D., Voigt, C. A. & Weiss, R. Realizing the potential of synthetic biology. *Nat. Rev. Mol. Cell. Biol.* **15,** 289–294 (2014).
9. 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).
10. Barford, D., Takagi, Y., Schultz, P. & Berger, I. Baculovirus expression: tackling the complexity challenge. *Curr. Opin. Struct. Biol.* **23,** 357–364 (2013).
11. Mansouri, M. & Berger, P. Strategies for multigene expression in eukaryotic cells. *Plasmid* **75,** 12–17 (2014).
12. Berger, I., Fitzgerald, D. J. & Richmond, T. J. Baculovirus expression system for heterologous multiprotein complexes. *Nat. Biotechnol.* **22,** 1583–1587 (2004).
13. Bieniossek, C. *et al.* Automated unrestricted multigene recombineering for multiprotein complex production. *Nat. Methods* **6,** 447–450 (2009).
14. Kriz, A. *et al.* A plasmid-based multigene expression system for mammalian cells. *Nat. Commun.* **1,** 120 (2010).
15. Haffke, M., Viola, C., Nie, Y. & Berger, I. Tandem recombineering by SLIC cloning and Cre-LoxP fusion to generate multigene expression constructs for protein complex research. *Methods Mol. Biol.* **1073,** 131–140 (2013).
16. Vijayachandran, L. S. *et al.* Robots, pipelines, polyproteins: enabling multiprotein expression in prokaryotic and eukaryotic cells. *J. Struct. Biol.* **175,** 198–208 (2011).
17. Sokolenko, S. *et al.* Co-expression vs co-infection using baculovirus expression vectors in insect cell culture: Benefits and drawbacks. *Biotechnol. Adv.* **30,** 766–781 (2012).
18. Anliker, B. *et al.* Specific gene transfer to neurons, endothelial cells and hematopoietic progenitors with lentiviral vectors. *Nat. Methods* **7,** 929–935 (2010).
19. Kost, T. A., Condreay, J. P. & Jarvis, D. L. Baculovirus as versatile vectors for protein expression in insect and mammalian cells. *Nat. Biotechnol.* **23,** 567–575 (2005).
20. Ames, R. S., Kost, T. A. & Condreay, J. P. BacMam technology and its application to drug discovery. *Expert Opin. Drug Discov.* **2,** 1669–1681 (2007).
21. Kost, T. A., Condreay, J. P. & Ames, R. S. Baculovirus gene delivery: a flexible assay development tool. *Curr. Gene. Ther.* **10,** 168–173 (2010).
22. Hofmann, C. *et al.* Efficient gene transfer into human hepatocytes by baculovirus vectors. *Proc. Natl Acad. Sci. USA* **92,** 10099–10103 (1995).
23. Boyce, F. M. & Bucher, N. L. Baculovirus-mediated gene transfer into mammalian cells. *Proc. Natl Acad. Sci. USA* **93,** 2348–2352 (1996).
24. Airenne, K. J. *et al.* Baculovirus: an insect-derived vector for diverse gene transfer applications. *Mol. Ther.* **21,** 739–749 (2013).
25. Fitzgerald, D. J. *et al.* Protein complex expression by using multigene baculoviral vectors. *Nat. Methods* **3,** 1021–1032 (2006).
26. Luckow, V. A., Lee, S. C., Barry, G. F. & Olins, P. O. Efficient generation of infectious recombinant baculoviruses by site-specific transposon-mediated insertion of foreign genes into a baculovirus genome propagated in *Escherichia coli*. *J. Virol.* **67,** 4566–4579 (1993).
27. Barsoum, J., Brown, R., McKee, M. & Boyce, F. M. Efficient transduction of mammalian cells by a recombinant baculovirus having the vesicular stomatitis virus G glycoprotein. *Hum. Gene. Ther.* **8,** 2011–2018 (1997).
28. Berger, I. *et al.* The multiBac protein complex production platform at the EMBL. *J. Vis. Exp.* **77,** e50159 (2013).
29. Gossen, M. & Bujard, H. Tight control of gene expression in mammalian cells by tetracycline-responsive promoters. *Proc. Natl Acad. Sci. USA* **89,** 5547–5551 (1992).
30. Bieniossek, C. *et al.* The architecture of human general transcription factor TFIID core complex. *Nature* **493,** 699–702 (2013).
31. Trowitzsch, S. *et al.* Cytoplasmic TAF2-TAF8–TAF10 complex provides evidence for nuclear holo-TFIID assembly from preformed submodules. *Nat. Commun.* **6,** 6011 (2015).
32. Marx, V. Cell biology: delivering tough cargo into cells. *Nat. Methods* **13,** 37–40 (2016).
33. Ratz, M., Testa, I., Hell, S. W. & Jakobs, S. CRISPR/Cas9-mediated endogenous protein tagging for RESOLFT super-resolution microscopy of living human cells. *Sci. Rep.* **5,** 9592 (2015).
34. Rubio Demirovic, A. *et al.* Targeting human cancer cells with VEGF receptor-2-directed liposomes. *Oncol. Rep.* **13,** 319–324 (2005).
35. Zeisberger, S. M. *et al.* Clodronate-liposome-mediated depletion of tumour-associated macrophages: a new and highly effective antiangiogenic therapy approach. *Br. J. Cancer* **95,** 272–281 (2006).
36. Wagle, M. & Jesuthasan, S. Baculovirus-mediated gene expression in zebrafish. *Mar. Biotechnol.* **5,** 58–63 (2003).
37. van Oers, M. M., Pijlman, G. P. & Vlak, J. M. Thirty years of baculovirus-insect cell protein expression: from dark horse to mainstream technology. *J. Gen. Virol.* **96,** 6–23 (2015).
38. Palmberger, D., Wilson, I. B., Berger, I., Grabherr, R. & Rendic, D. SweetBac: a new approach for the production of mammalianised glycoproteins in insect cells. *PLoS ONE* **7,** e34226 (2012).
39. Palmberger, D., Klausberger, M., Berger, I. & Grabherr, R. MultiBac turns sweet. *Bioengineered* **4,** 78–83 (2013).
40. Heikura, T. *et al.* Baculovirus-mediated vascular endothelial growth factor-D(DeltaNDeltaC) gene transfer induces angiogenesis in rabbit skeletal muscle. *J. Gene Med.* **14,** 35–43 (2012).
41. Takata, Y. *et al.* Generation of iPS cells using a BacMam multigene expression system. *Cell Struct. Funct.* **36,** 209–222 (2011).
42. Kriz, A., Schmid, K., Ballmer-Hofer, K. & Berger, P. Integration of multiple expression cassettes into mammalian genomes in a single step. *Protocol Exchange*, doi:10.1038/protex.2011.249 (2011).
43. Becke, C., Haffke, M. & Berger, I. Ce-ACEMBLER Software User Manual. 10.13140/2.1.1068.11278; https://github.com/christianbecke/Cre-ACEMBLER (2012).
44. Bieniossek, C., Richmond, T. J. & Berger, I. MultiBac: multigene baculovirus-based eukaryotic protein complex production. *Curr. Protoc. Protein Sci.* Chapter **5,** Unit 5 20 (2008).
45. O'Reilly, D. R., Miller, L. K. & Luckov, V. A. *Baculovirus Expression Vectors: A laboratory Manual* (Oxford University Press, 1994).
46. Dee, K. U. & Shuler, M. L. Optimization of an assay for baculovirus titer and design of regimens for the synchronous infection of insect cells. *Biotechnol. Prog.* **13,** 14–24 (1997).
47. Devine, M. J. *et al.* Parkinson's disease induced pluripotent stem cells with triplication of the alpha-synuclein locus. *Nat. Commun.* **2,** 440 (2011).
48. Chanda, S. *et al.* Generation of induced neuronal cells by the single reprogramming factor ASCL1. *Stem cell Rep.* **3,** 282–296 (2014).
49. Rizk, A. *et al.* Segmentation and quantification of subcellular structures in fluorescence microscopy images using Squassh. *Nat. Protoc.* **9,** 586–596 (2014).
50. Westerfield, M. *The Zebrafish Book, A guide for the laboratory use of zebrafish (Danio rerio)*. 5th edn (University of Oregon Press, 2007).
51. Nie, Y., Bellon-Echeverria, I., Trowitzsch, S., Bieniossek, C. & Berger, I. Multiprotein complex production in insect cells by using polyproteins. *Methods Mol. Biol.* **1091,** 131–141 (2014).

Acknowledgements

This work was supported by the Swiss National Science Foundation (31003A-146975 and 3100A0-118351 to P.B.), the European Commission (EC) Framework programme (FP) seven projects ComplexINC, SynSignal and Biostruct-X (contracts nr. 279039, 613879 and 283570 to I.B.) and a fellowship of the Iranian Ministry of Science (to M.M.). We acknowledge the assistance and support of the Center for Microscopy and Image Analysis, University of Zürich for performing zebrafish *in vivo* imaging. We would like to thank Mayanka Asthana for help with VEGF labelling and Dr Stefan Jakobs for providing gRNA plasmids.

Author contributions

I.B. and P.B. conceived the study. M.M., I.B.E., Z.E., C.C.C., C.S.S., Y.X. and P.B. performed experiments; A.R. and M.W.D. performed data analysis or developed software; M.M., F.M.B., S.C.F.N., V.T., K.B-H., I.B. and P.B. designed experiments; M.M., K.B-H., I.B. and P.B. wrote the paper with input from all authors.

Additional information

Accession codes: Nucleotide sequences have been deposited in GenBank under the following accession codes: pSI-AGR10: KX001770; pSI-AGZ10: KX001771; pSI-DSZ2cx-TRE: KX001772; pSI-DSZ2cx-DI: KX001773; pSI-DSZ2x-DM: KX001774; pSI-DSZ2cx-PGK: KX001775; pSI-DSZ2cx-UBC: KX001776; pSI-DSZcx-SV40: KX001777; pSI-AGL10: KX001778; pSI-AGL10-DM: KX001779; pSI-AGL10-DI: KX001780; pSI-DAZ-DM: KX001781; pSI-DAZ-DI: KX001782.

Supplementary Information accompanies this paper at http://www.nature.com/naturecommunications

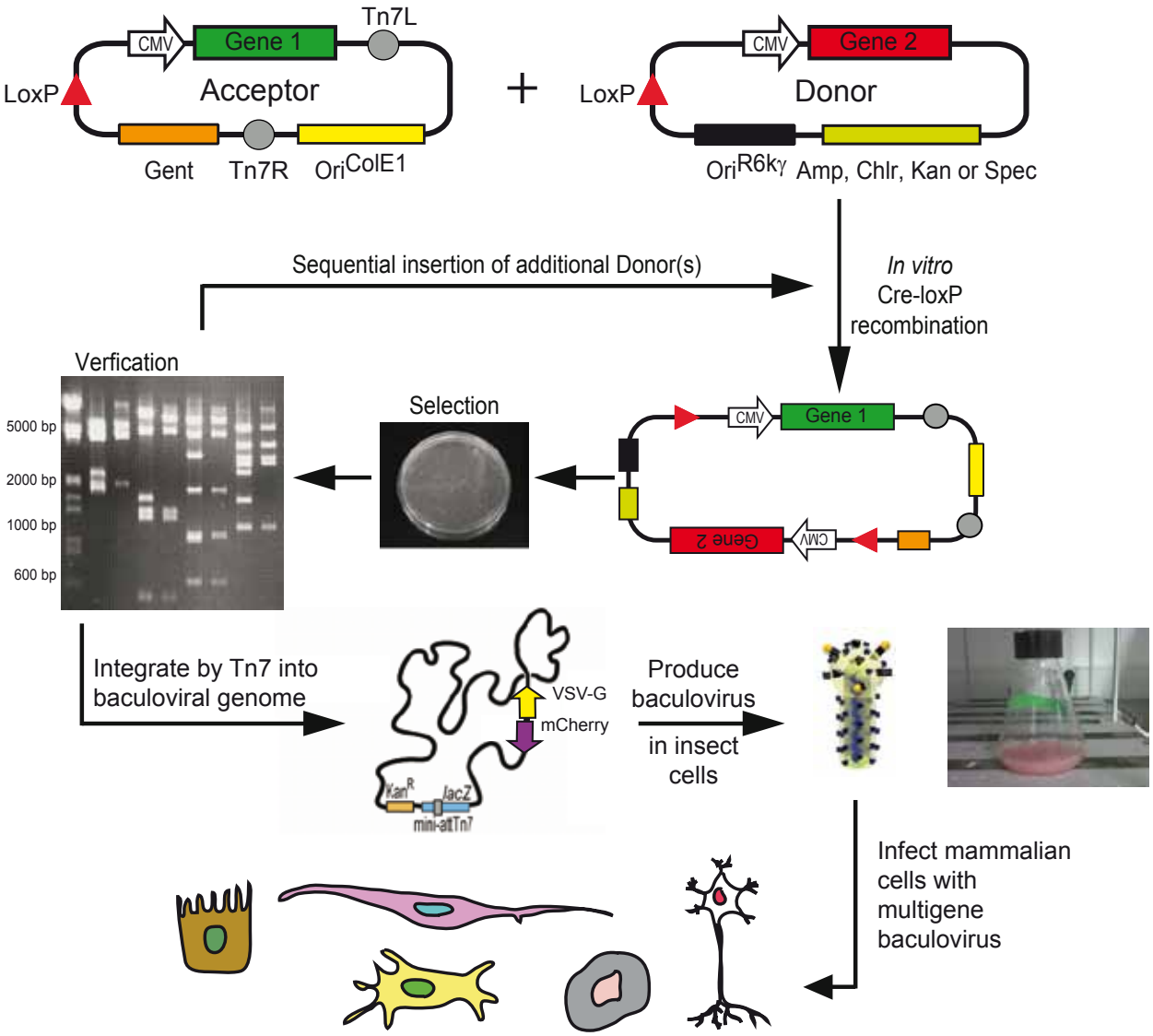
Competing financial interests: The authors declare competing financial interest. I.B. is inventor on patents that comprise components of the systems here described (EP 1945773, EP 2403940).

Reprints and permission information is available online at http://npg.nature.com/reprintsandpermissions/

How to cite this article: Mansouri, M. *et al.* Highly efficient baculovirus-mediated multigene delivery in primary cells. *Nat. Commun.* 7:11529 doi: 10.1038/ncomms11529 (2016).

Supplementary Figure 1

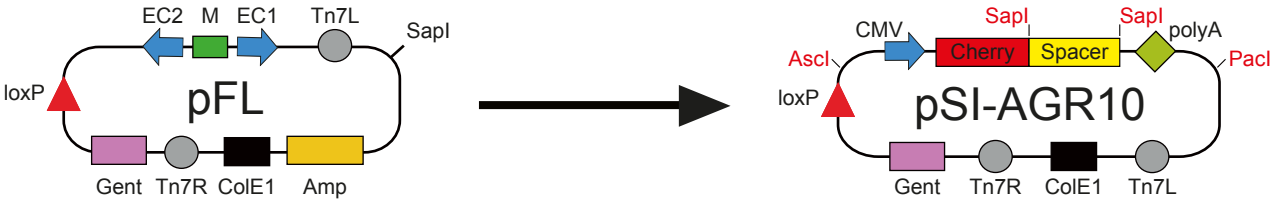
Workflow of MultiPrime



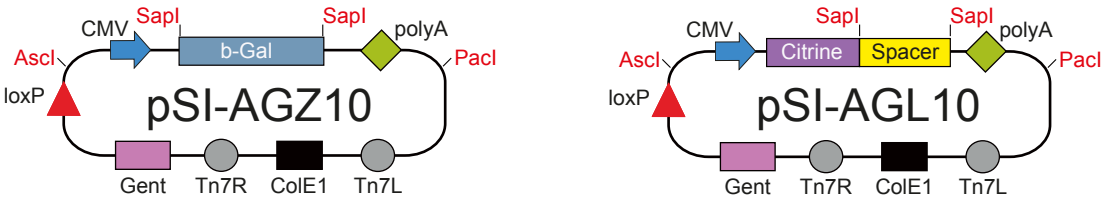
Supplementary Figure 1: Schematic presentation of the MultiPrime workflow.
One Acceptor is fused to one or several Donors (each with a different resistance marker). Acceptor and Donors can each contain one or several genes of interest (one gene shown for clarity). Acceptors contain the Tn7L and Tn7R sequences for transposition mediated by Tn7 transposase. Acceptor-Donor fusions are selected on plates with appropriate antibiotics, and verified by restriction mapping. Cre mediated fusion of further Donors, each with one or several genes of interest, can be performed iteratively. Verified Acceptor-Donor fusions containing all genes and DNA elements (regulatory elements, enhancers, etc) of interest are then integrated into a baculoviral genome (e.g. EMBacY or MultiBacMam) *in vivo* by Tn7 transposition in bacterial cells. Integrands are selected by blue-white screening, and baculovirus produced in insect cell cultures displaying a red or yellow color depending on the used bacmid. MultiBacMam derived baculovirions display a VSV-G protein (yellow spheres) resulting in enhanced infectivity. Recombinant baculovirus is then harvested and can be used to infect a wide range of mammalian cells with high efficiency.
Abbreviations: Gent: Gentamycin; Amp: Ampicillin; Chlr: Chloramphenicol; Kan: Kanamycin; Spec: Spectinomycin; CMV: Cytomegalovirus promoter; Tn7R, Tn7L: Tn7 tranposition elements; loxP: Cre recognition site; ori: origin of replication.

Supplementary Figure 2

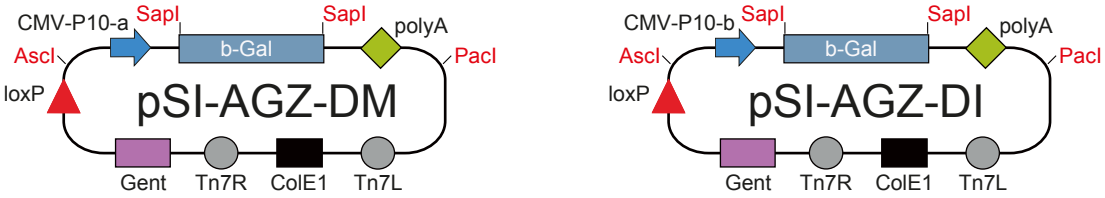
A) Development of the initial Acceptor vector



B) Additional CMV-based Acceptors



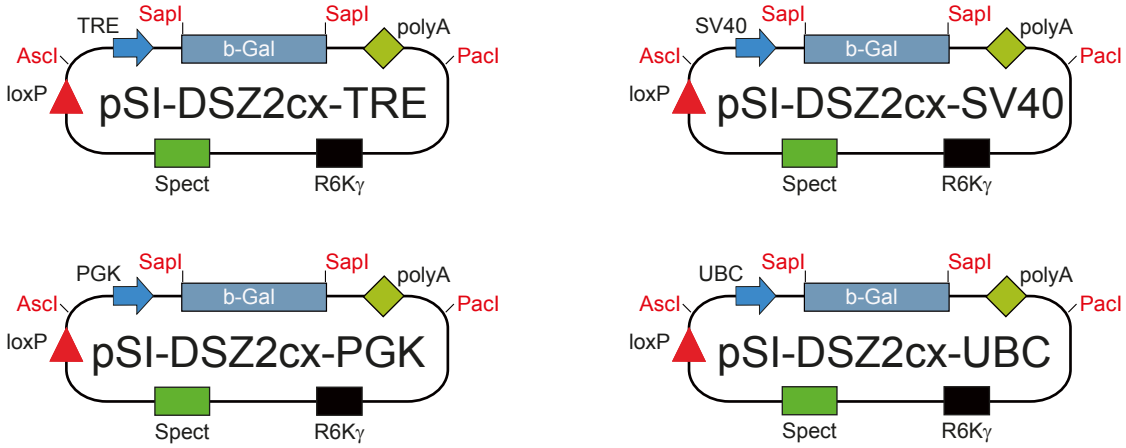
C) Acceptor with dual-host promoters



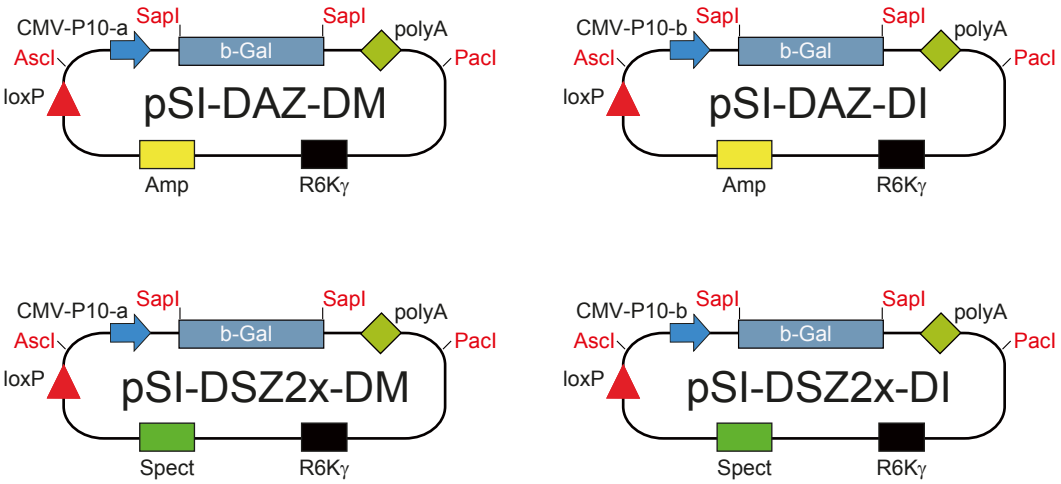
Supplementary Figure 2: MultiPrime compatible Acceptor vectors.
(A) Development of the initial MultiPrime Acceptor vector. This Acceptor plasmid is based on the previously described pFL vector (Fitzgerald et al., 2006). The following changes were made: (i) the ampicillin resistance gene was removed, (ii) an internal SapI site was deleted, and (iii) a mammalian expression cassette was inserted. The mammalian expression cassette derived from the previously described pSI-AKR1 plasmid (Kriz et al., 2010). The cloning sites (red) are compatible within the MultiLabel plasmid family. (B) Additional Acceptors with CMV promoter. (C) Acceptors with dual-host promoters for expression in insect and mammalian cells.
Abbreviations: Gent: Gentamycin; Amp: Ampicillin; M: Multiplication module; EC1,EC2: original pFL expression cassettes; CMV: Cytomegalovirus promoter; polyA: polyadenylation site; lacZ: Marker gene for blue/white screening; Tn7R, Tn7L: Tn7 tranposition elements; loxP: Cre recognition site

Supplementary Figure 3

A) Donors with alternative promoters



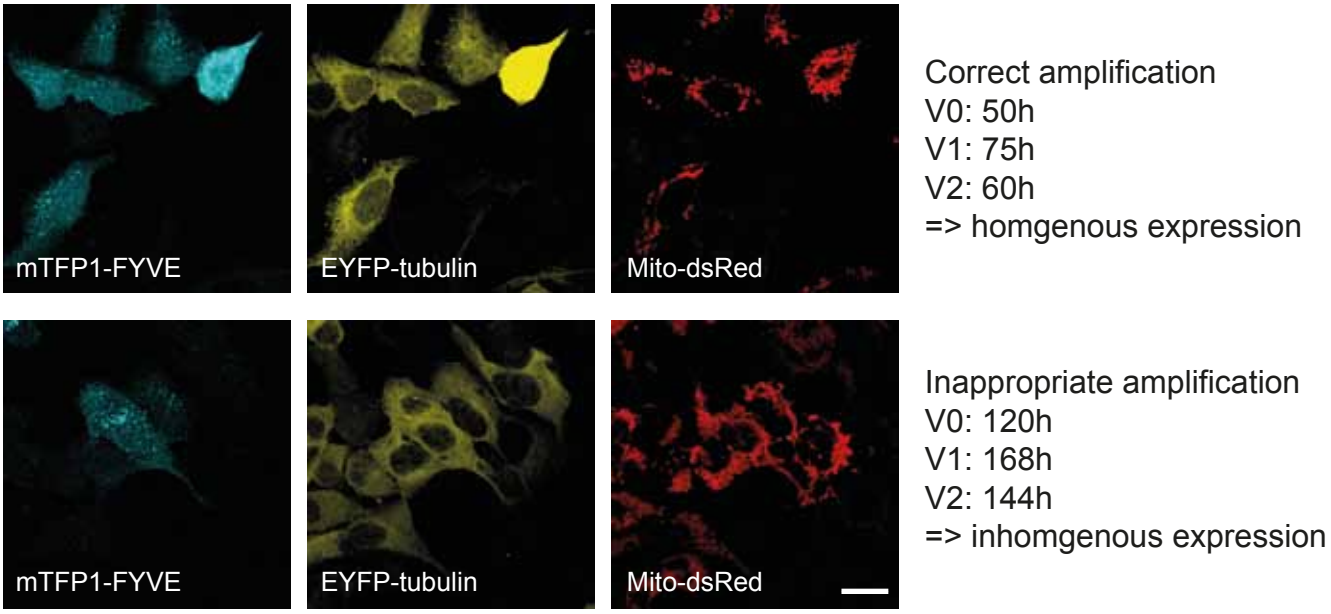
B) Donors with dual-host promoters



Supplementary Figure 3: MultiPrime compatible Donor vectors.
(A) Donor vectors with alternative promoters for expression in mammalian cells. TRE: tetracycline-inducible promoter; SV40: early promoter from simian virus 40; PGK: phosphoglycerate kinase promoter; UBC: ubiquitin promoter. Please note that all previously described CMV-based promoters are also compatible with MultiPrime (Kriz et al., 2010). (B) Donors with dual-host promoters suitable for insect and mammalian cell expression. Abbreviations: Amp: Ampicillin; Spect: Spectinomycin; b-Gal: spacer with expression of b-galactosidase; R6Kγ: conditional origin of replication; CMV: Cytomegalovirus promoter; polyA: polyadenylation site; loxP: Cre recognition site.

Supplementary Figure 4

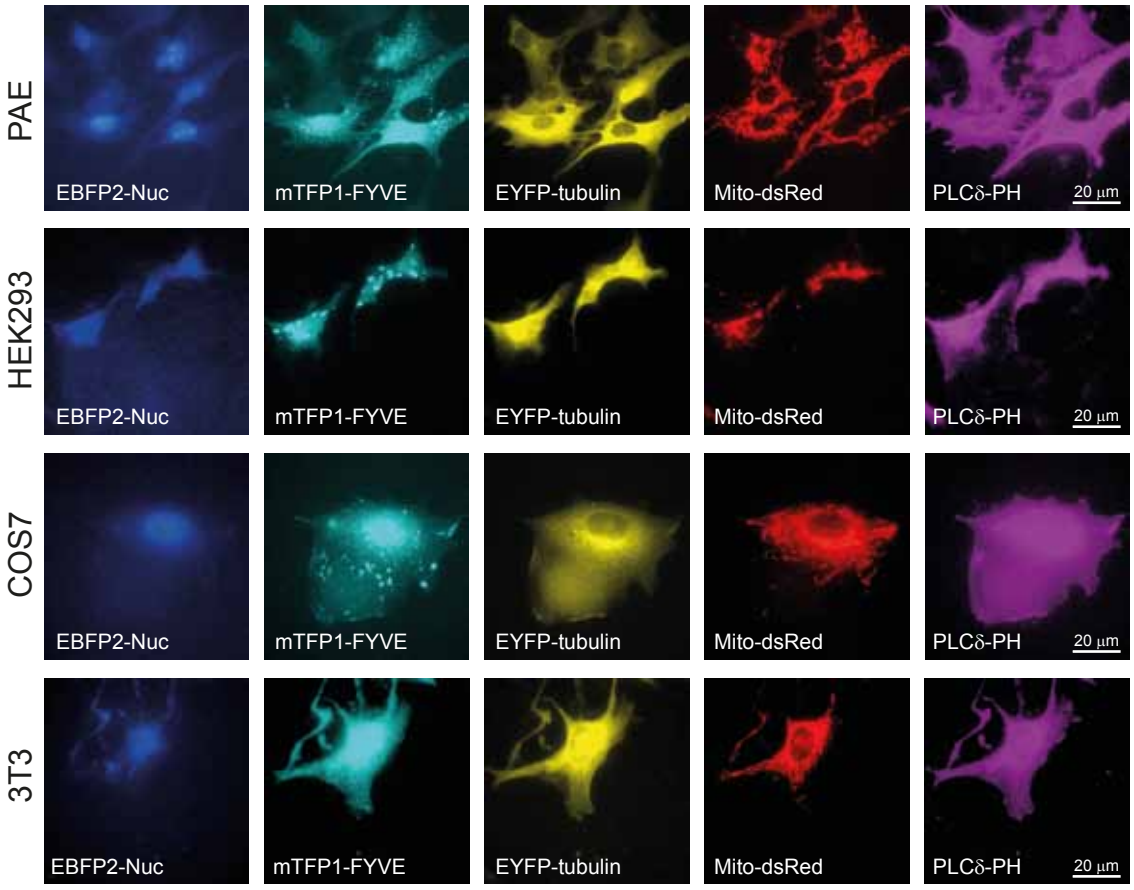
Importance of the virus amplification protocol

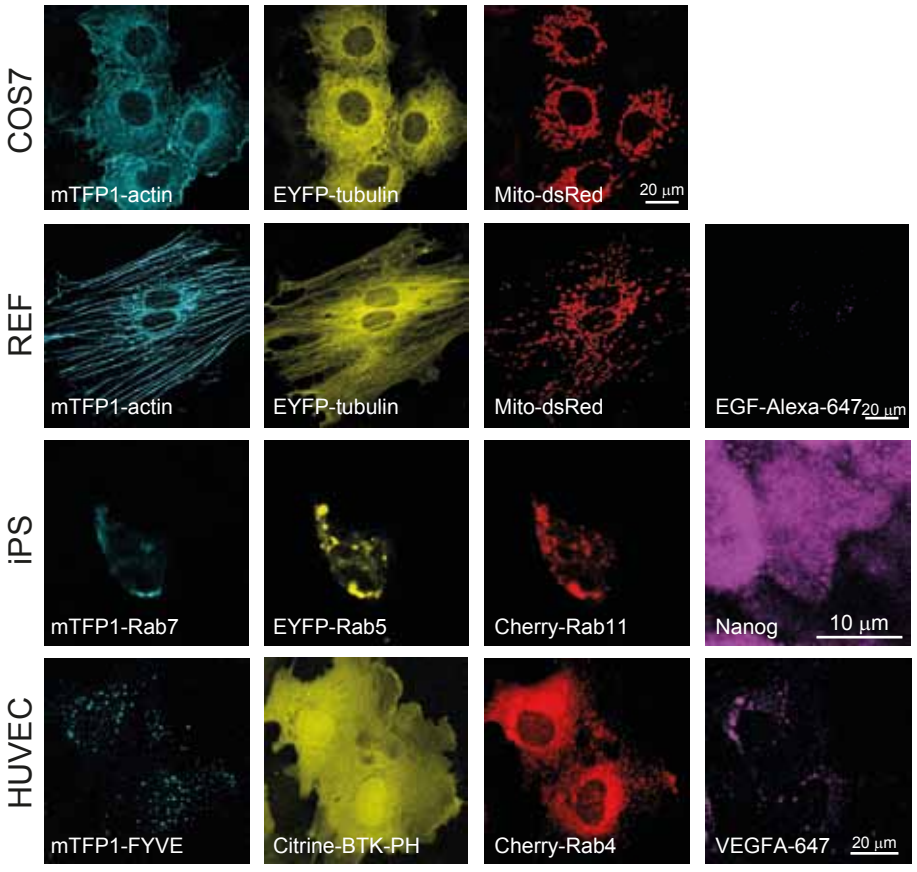


Supplementary Figure 4: Importance of virus amplification protocol
(Top) Virus amplification according to the protocol in Materials and Methods leads to a homogenous HEK293 cell population. (Bottom) Virus amplification with prolonged incubation times leads to rearrangement of the baculovirus genome and inhomogenous HEK293 cell population after transduction. Scale bar: 20 μm.

Supplementary Figure 5

Additional cell types and constructs





Supplementary Figure 6

Sequences of alternative promoters

Tet responsive promoter (TRE)

GGCGCGCCTAGTTATTAATAGTAATCAATTACGGGGTCGAGCTCGACTTTCACTTTTCTCTATCACTGATAGGGAGTGGTAAACT
CGACTTTCACTTTTCTCTATCACTGATAGGGAGTGGTAAACTCGACTTTCACTTTTCTCTATCACTGATAGGGAGTGGTAAACTC
GACTTTCACTTTTCTCTATCACTGATAGGGAGTGGTAAACTCGACTTTCACTTTTCTCTATCACTGATAGGGAGTGGTAAACTCG
ACTTTCACTTTTCTCTATCACTGATAGGGAGTGGTAAACTCGACTTTCACTTTTCTCTATCACTGATAGGGAGTGGTAAACTCGA
CCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACCG
GGACCGATCCAGCCTCCGCGGCCCCGAATTGAAGCTT

Mouse phosphoglycerate kinase (PGK)

GGCGCGCCTAGTTATTAATAGTAATCAATTACGGGGTCTTCTACCGGGTAGGGGAGGCGCTTTTCCCAAGGCAGTCTGGAGCATG
CGCTTTAGCAGCCCCGCTGGGCACTTGGCGCTACACAAGTGGCCTCTGGCCTCGCACACATTCCACATCCACCGGTAGGCGCCAA
CCGGCTCCGTTCTTTGGTGGCCCCTTCGCGCCACCTTCTACTCCTCCCCTAGTCAGGAAGTTCCCCCCCGCCCCGCAGCTCGCGT
CGTGCAGGACGTGACAAATGGAAGTAGCACGTCTCACTAGTCTCGTGCAGATGGACAGCACCGCTGAGCAATGGAAGCGGGTAGG
CCTTTGGGGCAGCGGCCAATAGCAGCTTTGCTCCTTCGCTTTCTGGGCTCAGAGGCTGGGAAGGGGTGGGTCCGGGGGCGGGCTC
AGGGGCGGGCTCAGGGGCGGGGCGGGCGCCCGAAGGTCCTCCGGAGGCCCGGCATTCTGCACGCTTCAAAAGCGCACGTCTGCCG
CGCTGTTCTCCTCTTCCTCATCTCCGGGCCTTTCGACCTAAGCTT

Human Ubiquitin C promoter

GGCGCGCCTAGTTATTAATAGTAATCAATTACGGGGTCGGTGCAGCGGCCTCCGCGCCGGGTTTTGGCGCCTCCCGCGGGCGCCC
CCCTCCTCACGGCGAGCGCTGCCACGTCAGACGAAGGGCGCAGGAGCGTTCCTGATCCTTCCGCCCGGACGCTCAGGACAGCGGC
CCGCTGCTCATAAGACTCGGCCTTAGAACCCCAGTATCAGCAGAAGGACATTTTAGGACGGGACTTGGGTGACTCTAGGGCACTG
GTTTTCTTTCCAGAGAGCGGAACAGGCGAGGAAAAGTAGTCCCTTCTCGGCGATTCTGCGGAGGGATCTCCGTGGGGCGGTGAAC
GCCGATGATTATATAAGGACGCGCCGGGTGTGGCACAGCTAGTTCCGTCGCAGCCGGGATTTGGGTCGCGGTTCTTGTTTGTGGA
TCGCTGTGATCGTCACTTGGTGAGTTGCGGGCTGCTGGGCTGGCCGGGGCTTTCGTGGCCGCCGGGCCGCTCGGTGGGACGGAAG
CGTGTGGAGAGACCGCCAAGGGCTGTAGTCTGGGTCCGCGAGCAAGGTTGCCCTGAACTGGGGGTTGGGGGGAGCGCACAAAATG
GCGGCTGTTCCCGAGTCTTGAATGGAAGACGCTTGTAAGGCGGGCTGTGAGGTCGTTGAAACAAGGTGGGGGGCATGGTGGGCGG
CAAGAACCCAAGGTCTTGAGGCCTTCGCTAATGCGGGAAAGCTCTTATTCGGGTGAGATGGGCTGGGGCACCATCTGGGGACCCT
GACGTGAAGTTTGTCACTGACTGGAGAACTCGGGTTTGTCGTCTGGTTGCGGGGGCGGCAGTTATGCGGTGCCGTTGGGCAGTGC
ACCCGTACCTTTGGGAGCGCGCGCCTCGTCGTGTCGTGACGTCACCCGTTCTGTTGGCTTATAATGCAGGGTGGGGCCACCTGCC
GGTAGGTGTGCGGTAGGCTTTTCTCCGTCGCAGGACGCAGGGTTCGGGCCTAGGGTAGGCTCTCCTGAATCGACAGGCGCCGGAC
CTCTGGTGAGGGGAGGGATAAGTGAGGCGTCAGTTTCTTTGGTCGGTTTTATGTACCTATCTTCTTAAGTAGCTGAAGCTCCGGT
TTTGAACTATGCGCTCGGGGTTGGCGAGTGTGTTTTGTGAAGTTTTTTAGGCACCTTTTGAAATGTAATCATTTGGGTCAATATG
TAATTTTCAGTGTTAGACTAGTAAAGCTT

Simian virus 40 early promoter (SV40)

Derived from pcDNA 3.1 using PCR

GGCGCGCCTAGTTATTAATAGTAATCAATTACGGGGTCCTGTGGAATGTGTGTCAGTTAGGGTGTGGAAAGTCCCCAGGCTCCCC
AGCAGGCAGAAGTATGCAAAGCATGCATCTCAATTAGTCAGCAACCAGGTGTGGAAAGTCCCCAGGCTCCCCAGCAGGCAGAAGT
ATGCAAAGCATGCATCTCAATTAGTCAGCAACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCG
CCCATTCTCCGCCCCATGGCTGACTAATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCTGCCTCTGAGCTATTCCAGAAGTA
GTGAGGAGGCTTTTTTGGAGGCCTAGGCTTTTGCAAAAAGAAGCTT

SV40-forward: GATCGGCGCGCCTAGTTATTAATAGTAATCAATTACGGGGTCCTGTGGAATGTGTGTCAG
SV40-reverse: GATCAAGCTTCTTTTTGCAAAAGCCTAGG

Supplementary Figure 6: Alternative mammalian promoter sequences.

The DNA sequences of promoters used as an alternative to CMV in this study are listed, including the Tet responsive promoter (TRE), the mouse phosphoglycerate kinase promoter (PGK), the human ubiquitin C promoter and the simian virus 40 early promoter (SV40). Insertion of promoters into the MultiPrime plasmids was done usin AscI (red) and HindIII (green).

Supplementary Figure 7

Sequences of dual-host promoters

CMVP10 (based on Philipps B. et al, Biotechnol. Prog., 2005)

Black: CMV sequence
Blue: P10 sequence

GGCGCGCCTAGTtattaatagtaatcaattacggggtcattagttcatagcccatatatggagttccgcgttacataacttacgg
taaatggcccgcctggctgaccgcccaacgacccccgcccattgacgtcaataatgacgtatgttcccatagtaacgccaatagg
gactttccattgacgtcaatgggtggagtatttacggtaaactgcccacttggcagtacatcaagtgtatcatatgccaagtccg
cccccattgacgtcaatgacggtaaatggcccgcctggcattatgcccagtacatgaccttacgggactttcctacttggcagt
acatctacgtattagtcatcgctattaccatgctgatgcggttttggcagtacaccaatgggcgtggatagcggtttgactcacg
gggatttccaagtctccaccccattgacgtcaatgggagtttgttttggcaccaaaatcaacgggactttccaaaatgtcgtaat
aaccccgccccgttgacgcaaatgggcggtaggcgtgtacggtgggaggtctatataagcagacgtcgtttagtgaaccgtcaga
tcaccatccagtatacggacctttaattcaacccaacacaatatattatagttaaataagaattattatcaaatcatttgtatat
taattaaaatactatactgtaaattacattttatttacaatcactcgacgaagacttgatcacccgggatctcgaAAGCTT

CMVintP10 (including an intron with the P10 promoter)

Black: CMV sequence
Blue: P10 sequence
yellow: intron sequence

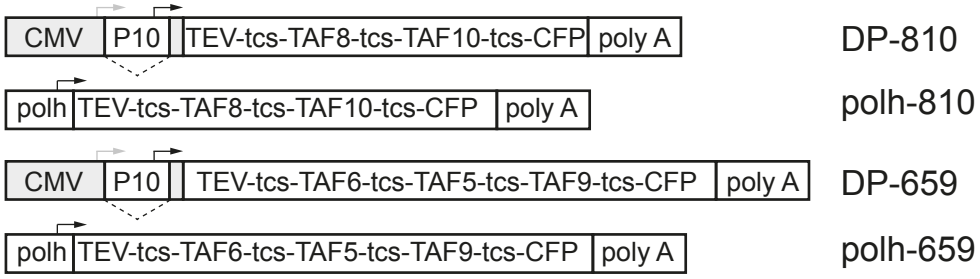
GGCGCGCCTAGTTATTAATAGTAATCAATTACGGGGTCATTAGTTCATAGCCCATATATGGAGTTCCGCGTTACATAACTTACGG
TAAATGGCCCGCCTGGCTGACCGCCCAACGACCCCCGCCCATTGACGTCAATAATGACGTATGTTCCCATAGTAACGCCAATAGG
GACTTTCCATTGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCATATGCCAAGTACG
CCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTATGCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGT
ACATCTAGTATTAGTCATCGCTATTACCATGGTGATGCGGTTTTGGCAGTACATCAATGGGCGTGGATAGCGGTTTGACTCACGG
GGATTTCCAAGTCTCCACCCCATTGACGTCAATGGGAGTTTGTTTTGGCACCAAAATCAACGGGACTTTCCAAAATGTCGTAACA
ACTCCGCCCCATTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGAT
CGCCTGGAGACGCCATCCACGCTGTTTTGACCTCCATAGAAGACACCGGGACCGATCCAGCCTCCGCGGCCGGGAACGGTGCATT
GGAACGCGGATTCCCCGTGCCAAGAGTGACGTAAGTACCGCCTATAGACTCTATAGGCACACCCCTTTGGCTCTTggacctttaa
ttcaacccaacacaatatattatagttaaaTAAGaattattatcaaatcatttgtatattaattaaaatactatactgtaaatta
cattttatttacaatcAGACAGACTGTTCCTTTCCTGGGTCTTTTCTGCAGGCACCGTCGTCGACTTAACAGATCTGTTTAAGCT
T

Supplementary Figure 7: Dual-host promoter sequences for expression in insect and mammalian cells.
The DNA sequences of dual-host promoters CMVP10 and CMVintP10 are shown, which support heterologous expression of target proteins of choice both in insect cells and in mammalian cells. Promoter structures are depicted in a schematic representation. Promoter DNAs were inserted into MultiPrime Acceptor and Donors by using restriction sites AscI (red letters) and HindIII (green letters). CMV stands for cytomegalovirus promoter, P10 for baculoviral very late promoter p10, polyA for the SV40 polyadenylation site. CMVintP10 contains in addition an intron sequence.

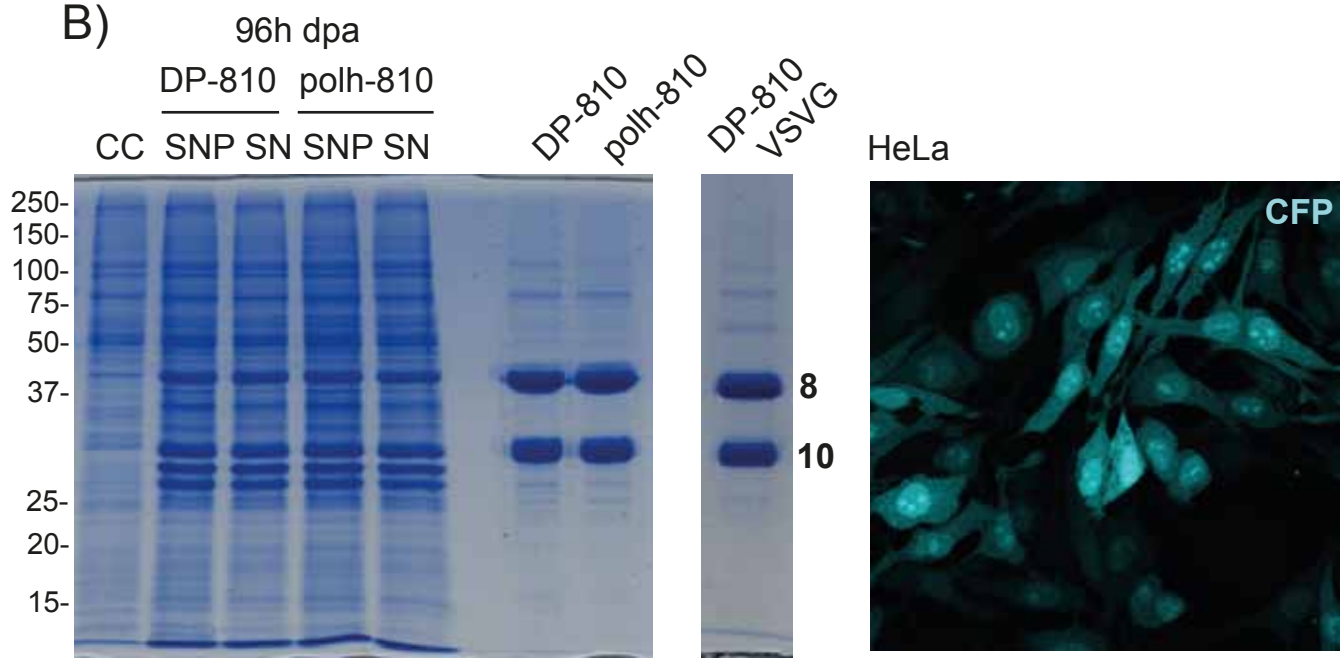
Supplementary Figure 8

Human transcription factor complex expression

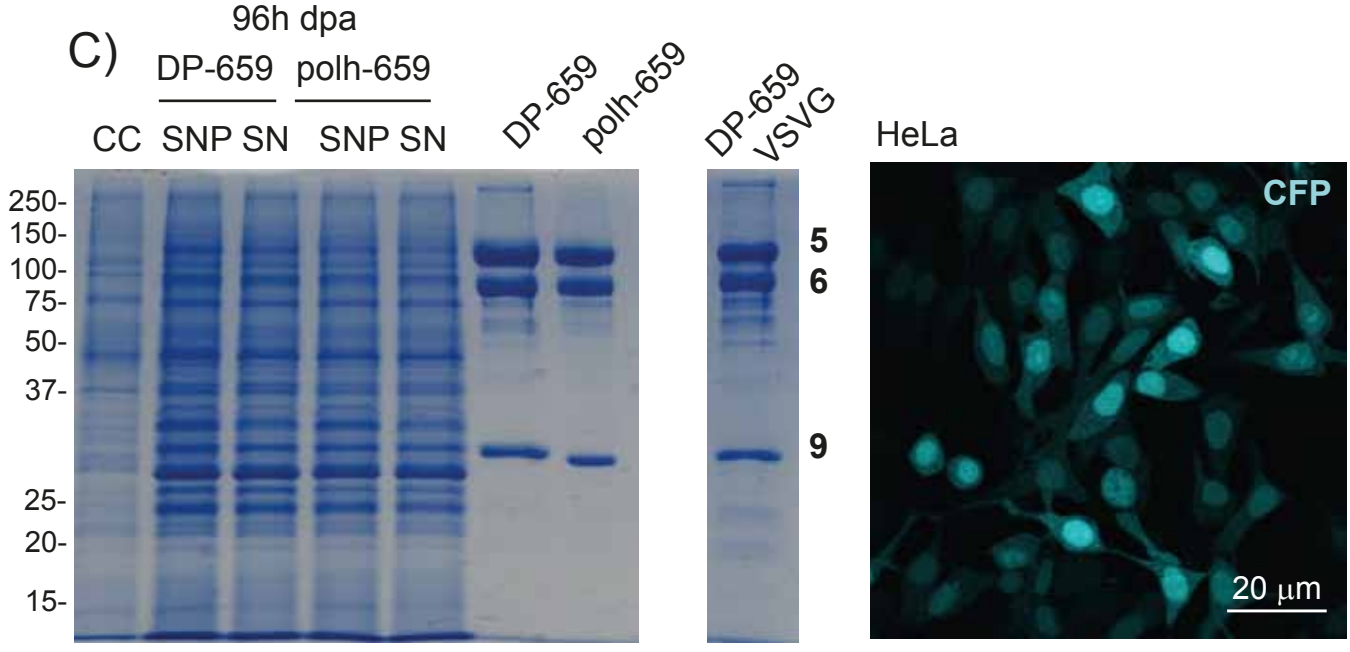
A)



B)



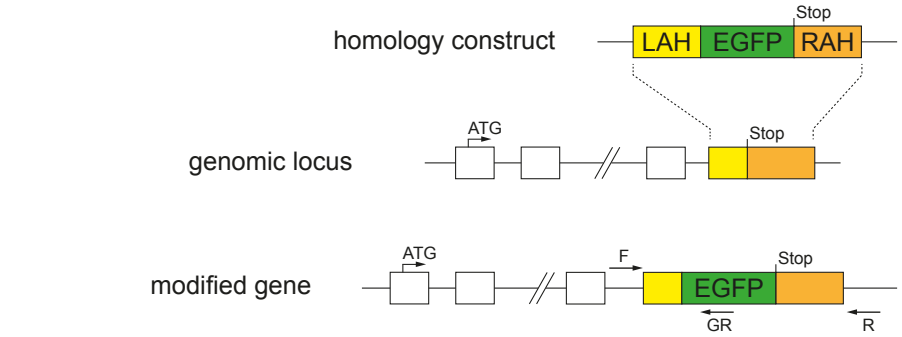
C)



Supplementary Figure 8: Validation of bifunctional dual promoter expression cassettes. Human General Transcription (GTF) factor TFIID subcomplex expression from expression cassettes driven by the bifunctional dual promoter CMVintP10-b was compared to expression from baculoviral very late promoter polyhedrin (polh) driven cassettes which are only active in insect cells. (**A**) Expression cassettes tested are shown in a schematic representation. CMV stands for cytomegalovirus late promoter, P10 for baculoviral very late promoter p10. TEV stands for tobacco etch virus NIa protease, tcs for TEV cleavage site, TAF stands for TBP associated factor, CFP for cyan fluorescent protein and poly A for the SV40 polyadenylation site. Small arrows indicate transcription start sites. An intron sequence enhancing transcription following the p10 promoter is represented as a box colored in gray. TFIID subcomplexes TAF8-TAF10 and TAF5-TAF6-TAF9 were expressed as self-processing polyproteins tailored into the individual subunits by TEV. (**B**) SDS-PAGE analysis of the expression of TAF8-TAF10 complex in baculovirus-infected insect cell cultures is shown (left). CC stands for uninfected cell control, SNP for whole cell extract, SN for cleared lysate. DP-810 denotes expression from the bifunctional dual promoter, polh-810 denotes expression from the baculoviral late promoter. EMBacY baculovirus was used. Virtually identical amounts of purified TAF8-TAF10 complex were obtained in the expression experiments. DP-810 VSVG denotes purified TAF8-TAF10 complex expressed from the bifunctional promoter, using MultiBacMam virus which displays VSV-G protein on its surface to enhance mammalian cell infectivity (middle). This virus was used to efficiently infect HeLa cells as evidenced by CFP fluorescence (right). (**C**) SDS-PAGE analysis of the expression of TAF5-TAF6-TAF9 complex in baculovirus-infected insect cell cultures is shown (left). CC stands for uninfected cell control, SNP for whole cell extract, SN for cleared lysate. DP-659 denotes expression from the bifunctional dual promoter, polh-659 denotes expression from the baculoviral late promoter. EMBacY baculovirus was used. Virtually identical amounts of purified TAF5-TAF6-TAF9 complex were obtained in the expression experiments. DP-659 VSVG denotes purified TAF5-TAF6-TAF9 complex expressed from the bifunctional promoter, using MultiBacMam virus which displays VSV-G protein on its surface to enhance mammalian cell infectivity (middle). This virus was used to infect HeLa cells with high efficiency as evidenced by CFP fluorescence (right).

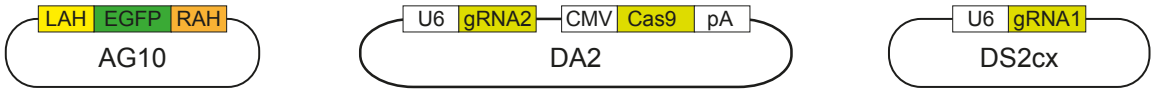
Supplementary Figure 9
Modification of the HMGA1 locus with CRISPR

modified from Ratz et al., Scientific Reports, 2015
(EGFP instead of rsEGFP2 was used)



HMGA1_out_F:	GTCCCTCCCTCTCCTGCTCCTAGAATACTCAG
HMGA1_out_F3:	TTGTCCTTCCTATGAGCCTCTGCAG
HMGA1_out_R:	CAATGACGGATGTCGAAGAATGGAACATTGAAC
EGFP_R:	CACGCTGCCGTCCTCGATGTTG
EGFP-N-Rev-long:	GTCAGCTTGCCGTAGGTGGCATC

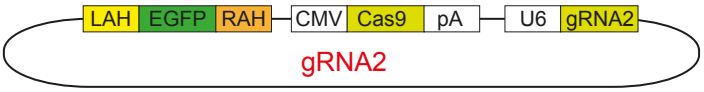
Plasmids for modification of the human HMGA1 locus
Plasmids were obtained from Stefan Jakobs (Ratz et al., 2015). The homology arm was cloned into the AscI/ PacI sites of AG10. rsEGFP2 was replaced by EGFP (left). U6-driven HMGA1-gRNA2 and Cas9 expression cassette from px330 were transfered to AscI/ PacI sites of plasmid DA2 (middle). U6-driven HMGA1-gRNA1 from px330 was transfered to AscI/ PacI sites of plasmid DS2cx (right).



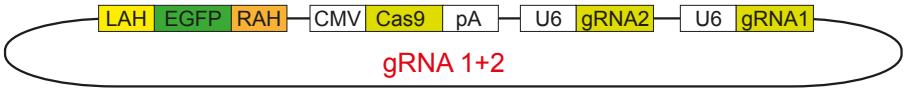
Baculovirus constructs for modification of the human HMGA1 locus

Homology construct („Homology“ in figures)

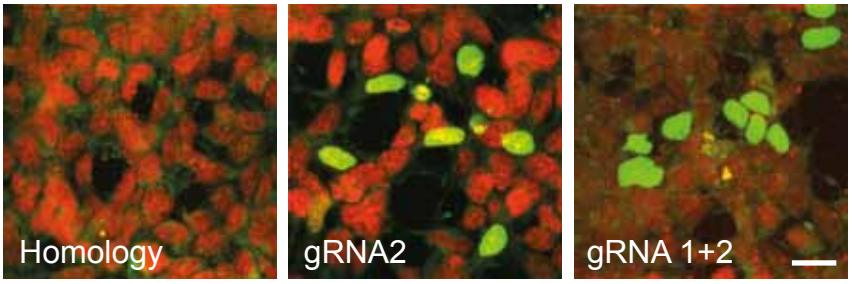
Homology construct / Cas9 - gRNA2 („gRNA2“ in figures)

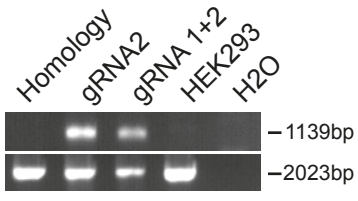


Homology construct / Cas9 - gRNA2 / gRNA1 („gRNA 1+2“ in figures)



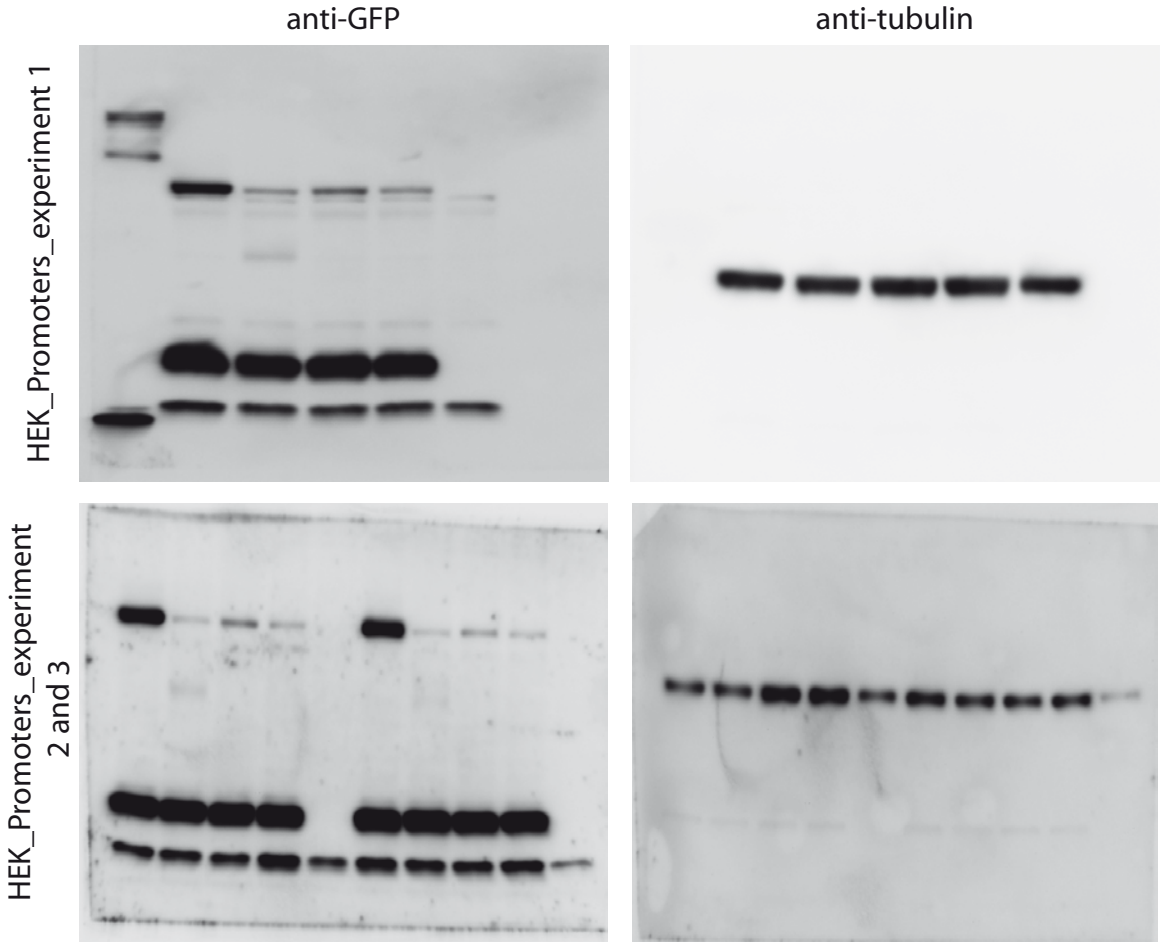
MultiPrime-based genome engineering of HEK293 cells. (left) HEK293 cells were infected with the indicated viruses to modify the HMGA1 locus. Cells were fixed after five days and the nuclei were counterstained with NucRed Dead 647 (Molecular Probes, red). Cells with a modified genome have a green nuclei. Scale bar: 20 μm. (right) PCR analysis of the infected cultures. The fragment with 1139bp for the mutant allel was only obtained from cultures that express the Homology construct together with gRNAs and Cas9.



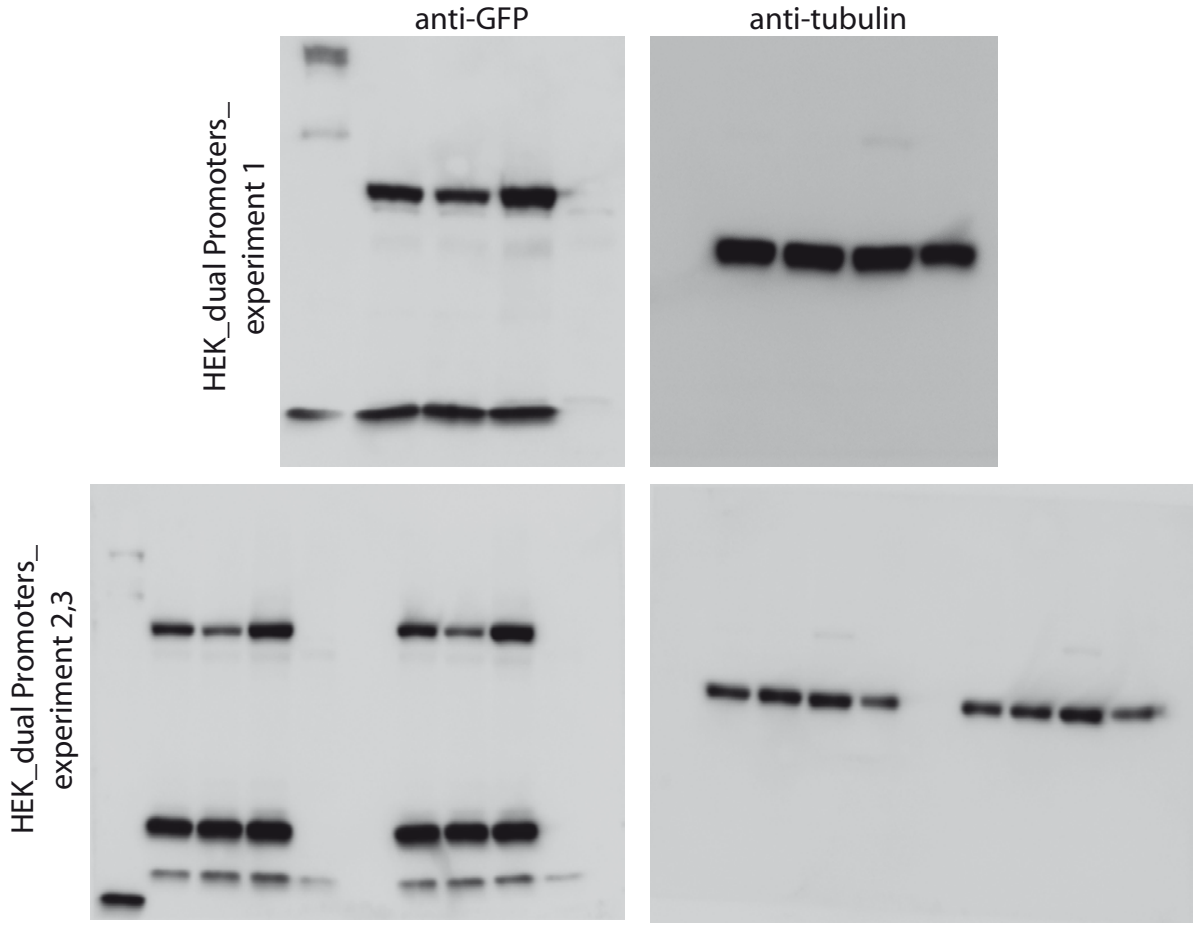


Supplementary Figure 10a

i) Original western blots of HEK293 cells (modulation of expression levels)

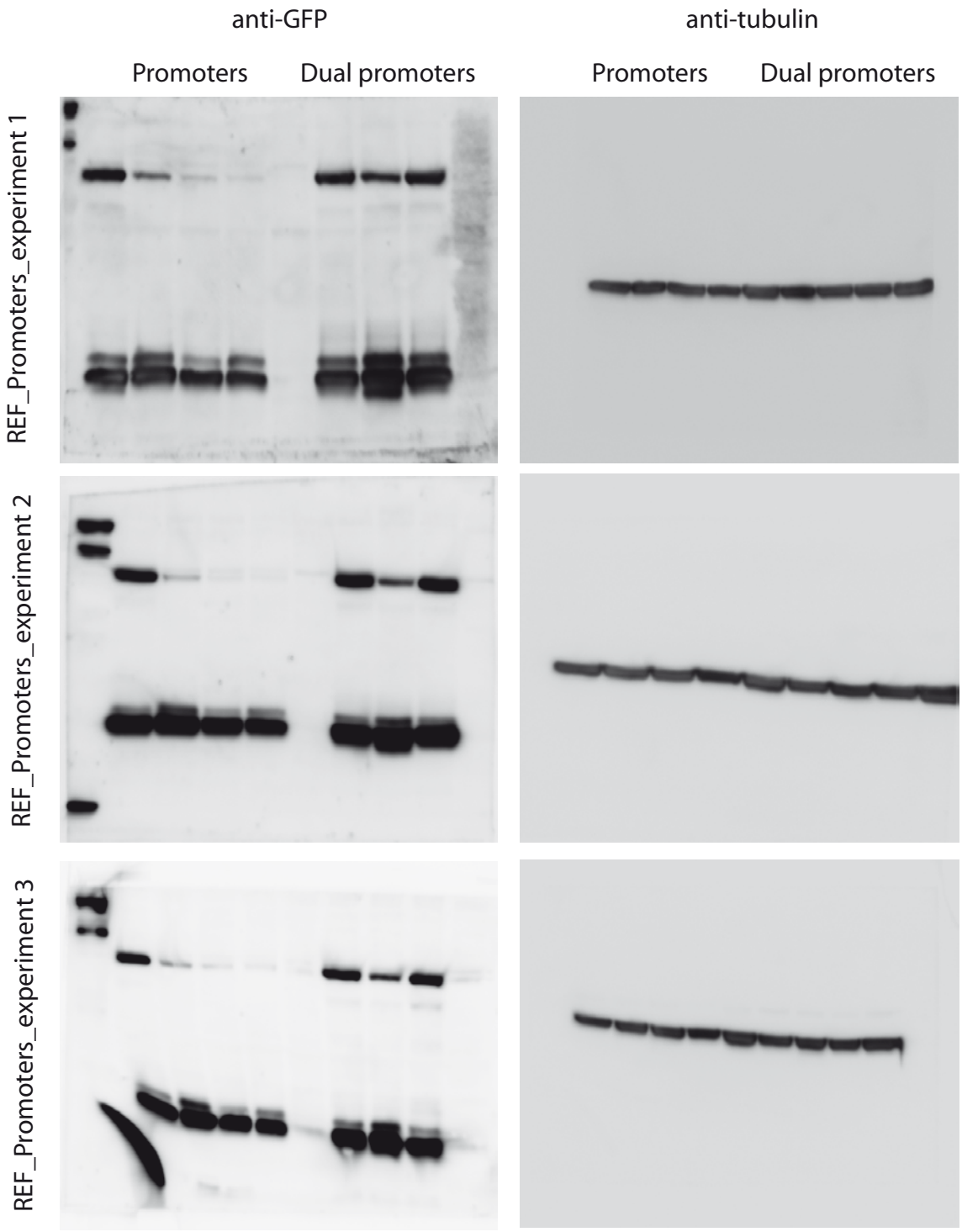


ii) Original western blots of HEK293 cells (dual-host promoters)



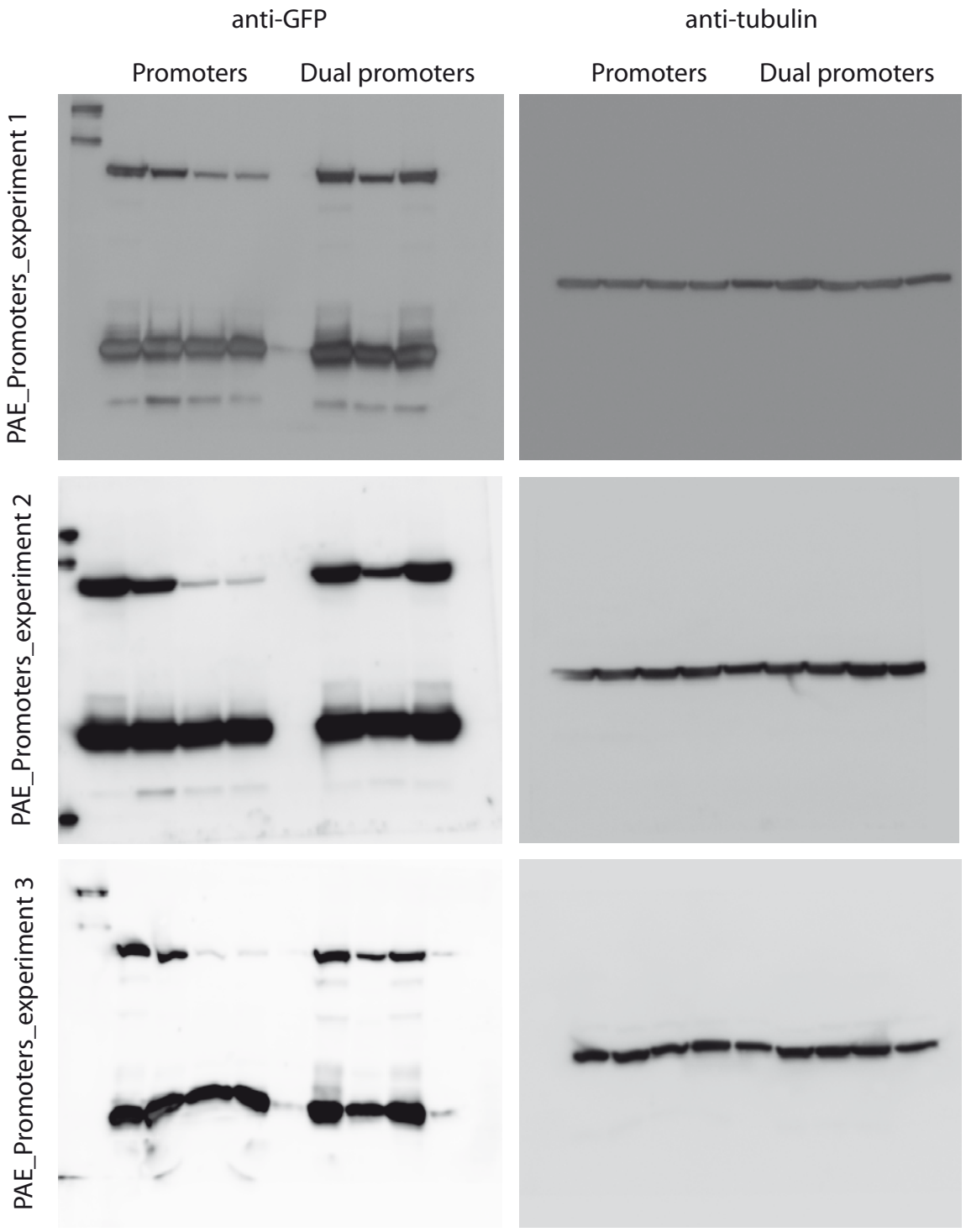
Supplementary Figure 10b

Original western blots of REF cells (modulation of expression levels and dual-host promoters)



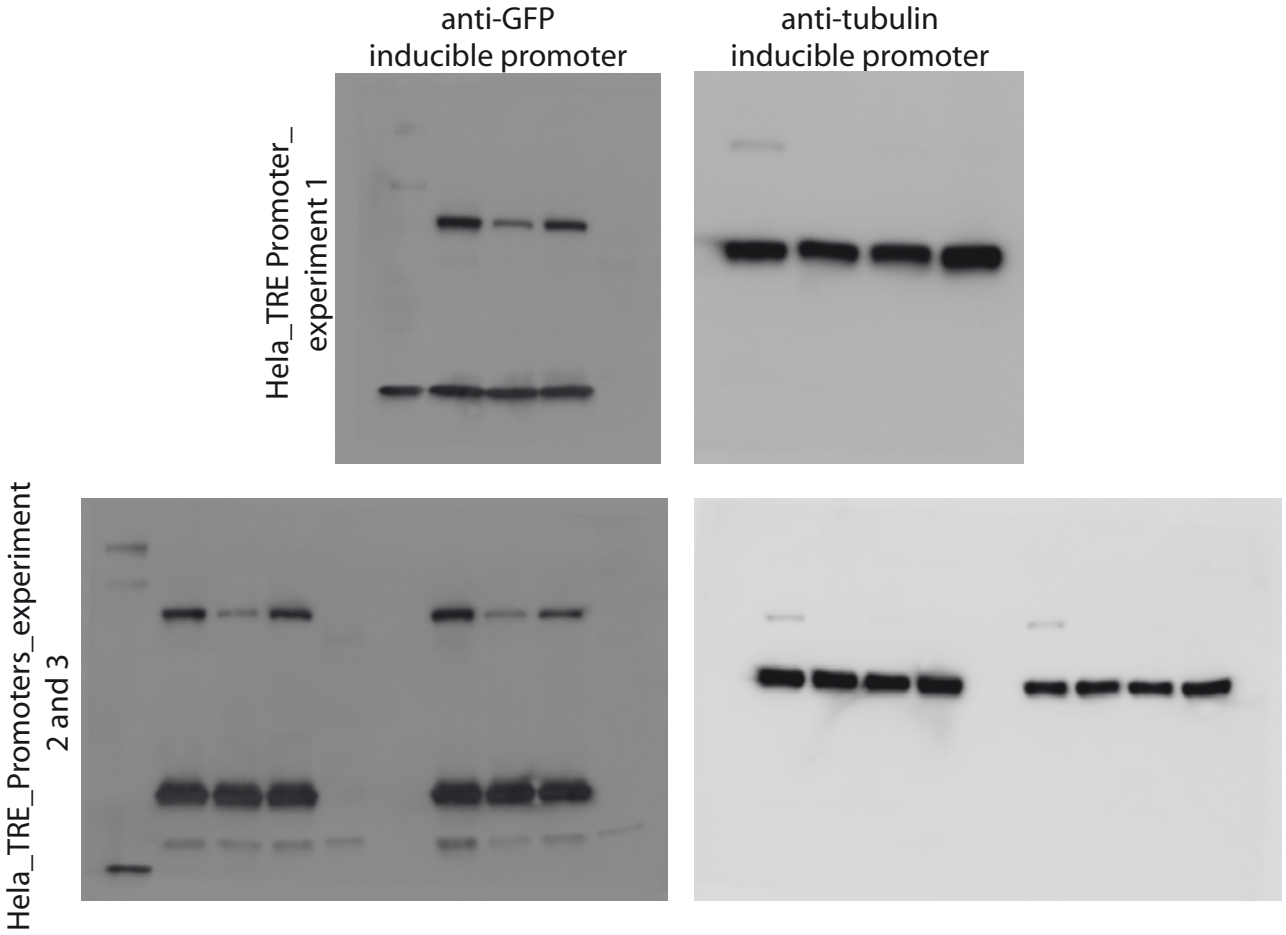
Supplementary Figure 10c

Original western blots of PAE cells (modulation of expression levels and dual-host promoters)

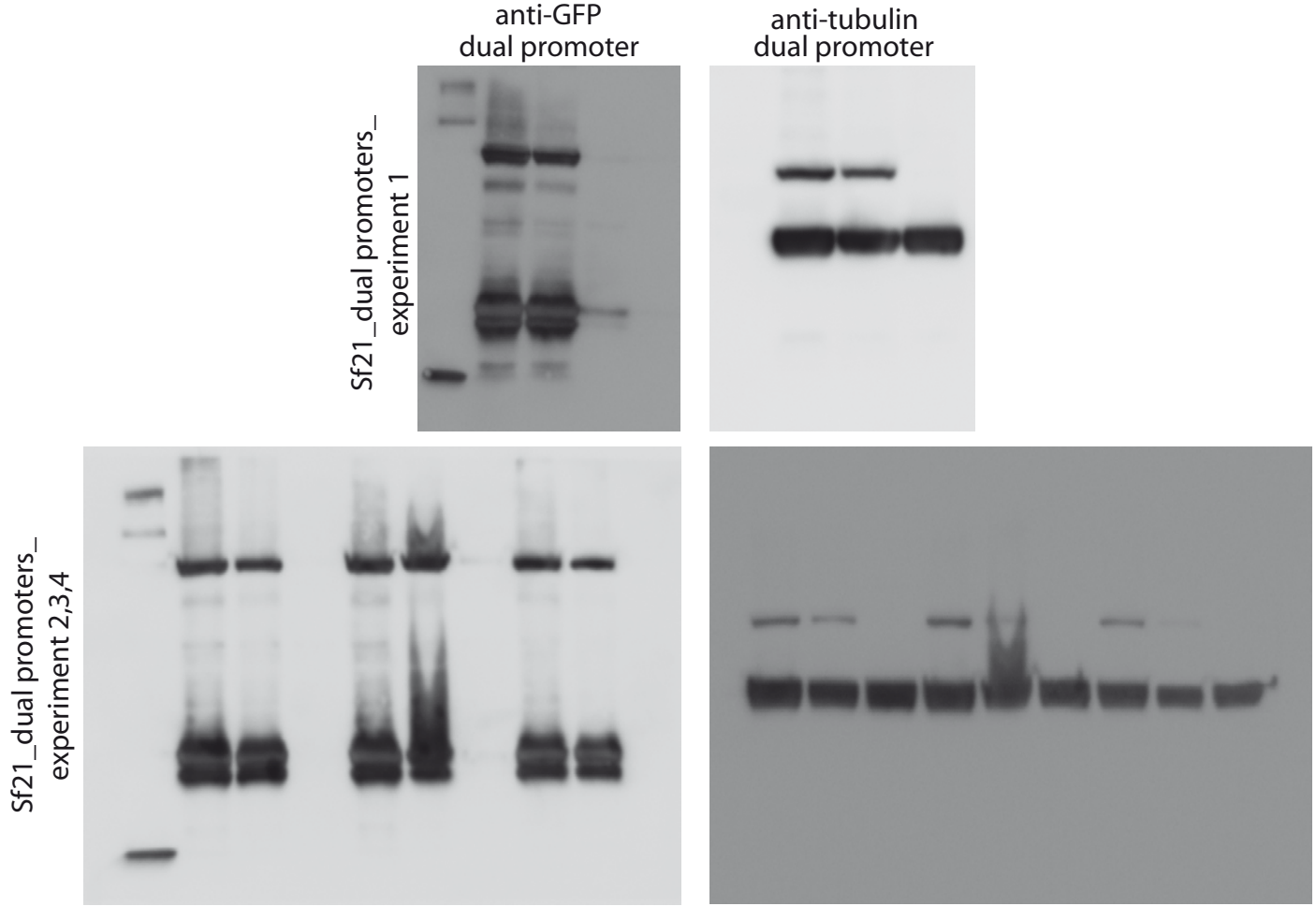


Supplementary Figure 10d

i) Original western blots of Hela tTA cells (modulation of expression levels)

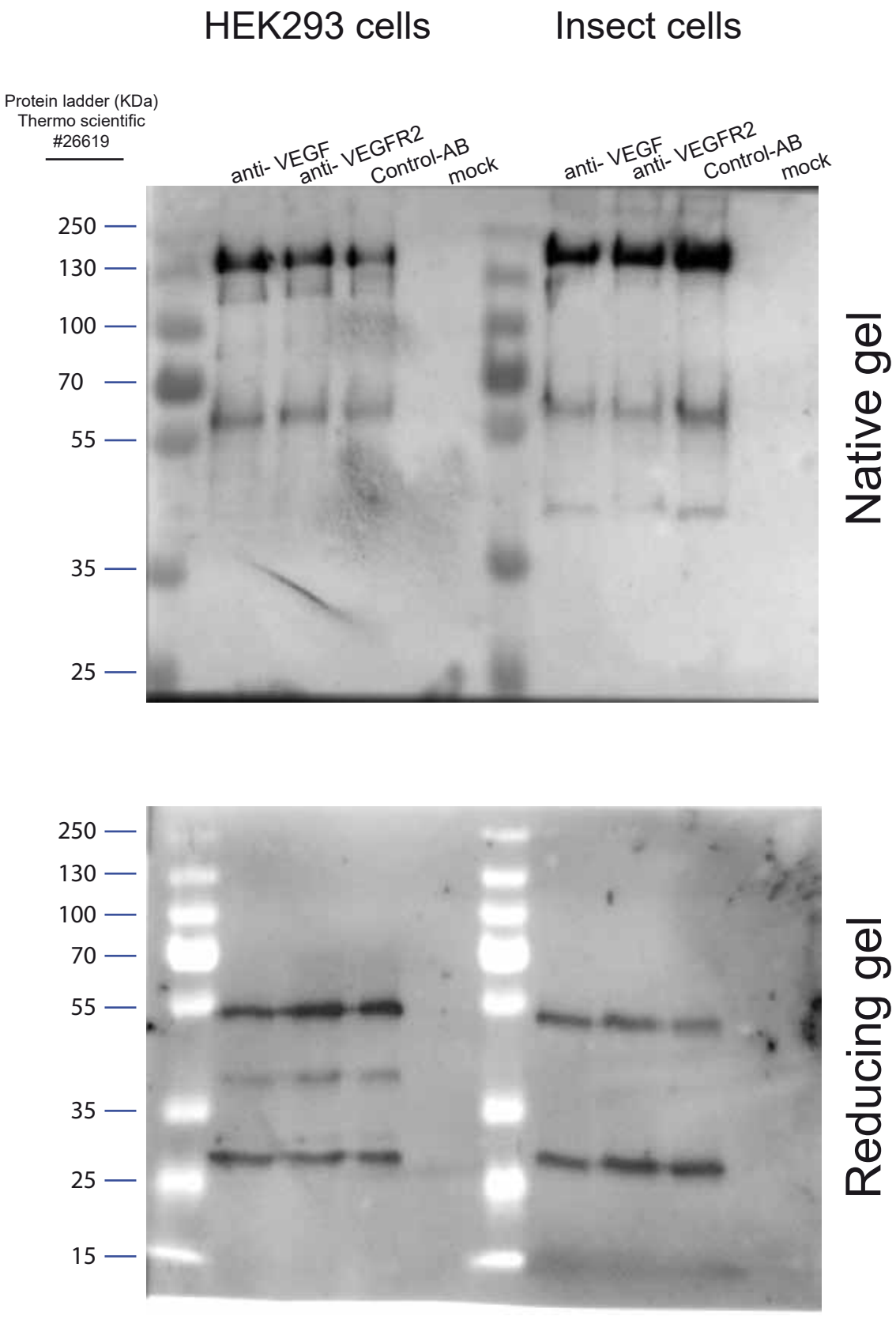


ii) Original western blots of Sf21(insect) cells (dual-host promoters)



Supplementary Figure 10e

Original western blots related to conversion of single chains to full length human antibodies



CHAPTER 3

SIMPLE PROTOCOL FOR EFFICIENT MULTIPRIME VIRUS GENERATION

Baculovirus is a promising vector for transducing numerous types of mammalian cells. Anyway production of functional baculovirus, harbouring all expected genes is challenging due to rearrangement of genetic materials in genome of virus during amplification.

In this chapter, I describe a step-by-step protocol for the cloning and assembly of genes of interest into MultiPrime system, and for generation of MutiPrime baculovirus. Furthermore, the protocol for efficient transduction of mammalian cells is included. I also provide some key steps which are critical for users from other labs to implement MultiPrime system. Finally, a troubleshooting table consisting of possible problems and their solutions is introduced.

Manuscript related to this chapter is planned for publication (Mansouri et al., in preparation).

Simple Protocol for Efficient MultiPrime Virus Generation

Maysam Mansouri[1], Aurélien Rizk[1], Kurt Ballmer-Hofer[1], and Philipp Berger[1]

[1] Paul Scherrer Institute, Biomolecular Research, Molecular Cell Biology, CH-5232 Villigen, Switzerland

[*] Corresponding author:
Dr. Philipp Berger
Paul Scherrer Institute
Molecular Cell Biology
CH-5232 Villigen
Phone: 0041 56 310 4728
FAX: 0041 56 310 5288
E-mail: Philipp.Berger@psi.ch

1 Abstract

Efficient, simple and reliable method for multigene delivery to mammalian cells is a crucial topic in many biological research laboratories. Baculovirus is a good choice for highly efficient delivery of multitude gene to mammalian cells due to its large cargo capacity. However, instability of genetic material during over-amplification of virus can restrict utility of this vehicle for gene delivery. Here, we introduce a step-by-step protocol for assembly and generation of intact baculovirus harboring multigene plasmid. Protocol for transduction of a variety of mammalian cells is also provided.

2 Introduction

Baculovirus is a promising vector for transducing numerous types of mammalian cells (Sung et al., 2014). Bacoluvirus, which normally infects insect cells, is a non-integrating virus that has a large cargo capacity, low toxicity and broad ability to transduce many mammalian cells (Kost et al., 2010). Although Baculovirus has unique advantages in comparison with other viruses, generation of intact baculovirus, harbouring all expected gene cassettes, can be quite challenging. In principle, the transgene in the genome of baculovirus is prone to spontaneous excision owing to instability and rearrangement of genetic material during amplification of virus within host cells (Chen et al., 2010; Fitzgerald et al., 2006; Sung et al., 2014). Mutation and rearrangement of genome is a general strategy in viruses and allow them to obtain higher variation in the next generation. Therefore, uncontrolled baculovirus generation can lead to production of defective virus resulting in incomplete gene delivery. This drawback can be even worse when baculovirus is carrying multigene constructs. In fother words, multigene baculoviral vectors are composed of numerous repetitive elements (promoters, terminators, recombination sequences) which may be a preferred site for eliminating recombinant cassettes during virus generation in insect cells (Fitzgerald et al., 2006). Therefore, development of a highly optimized protocol for production of healthy baculovirus, bearing all the transgene constructs is highly desirable. We have recently developed an efficient baculovirus-based multigene delivery system, called MultiPrime, for mammalian cells (Mansouri et al., 2016). MultiPrime is established to tackle the difficultly of multigene delivery to mammalian cells particularly primary and stem cells, which are recalcitrant to plasmid transfection. Here, we introduced a simple protocol for production of multigene baculoviral vectors. We optimized our protocol for minimum gene deletion with maximum virus titration. This feature allows baculovirus supernatant to be directly added to the mammalian cell and eliminate the requirement to concentrate virus by ultracentrifugation or special kits for virus purification. Furthermore, this protocol is faster than other introduced protocols. The entire

MultiPrime procedure, from assembly of constructs to baculovirus generation and transduction, can be completed in one month. This protocol can be used for generation of virus for multiple protein expression purposes in structural biology as well.

3 Reagents

E. coli competent cells for normal cloning: TOP10, DH10β, Mach1, BW23473 (pir+)
E. coli competent cells for MultiPrime baculovirus generation: DH10Bac EMBacY and DH10MultiBacMam competent cells
Spodoptera frugiperda (Sf21) cells (Life Technologies)
Mammalian cells to be transduced, for example;
Established cell lines: HEK293, Hela, COS-7, PAE, Swiss 3T3, etc. Primary cells: HUVEC (Life Technologies), REF (isolated from rat embryos), Rat cortex Neuron (Life Technologies). Stem cell: induced- pluripotent stem (iPS) cells (we have used the human iPS cells, supplied by the Taylor's lab (University of Basel).
SF-900 II serum-free (SFM) medium or SF-900 II 1%FBS and +1% Penicillin/Streptomycin medium
DMEM (high glucose 4.5g/L with L-Glutamine, Amimed)
Trypsin-EDTA PBS 1:250 (0.05%/0.02%) without Ca++/Mg++, with Phenol red (Amimed)
Gentamycin, Ampicillin, Chloramphenicol, Kanamycin, Tetracycline
X-gal, IPTG
Cell fectin II reagent (Invitrogen)
Penicillin-streptomycin, 100× (10,000 U/ml, Amimed)
MultiPrime vectors:
Acceptor vectors with CMV promoter: pSI-AGZ10 (Accession code at NCBI: KX001771), pSI-AGR10 (KX001770), pSI-AGL10 (KX001778).
Donor vectors with CMV promoter; can be obtained from reference (Kriz et al., 2010).
Acceptor vectors with dual promoters; pSI-AGL10-DM (KX001779), pSI-AGL10-DI (KX001780).
Donor vectors with dual promoters; pSI-DSZ2cx-DI (KX001773), pSI-DSZ2cx-DM (KX001774), pSI-DAZ-DI (KX001782), pSI-DAZ-DM (KX001781).
Acceptor and donor vector with different/inducible promoters; pSI-DSZ2cx-TRE (KX001772, with a tetracycline- inducible promoter), pSI-DSZ2cx-PGK (KX001775, with a mouse phosphoglycerate kinase 1 or PGK promoter), pSI-DSZ2cx-UBC (KX001776, with a human Ubiquitin C or UBC promoter) and pSI-DSZ2cx-SV40 (KX001777, with a simian virus 40 early promoter or SV40).
Cre recombinase and 10× reaction buffer (NEB)
Phusion high-fidelity DNA polymerase in 2× HF or 2× GC buffer (NEB)
PCR primers

Paraffin wax beads (e.g. Fluka Chemika, #411671)
Restriction enzymes DpnI, AscI (SgsI), SapI (LguI), and PacI (NEB, Fermentas)
QIAprep Gel Extraction kit, QIAprep PCR purification kit (Qiagen)
Midiprep plasmid extraction kit (Genomid, Jetstar 2.0 plasmid Midiprep kit)
Miniprep plasmid extraction kit (Thermo scientific)
QIAprep PCR purification kit (QIAgene)
T4 DNA ligase, 10× T4 DNA ligase buffer (NEB)
Fetal calf serum (Amimed)
FuGENE HD (Promega)
Opti-MEM I Reduced Serum Medium (Gibco)

4 Equipment

Ultra-centrifuge (Optima^{TM} XL-100K, BECKMAN COULTER)
Confocal microscope (Leica SP5)
Fluorescent microscope (Olympus IX81 equipped with an Andor iXonEM camera)
Inverted microscope equipped with a phase-contrast ring and fluorescence filters
Thermocycler
Equipment for agarose gel electrophoresis
Electroporator and cuvettes
Tissue culture Laboratory
Incubator, 37 °C, 5% CO2
Incubator, 27 °C, no CO2
Refrigerator, 4 °C
Freezer, −20 °C
Tissue culture plates (10 cm and 6-, 12- and 96-well)
Tissue culture flasks (T-25, T-75 and T-150)
Micro centrifuge
Neuobar counting system
Syringes (BD Biosciences)
Syringe filter 0.22µm (Merck Millipore Ltd.)
Filter 0.22 µm and 0.45 µm (SARSTEDT, Germany)

5 Procedures

A: Cloning genes of interest into MultiPrime vectors and mutigene assembly

1. Make your plan for cloning of interested genes into MultiPrime vectors. First gene must be clone into an acceptor and the rest into donor vectors. Different methods of cloning (A, B, C, D) can be used to achieve this goal. Use option A, if a SapI (LguI) recognition site dose not present inside gene of interest sequence. Option B, is suitable when an expression cassette including promoter, gene of interest and a terminator signal already exists. Use option C, for a Gibson assembly based cloning in a free restriction enzyme (RE) reaction. Option D is a Sticky end based PCR cloning and is suitable when one (or more than one) internal recognition site for RE is present.
   CRITICAL STEP *Plasmid prepared with option C can directly jump to step 7. Meaning that it does not need to further treatments.*
2. Amplify DNA fragments via PCR reaction. We amplify DNA fragments using this PCR reaction and program:

Template DNA (100 ng/µl)	1 µl
5′ primer (5 µM stock)	3 µl
3′ primer (5 µM stock)	3 µl
ddH2O	8 µl
One paraffin wax bead	20 µl
2× Phusion HF Reaction mix	15 µl

1x	98 ^{oc}	5 min
25-30x	98 ^{oc}	30s (denaturation)
	52 ^{oc}	30s (annealing)
	72 ^{oc}	1min/kb (extension)
1x	72 ^{oc}	10 min (final extension)

   CRITICAL STEP *It is crucial to use a high fidelity proofreading polymerase with a low error rate to amplify your DNA fragments. We used Phusion high-fidelity DNA polymerase in 2x HF buffer.*

   Troubleshooting

   CRITICAL STEP *Treatment of PCR product with DpnI (directly add 1 ul DpnI to 30 ul PCR product, overnight at 37^{oc}) leading to elimination of methylated parental DNA. Therefore, it can decrease rate of background after transformation.*
3. Digest PCR product with suitable restriction enzymes at 37^{oc} for Overnight (O/N).
4. Clean up your PCR products by QIAprep PCR purification kit and run 2 ul of final products on 1.2% agarose gel to evaluate their qualities.
5. Prepare your acceptor and/or donor MultiPrime vectors according to your ligation strategies. For option A and B, MultiPrime vectors must be a cleaned up product of a SapI or AscI /PacI digestion, respectively. For option D, it should be prepared according to restriction enzymes that has been used for amplification of PCR fragments.

6. Ligate MultiPrime vector backbone with prepared PCR products through T4 DNA ligase for O/N in room temperature (RT) or Fast T4 Ligase for 20 min at RT.
7. Transform 1 ul of ligation or Gibson assembly product to 100 ul of E. coli competent cells (TOP10, DH10β, Mach1 or BW23474). A mixture of DNA and competent cells should be keeping on ice for 20 min and then use it for electroporation (1.7 kV pulse, 25 μF, 200Ω). Transformed cells are recovered in 400 ul of SOC medium and 2h at 37^{oc}. Plate 250 ul of them into LB agar containing appropriate antibiotic (List and concentration of antibiotics are available in table 2a). Incubate the plates at 37^{oc} for O/N.
   CRITICAL STEP *Donor plasmids are containing a conditional origin of replication derived from R6Kγ and then have to be propagating in cell strains expressing the pir gene (pir+ strains) like BW23474. But acceptor plasmids and recombined plasmids can be propagated in common bacterial strains such as TOP10, Mach1, DH10 β.*

   Troubleshooting
8. Pick 2-10 colonies and inoculate to 5 ml LB broth containing appropriate antibiotic. Incubate the tubes in 37^{oc} with an orbital shaker (150 rpm) for O/N.
9. Isolate recombinant plasmids via miniprep plasmid extraction kit from 4.5 ml of culture and keep the rest (500 ul) in 4^{oc}.
10. Validate correct clones through restriction enzyme digestions and/or PCR. It is also recommended to verify the cloned gene by sequencing, as well.
11. Inoculate 50 ul of positive clones into 100 ml LB broth culture comprise correct antibiotic. Incubate them at 37^{oc} for O/N.
12. Apply a midiperep kit to isolate recombinant plasmid. Measure DNA concentration through a nanodrop.
    CRITICAL STEP *Usually concentration of DNA in this step should be around 1-2 ug/ul with a ratio of absorbance (A260/280) 1.8-2.*
13. Combine MultiPrime acceptor and donor plasmids through in vitro cre/loxP recombination reaction:

Acceptor (1ug)	1ul
Donor (1 ug)	1ul
10x Buffer	1ul
Cre recombinase	1ul
ddH2O	6ul
Total	**10 ul**

    Incubate reaction for 30 min at 37^{oc} and then 5 min on ice.
    CRITICAL STEP *One acceptor plasmid can be assembled with one donor plasmid in an in vitro cre/loxP recombination reaction. The second, third and fourth donor plasmids should be inserted in a sequential manner with double-, triple-, four-assembled MultiPrime plasmids, respectively.*
14. Repeat step 7. Add appropriate combination of antibiotics to the LB agar plates (Tabel 2a).
15. Pick 5 colonies and inoculate to small-scale cultures of LB broth media containing all combined antibiotics in 32^{oc} and 110 rpm orbital shaker.

CRITICAL STEP *Grow bacterial in 32^{oc} (instead of 37^{oc}) to avoid unwanted recombination that can happen over fast bacterial amplification. This type of recombination can lead to deletion of expression cassette(s) completely or partially.*

16. Repeat step 9.
17. Validate correct clones through restriction enzyme digestions (restriction mapping).
    CRITICAL STEP *Although assembly of acceptor and donor (and a combined acceptor-donor with the next donor vector) plasmids can be done with any DNA sequence analyzer software but we recommend Ape plasmid editor software. This software is an open source software and can be acquired from http://biologylabs.utah.edu/jorgensen/wayned/ape/. The most convenient way to assemble two vectors is to open their sequence with Ape. Then select Multi-cre recombination from Tools menu. Next, determine acceptor and donor vector sequence files and press OK. Finally, save your fused acceptor-donor sequence in your desired place.*
18. Repeat 11 and 12.
19. **(Optional)** Functional analysis of recombinant MultiPrime plasmid can be done in this step. This step can validate expression of all interested genes in mammalian cells via tranfection. Many procedures for transfection are available. In the simplest one, plate 60 x 10^4 HEK293 cells/6-well plate about 6-8 h before transfection. If genes of interest are fused to fluorescent proteins, poly-L-lysine treated coverslips can be put in the 6-well plates, as well. Mix 2 ug of assembled MultiPrime plasmid with 6 ul of Fugene HD in 100 ul Optimum medium and keep the mixture for 15 min at RT. Add the cocktail of DNA and transfection reagent to the 6 well plate as drop wise and analyze cells through WB, RT PCR or fluorescent microscopy after 24-48h.

B: Generation of MultiPrime baculovirus

20. Transform 100 ngr/ul of recombinant multigene plasmid into 100 ul of E. coli DH10 BacEMBacY and/or E. coli DH10 MultiBacMam competent cells. We apply 2 kV pulse (instead of 1.7 which were mentioned in step 6), 25 µF, and 200 Ω for transposition of MultiPrime plasmid into E. coli harboring bacmids. Add 400 ul of SOC medium and incubate them for O/N at 37^{oc}.
    CRITICAL STEP *E. coli DH10 MultiBacMam is a pseudotyped baculovirus with VSVg (Vesicular Stomatitis Virus glycoprotein). VSVg has been shown to improve mammalian cell transduction efficiency. E. coli DH10 EMBacY does not express this glycoprotein on its surface. Also, EMBacY and MultiBacMam baculoviruses express EYFP and mCherry, respectively. These fluorescent markers simplify and standardize monitoring the generation of baculovirus during infection in insect cells (Fig.3)*

21. Plate 250 and 50 ul of transformed bacteria on LB agar contain all appropriate antibiotics, X-gal and IPTG. Incubate the plates in 32^{oc} for O/N.
    CRITICAL STEP *LB agar should containing Gentamycine, Ampicilin, Kanamycin, tetracycline, chloramphenicol, IPTG and X-Gal (Table 2b). Prepare at least three plates for each MultiPrime constructs. After preparation, plate should be keeping at 4^{oc} and away from light. Before spreading of bacteria, pre-warm the plates at 37^{oc} for 30 min.*
22. Pick 5 white colonies as well as one blue colony (as control negative) and re-streak them down on new LB agar (with all the antibiotics as well as IPTG and X-gal).
23. Pick at least two white re-streaked colonies and inoculate them individually into 100 ml LB broth consist of all antibiotics but not X-gal and IPTG. Incubate them at 32^{oc} and 110 rpm orbital shaker for O/N.
    CRITICAL STEP *Make 2-3 bacterial stocks from each white colony through mixing with 100% glycerol (1:1) and keep them at -80^{oc}.*
24. Apply a midiperep kit to isolate recombinant bacmid plasmid. Use a nanodrop to measure concentration of extracted Plasmid. Usually concentration of DNA in this step is about 0.5-1 ug/ul.
    CRITICAL STEP *We recommend to dry DNA pellet (in the last step of plasmid isolation) under a bacterial hood. Also, use a sterile ddH2O or TE for re-suspension of DNA pellet).*
    CRITICAL STEP *Use at least two different bacterial white colonies to generate MultiPrime baculovirus.*
25. For each MultiPrime bacmid DNA, take two tubes; A and B. Add 100 ul of SF-900 II SFM medium plus 8 ul of Cell fectin II to the tube A. Add 100 ul of SF-900 II SFM medium plus 2 ugr recombinant bacmid in the tube B. Keep them for 5 minutes away and then mix them together gently. Keep the mixed tubes (A+B) for 30-40 min at R/T.
    CRITICAL STEP *From step 25, the entire steps must be performed in a sterile cell culture hood.*

Troubleshooting

26. For every MultiPrime bacmid DNA, seed 2 ml of 0.5 x 10^6 insect Sf21cells/ml in a 6-well plate for V0 virus generation. Seal the plate with parafilm and incubate it in a 27^{oc} incubator for 30-40 min.
    CRITICAL STEP *Use newly thawed insect cells (Sf21 or Sf9). Cells should proliferate properly every 18h. Their shape and numbers are very important. Keep the stock of cells between 1-1.5 x 10^6 cells/ml.*
    CRITICAL STEP *Keep the cells in SF-900 II SFM medium during transfection.*
27. Transfect Sf21 cells with a mixture of DNA from step 25 in 6-well plate as drop wise. Seal the 6-well plate with parafilm and incubate it for 5 h at 27^{oc}.
28. After 5 hours, discard the supernatant (contain transfection reagent) and replace it with 2 ml either SF-900 II SFM or fresh SF-900 II 1%FBS-1%PS

medium. Return the 6-well plate to 27^{oc} incubator and continue it with more 47 hours incubation (52 hours after beginning of transfection).

Troubleshooting

29. Collect supernatant and centrifuge it in 1200 rpm for 10 min. Pass the supernatant through a 0.45 uM and collect them in 2-ml eppendorf tubes. This is the initial virus V0.

    CRITICAL STEP *Baculovirus generated in steps V0, V1 and V2 are stable in 4 ^{oc} and dark place for 4-6 months. Also, baculovirus stocked with 5% FBS and 2.5% Glycerol can be keep in -80^{oc} for long time (months).*

30. Seed 4 ml of 0.5 x 10^6 insect Sf21cells/ml in a 25 cm T-flask. Add 600 ul of V0 stock into T-flask and put the flask in 27^{oc} incubator with 80 rpm orbital shaking. Keep the flask to culture for 72 h. This is generation 1 virus (V1). Monitor MultiPrime baculovirus generation inside the insect cells through fluorescent protein signals with a fluorescent microscope. In this step, at least 50 % of cells are fluorescent.

    Troubleshooting

    CRITICAL STEP *If 20-30 of cells is shiny, add 1.5 ml SF-900 II SFM medium to the T-flask and let them growth for more 18-22h. Cells must be harvest in this step even if they are not 60 % infected (Fig. 3).*

31. Harvest V1 MultiPrime virus with centrifugation of whole medium suspension in 1200 rpm, 10 min and 4^{oc}. Supernatant should be filtrate with a 0.45 ul filter and keep in 4 ^{oc}.

32. Seed 100 ml of 0.5 x 10^6 insect sf21cells/ml in a 75 cm T-flask. Add 4 ml of V1 stock into the T-flask and put the flask in 27^{oc} incubator with 80 rpm orbital shaking. Keep the flask to culture for 48-72.

    CRITICAL STEP *Monitor baculovirus generation by fluorescent signals. After 48 h should almost 50% of insect cells get infected and be fluorescent. After 72 h more that 90% of cells are infected. Insect cells should not keep longer than 72h in this step.*

    CRITICAL STEP *Harvest V2 MultiPrime virus with centrifugation of entire medium suspension in 1200 rpm, 10 min and 4^{oc}. Supernatant should be filtrate with a 0.45 uM or 0.22 uM filter and keep in 4^{oc}.*

    CRITICAL STEP *With this protocol usually we produce 1 x 10^8 to 3 x 10^8 pfu/ml. This amount of bacuvirus is usually enough to transduce mammalian cells. Therefore, this supernatant can be added directly to the mammalian cell plate. But anyway, sometimes a higher transduction rate is more desirable. We recommend concentrating of generated baculovirus through ultra-centrifugation (see box 1).*

    CRITICAL STEP *Determine titer of generated baculovirus concentration using either plaque assay or end-point dilution assay. It would be help you to fine-tune transduction level according to your interest. We recommend end-point dilution assay.*

C: Transduction of mammalian cells

In this part, we explain a general protocol that can be used for a various mammalian cells including established cell lines, primary cells and stem cells.

33. Seed mammalian cells in the desired density in the 24-, 12- or 6- well plate one O/N prior transduction.
    CRITICAL STEP *Density of mammalian cells depending on the aim of experiment, size of cell and their doubling time. HEK293 or Hela cells for microscopy purposes usually can be seed 2.5 x10^5- 3.6 x10^5 cells/ml in 6 well plate. We use a density 7.2 x10^5 and 8.4 x10^5 cells/ml for HEK293 and HeLa cells when a protein expression experiment is aimed. COS-7 and PAE cells need to be 50 x 10^4 cells/6-well confluent and HUVEC or REF need to be 70 x 10^4 cells/6-well. We used 2 x10^5 induced-Pluripotent Stem cell (iPS).*
    CRITICAL STEP *If cells are plated on the coverslips, make sure that they are stably mounted to the glass. Otherwise treat coverslips before cell seeding for 30 min at RT with suitable reagent (like Poly-L-Lysine for HEK cells, Poly-D-Lysine for neurons, Matrigel for iPS and etc).*
    CRITICAL STEP *Seed mammalian cells in their usual medium along with their additives (e.g., FBS, growth factors, antibiotics and etc).*
34. Check the cells with normal phase contrast microscope before starting transduction. Cells should look spread and completely steak to the plate.
35. Next day after seeding, aspirate medium and add desired volume of MultiPrime baculovirus to mammalian cell culture according to planed MOI for transduction. We suggest using either 2 ml un-concentrated plus 500 ul normal mammalian medium (without FBS and PS) or 250 ul concentrated virus along with 1 ml normal mammalian cell medium (without FBS and PS). Keep the cells under transduction at 37^{oc} incubator for 8h.
    CRITICAL STEP *Pre-warm baculovirus stock with keeping them at RT and dark place half an hour before transduction.*
    CRITICAL STEP *If viruses are going to be thawed from -80, it is better to thaw them on the ice at 4^{oc} for O/N.*
    Troubleshooting
36. Discard supernatant when incubation time is over. Replace transduction medium with fresh and complete medium containing all the necessary antibiotics or FBS (e.g. DEME + 10% FBS + 1% PS).
    CRITICAL STEP *We recommend keep the antibiotics in culture as could as possible.*
37. Proceed with your experiment; fix the cells after 48 h and analyze the expression level through immunostaining or western bloting. Also, lysed cells can be used for mRNA expression level analysis and etc.
    Troubleshooting

6 conclusion

We described a rapid, simple and efficient for production of baculovirus harboring multitude gene cassettes. Using this protocol, generation of bacilovirus take place relatively fast (10 days for V0 to V2). Furthermore, V2 virus can transduce mammalian cells directly without need to an extra ultra-centrifugation step when a super high rate of transduction is not aimed. In contrast with other protocols that mainly transduce mammalian cells by baculovirus in 25^{oc} for a higher efficiency (Sung et al., 2014), we transduced all the mammalian cells in 37^{oc} and obtained even higher that 95% transduction efficiency. Finally, this protocol is not only restricted for generation of baculovirus, that harboring multigene for mammalian cells, but also can be used for production of baculoviurs with either single or multigene for protein production in insect cells.

7 troubleshooting

Troubleshooting advices can be found in Tabel1.

8 References

Chen, X., Pham, E., and Truong, K. (2010). TEV protease-facilitated stoichiometric delivery of multiple genes using a single expression vector. Protein Sci. *19*, 2379–2388.

Fitzgerald, D.J., Berger, P., Schaffitzel, C., Yamada, K., Richmond, T.J., and Berger, I. (2006). Protein complex expression by using multigene baculoviral vectors. Nat. Methods *3*, 1021–1032.

Kost, T.A., Condreay, J.P., and Ames, R.S. (2010). Baculovirus gene delivery: a flexible assay development tool. Curr. Gene Ther. *10*, 168–173.

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. Nat. Commun. *1*, 120.

Mansouri, M., Bellon-Echeverria, I., Rizk, A., Ehsaei, Z., Cianciolo Cosentino, C., Silva, C.S., Xie, Y., Boyce, F.M., Davis, M.W., Neuhauss, S.C.F., et al. (2016). Highly efficient baculovirus-mediated multigene delivery in primary cells. Nat. Commun. *7*, 11529.

Sung, L.-Y., Chen, C.-L., Lin, S.-Y., Li, K.-C., Yeh, C.-L., Chen, G.-Y., Lin, C.-Y., and Hu, Y.-C. (2014). Efficient gene delivery into cell lines and stem cells using baculovirus. Nat. Protoc. *9*, 1882–1899.

Tabel1: Troubleshooting

	Problem	Reason	Solution
2	Nonspecific band	Mispriming during annealing step	Increase annealing temperature, use wax beads
2	No PCR product	One (or more than one) of the PCR component(s) are missed	Make sure that all the components are present in the PCR reaction
		Concentration of template is too low or too high	Adjust template concentration between 50-200 ng/ul
		PCR Product has GC rich templates or it is long and has complex secondary structures	Use Phusion GC Buffer
7	No colony growth on the plate	Transformed bacteria plated on inappropriate antibiotic plate	Repeat transformation and make sure that the bacteria are plating on appropriate antibiotic plates
		Donor plasmid is transformed into incorrect competent cells.	Transform donor plasmid into BW23474
25	One day after transfection, insect cells got contaminated	MultiPrime Plasmid is not sterile	Filtrate plasmid through 0.22uM filter, or/and adding 1% P/S to SFII 900 medium
28	One day after transfection, a lots of dead cells are observed	Present P/S in the SF-900 II SFM medium	Repeat transfection with SF-900 II SFM medium without any P/S
		Use more than 8ul of Cell fectin II	Use only 8ul of cell fectin for transfection
30	Insect cells are not infected at all	Transfection was not efficient	Throw away this virus and come back to step 25 and Transfect new insect cells again. Just before adding A+B mixture to the 6-well plate, remove 1000 ul of the 6-well medium (now total volume is 1000 ul) and then add A+B cocktail to medium as drop wise
	20-30% (less than 50%) of insect cells are infected	Titer of V0 virus was not high	Add 1.5 ml SF-900 II SFM or 1% FBS + 1% PS medium to the T-flask and keep them for 18-22 h more.
35	Mammalian cells are	Insect cell medium is not good for mammalian cells	[1] use less amount of un-concentrated (e.g, 500-1000

	rounded and starting to detach	Baculovirus titration is very high	ul) and add more mammalian cell medium (1-1.5 ml) [2] use concentrated virus (250 ul virus+1750 ul medium) [3] decrease incubation time from 8h to 4h or even 2h at 37^{oc} Use a lower virus concentration for mammalian cell transduction
38	Infection rate is not high	Virus titer is not high	Titer virus and use a higher MOI for infection
38	One or some of interested genes do(se)not express	Over amplification of virus let to deletion of genes	Throw away virus and repeat virus production again. Use a maximum incubation time of 5h for V0, 72h for V1 and 48 h for V2

Table 2.a: Concentration of antibiotics and essential components require for MultiPrime vectors

Vector name	Acceptor or donor	Antibiotic resistance gene	Stock of antibiotic concentration (mg/ml)	Dilution in use
pSI- AGZ10, pSI- AGR10, pSI- AGL10, pSI- AGL10-DM, pSI- AGL10-DI	Acceptor	Gentamycin	10	1:1000
pSI- DSZ2cx-DI, pSI-DSZ2cx-DM, pSI- DSZ2cx-PGK, pSI- DSZ2cx-SV40, pSI- DSZ2cx-UBC,	Donor	Spectinomycin	50	1:1000
pSI- DAZ2cx-DI, pSI-DAZ2cx-DM	Donor	Ampicillin	100	1:1000
pSI- DCP-DI, pSI- DCP-DM	Donor	Chloramphenicol	34	1:1000
pSI- DKL-DI, pSI- DKL-DM	Donor	Kanamycin	50	1:1000

* Add a combination of antibiotics when they are assembled. For example when pSI-AGR10 and pSI-DAZ2cx are assembled, you should add Gentamycin plus Ampicillin to LB agar.
** Reduce antibiotics concentration to 50% when more than 3 MultiPrime plasmids are assembled

Table 2.b: Concentration of antibiotics and essential components require for MultiPrime baculovirus generation in LB agar

Component	Concentration	Dilution in use	Note
Antibiotics	Ampicillin (100 mg/ml), Kanamycin (50 mg/ml), Gentamycin (10 mg/ml), Chloramphenicol (34 mg/ml), tetracycline (15 mg/ml)	1:1000	Reduce antibiotics concentration to 50% when more than 3 MultiPrime plasmids are assembled
X-gal	20 mg/ml	Add 100 ul to each plate (100ul per 20 ml LB agar)	Add in last step and should be light protected
IPTG	1000 mM	0.5 mM	Add in last step and should be light protected

CHAPTER 4

SOFTWARE FOR IMAGE ANALYSIS

Quantitative image analysis is the extraction of meaningful information from biological images or movies obtained from fixed or live-cell imaging. In this chapter, I introduced two original papers related the development of image analysis software including Squassh, Squassh3C and SquasshAnalyst. These tools allow quantification of spatial colocalization between fluorescently tagged species with known labeled subcellular markers in two fluorescent channels (Squassh, Rizk et al., 2014) or three fluorescent channels (Squassh3C, Rizk et al., 2015). Both are based on a segmentation technique combining segmentation and deconvolution in a single step, which was shown to yield better results than computing methods. SquasshAnalyst is a data analysis software distributed with user friendly interface providing data browsing, statistical analysis and figure generation.

My role in these papers included construction of biosensors (e.g. fluorescently labeled Rab GTPases), generation of related MultiPrime virus, transfection and/or transduction of different mammalian cells, acquisition of fixed and live cells images and movies and interpretation of results.

Segmentation and quantification of subcellular structures in fluorescence microscopy images using Squassh

Aurélien Rizk[1,2], Grégory Paul[2,5], Pietro Incardona[2], Milica Bugarski[1], Maysam Mansouri[1], Axel Niemann[3], Urs Ziegler[4], Philipp Berger[1] & Ivo F Sbalzarini[2]

[1]Paul Scherrer Institute, Biomolecular Research, Molecular Cell Biology, Villigen PSI, Switzerland. [2]MOSAIC Group, Center of Systems Biology Dresden, Max Planck Institute of Molecular Cell Biology and Genetics, Dresden, Germany. [3]ETH Zurich, Institute of Molecular Health Sciences, Zürich, Switzerland. [4]Center for Microscopy and Image Analysis, University of Zurich, Zürich, Switzerland. [5]Present address: ETH Zurich, Computer Vision Laboratory, Zürich, Switzerland. Correspondence should be addressed to P.B. (philipp.berger@psi.ch) or I.F.S. (ivos@mpi-cbg.de).

Published online 13 February 2014; doi:10.1038/nprot.2014.037

Detection and quantification of fluorescently labeled molecules in subcellular compartments is a key step in the analysis of many cell biological processes. Pixel-wise colocalization analyses, however, are not always suitable, because they do not provide object-specific information, and they are vulnerable to noise and background fluorescence. Here we present a versatile protocol for a method named 'Squassh' (segmentation and quantification of subcellular shapes), which is used for detecting, delineating and quantifying subcellular structures in fluorescence microscopy images. The workflow is implemented in freely available, user-friendly software. It works on both 2D and 3D images, accounts for the microscope optics and for uneven image background, computes cell masks and provides subpixel accuracy. The Squassh software enables both colocalization and shape analyses. The protocol can be applied in batch, on desktop computers or computer clusters, and it usually requires <1 min and <5 min for 2D and 3D images, respectively. Basic computer-user skills and some experience with fluorescence microscopy are recommended to successfully use the protocol.

INTRODUCTION

An increasing amount of biological and medical research relies on single-cell imaging to obtain information about the phenotypic response of cells to a variety of chemical, mechanical and genetic perturbations. Although it is possible to distinguish obvious phenotypes by eye, computational analyses enable the following: the processing of large data sets; the generation of quantitative and less biased results; and the detection of subtler changes in phenotype by statistical analysis. In addition, the ability to observe and quantify multiple fluorescent markers in the same cell under various conditions and over time opens doors for spatiotemporal modeling of biological processes[1].

Quantification of the shapes and spatial distributions of subcellular objects is an important task, as the fluorescent probes are usually related to cellular markers. Colocalization of objects between different color channels can be quantified by various methods that are either pixel-based or object-based[2]. Pixel-based methods compute an overlap measure between the pixel intensities of the different color channels and include the following: Pearson's correlation coefficient[3]; the overlap and Manders' overlap coefficients[4]; intensity correlation[5]; cross-correlation[6]; and techniques correcting for unspecific (random) colocalization[7]. Object-based methods first detect and delineate the objects represented in the image and then quantify their overlap[3,8] or nearest-neighbor distances[9]. This approach also allows correction for the cellular context and for unspecific colocalization[8]. Methods based on intensity correlation are not suitable when fluorophores are not in ratiometric numbers[10], whereas methods looking for co-occurrence are very sensitive to noise and background levels. A more robust method that indicates the fraction of colocalized molecules on the basis of cross-correlation and autocorrelation has been proposed, but it cannot be automatically applied to a set of images[11].

Object-based analysis provides access to additional features of biological relevance, such as the spatial distribution of objects within the cell and the shapes and sizes of objects. This allows interactions between objects to be inferred and statistical hypotheses about their distribution (e.g., random versus nonrandom) to be tested[9]. However, object-based methods require that the objects in the image be first detected and delineated, which is a nontrivial task that involves image segmentation. Object detection and image segmentation are still frequently done by hand or by using *ad hoc* heuristics such as thresholding or hand-crafted pipelines of filters. However, recent progress in computer vision has provided well-founded theories that can give justification to the methodology[12].

An overview of Squassh and its advantages

Here we present a protocol for Squassh and the colocalization of subcellular shapes. Squassh makes use of a segmentation method that directly connects the image-segmentation task with biological reality through prior knowledge about the imaged objects, the image-formation process and the noise present in the image. This allows the same method to be applied to a wide spectrum of images by adjusting the prior knowledge (i.e., changing parameter values). In addition, the segmentation method used here provides theoretical performance and robustness guarantees, is independent of manual initialization and directly corrects for microscope blur and detector noise, yielding optimally deconvolved segmentations[13]. This last feature is achieved by accounting for the microscope's point-spread function (PSF), improving the capacity to segment objects with sizes close to the resolution limit. The algorithm makes no assumptions about the expected shapes of the segmented objects, hence minimally biasing the results. The algorithm is not limited to spot-like or spherical objects, and it can

be applied to the segmentation of more complex shapes, as shown in ref. 13 for an endoplasmic reticulum and epithelial tissue.

Squassh uses the segmentation algorithm from ref. 13 by implementing it in a user-friendly multithreaded software plug-in for the free open-source bioimage processing frameworks ImageJ[14,15] and Fiji[16], as part of the MosaicSuite. We extend the original algorithm by accounting for the fact that different objects may have different fluorescence intensities[17], and by allowing subpixel accurate segmentations in both 2D and 3D. Subpixel segmentation has previously only been available in two dimensions[18,19], and it has been shown to be useful, e.g., for studying the live morphology of endosomes[20]. We further extend the method with cell masks that allow the analysis to be restricted to a subpopulation of cells in an image, e.g., to transfected cells. The workflow is completed with a script, automatically generated by the plug-in as part of its output, for the free open-source statistical software program R (ref. 21), thereby providing statistical significance tests for the quantitative data generated.

Limitations of Squassh

The quality and reliability of Squassh analysis mainly depends on the success of image segmentation. We recommend visually checking the image-segmentation outcome in at least one case per condition in order to confirm its quality. The algorithm used in Squassh accounts for the linear characteristics of the microscope, but imposes no further prior knowledge. Although this helps limit bias in the analysis, it also solely relies on the image data, and segmentation artifacts may occur. The most frequent problem is that objects in the same color channel that are separated by less than the half-width of the PSF will be fused and detected as a single object. In addition, very dim objects may be missed altogether. Although this could potentially be avoided by imposing a shape prior (e.g., that all objects should be roundish), we choose not to do so, as such priors always bias the result toward objects of the sought-for shape. We prefer segmenting all shapes with equal probability, and defer any shape-based filtering to the postprocessing step. There, objects below, e.g., certain sphericities can be filtered out by using the provided statistical analysis script.

The Squassh software also does not correct for chromatic aberration and other nonlinear optical effects. Positional shift between color channels must be corrected for either before (by image warping) or after (by object coordinate transformation) Squassh analysis.

Colocalization analysis is limited to detecting overlap between objects from two-color channels, and it cannot be used to infer patterns within a single channel or long-range order in the object distribution across channels. A separate method is available for that[9] and implemented in software[22].

The Squassh protocol is limited to fluorescence microscopy. Other imaging modalities are not currently supported.

Image segmentation

Image segmentation is a well-researched topic in computer vision, and many technological advances have successfully been applied to bioimage analysis[12]. Many user-friendly software tools are available for analyzing and quantifying fluorescence microscopy images[23]. They are either based on applying various filters to the pixels of the image (e.g., CellProfiler[24]), on using machine-learning techniques to classify pixels as belonging to an object or to the background (e.g., Ilastik[25]), or on including models of the imaged objects and the image-formation process (e.g., Region Competition[19]). Filter-based approaches require the user to construct an appropriate pipeline of filters and to select all filter parameters. Machine-learning approaches require the user to manually label or segment a part of the data for the algorithm to learn the task. Model-based approaches require the user to design or choose a model that appropriately describes the type of image to be segmented.

Model-based methods aim to find the segmentation that best explains the image. In other words, they compute the segmentation that has the highest probability of resulting in the actually observed image when imaged with the specific microscope used. We chose this framework because it is generic to a wide range of image types and segmentation tasks, and because it provides direct access to statistical quantitates that can be used in downstream analyses. We based our work on a recent extension of a family of image-segmentation models that include a variety of denoising and deconvolution tasks[13]. The image-segmentation method used in Squassh is independent of initialization and robustly finds the optimal solution, because the underlying optimization problem is convex and hence has only globally optimal solutions; see ref. 13 for theoretical and algorithmic details.

We consider an image model where the intensity distribution within each object is homogeneous and where the image locally around each object consists of two regions: a brighter foreground and a darker background. The noise in the image can be either Gaussian or Poisson. Three quantities are estimated for each object in a given image: the segmentation, which is the estimated outline of the object; the fluorescence intensity in the local background around the object; and the fluorescence intensity in the interior of the object. The method used here provides optimal segmentations using prior knowledge about the microscope optics (see **Supplementary Note** for details). This strategy hence combines image denoising, deconvolution and segmentation[13]; it is not necessary to separately denoise or deconvolve the images beforehand. The results provided by such a joint deconvolution-segmentation procedure are of higher quality and are more robust than those obtained by standard deconvolution, followed by segmentation[13]. This is because the two tasks naturally regularize each other when considered jointly. This helps cope with very noisy data and with small objects close to the diffraction limit.

We account for spatial variations in the background intensity and for different objects having different foreground intensities by iteratively applying the procedure to local windows in the image around each object. Hence, each object may have a different estimated intensity, and the background around each object may locally be different (**Fig. 1**).

Overview of the procedure

Segmentation procedure. We illustrate the segmentation procedure implemented in Squassh by using an example image of Cherry-tagged RAB5 endosomes (**Fig. 1a**,**b**). First, object detection and separation is performed (segmentation stages (S)1–4), followed by estimation of the local background and object intensities and computation of the optimal segmentation (S5–7). The individual stages are described in detail below and referred to in the PROCEDURE.

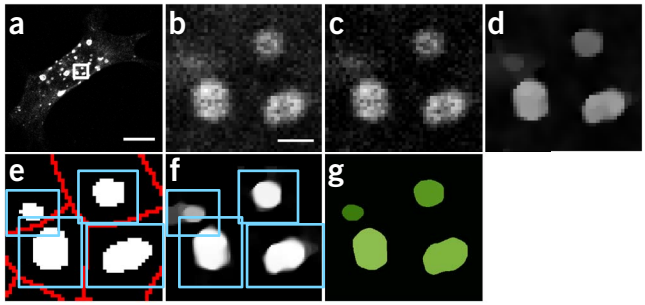


Figure 1 | Workflow of the Squassh protocol illustrated on endosome segmentation. (**a**) Original image of a Cherry-RAB5–transfected HEK293 cell; scale bar, 10 µm. (**b**) Close-up of the region highlighted by the white square in **a**; scale bar, 1 µm. (**c**) The same close-up after background subtraction. (**d**) Object model image found by Squassh. Intuitively, this is a denoised and deconvolved version of **c** taking into account the microscope's PSF. (**e**) Objects (white) obtained by thresholding image **d**. The image is decomposed into regions (blue boxes), and objects are separated by Voronoi decomposition (red lines). (**f**) Refined subpixel object model images obtained by applying the Squassh segmentation method inside each region with individual estimates for the local object and background intensities. (**g**) Final segmentation with the estimated object intensities displayed in shades of green.

- (S1) *Background subtraction* (**Fig. 1c**) is performed first, as the segmentation model assumes locally homogeneous intensities. Background variations are nonspecific signals that are not accounted for by this model. We correct uneven background intensity by using the rolling-ball algorithm[26]. This algorithm computes intensity histograms in a window moving across the image (i.e., a 'rolling ball'). The edge length of the window is set by the user (Step 4 of the PROCEDURE). In each window, the most frequently occurring intensity value is taken as the local background estimate. This is based on the assumption that the objects of interest are smaller than the window size and cover less than half of the total window area.
- (S2) *Object detection* is performed over the entire image (**Fig. 1d**). This stage is not the final segmentation, but only serves to find regions in the image that contain objects of interest. This initial rough detection is done with the model-based algorithm from ref. 13 (see **Supplementary Note** for an intuitive description), but with fixed values for the mean background and object intensities (Steps 6 and 7 of the PROCEDURE). We find these fixed values by performing *k*-means clustering of all the pixels in the image.
- (S3) *Thresholding* of the initial detection reveals the initial objects (**Fig. 1e**). The threshold value is set to the user-defined minimum intensity of objects to be included in the analysis (Step 7 of the PROCEDURE). All objects with peak intensities lower than this threshold are discarded, and the others are retained. This parameter allows the user to control the sensitivity of the analysis, and it can be set to zero without compromising segmentation accuracy. Connected regions are identified as individual objects[19]. Connected regions share a common object and common background intensity. They can, however, be separated into individual objects during the segmentation performed in stages S6 and S7 below.
- (S4) *Decomposition* of the image into smaller parts is done in order to allow locally different background and foreground intensities for different objects (**Fig. 1e**). We decompose the image by (possibly overlapping) 2D or 3D boxes around each detected object from the previous stage (blue boxes in **Fig. 1e**). We ensure that objects in different boxes do not influence each other by computing the Voronoi diagram (on the basis of the actual shapes of the objects and not just their centroids) of the binary mask obtained in stage S3 (red lines in **Fig. 1e**). The image region considered for each object is then the intersection of the box around it with the Voronoi cell containing it (**Fig. 1e**).
- (S5) *Local background and object intensities* are estimated in each image region (**Fig. 1e**). This determines what will be considered background and foreground in the subsequent segmentation stage, and it provides local analysis of the image (Step 9 of the PROCEDURE). Intensities are estimated by solving the optimization problem described in the **Supplementary Note**.
- (S6) *Individual object segmentation* is obtained by running the algorithm from Paul *et al.*[13] separately for each image region (**Fig. 1f**; Steps 5–7 of the PROCEDURE). This can be done with a spatial resolution that is higher than the pixel resolution of the image, thus providing subpixel accuracy[20] (Step 8 of the PROCEDURE). As this oversampling is only done in local patches around the objects, rather than on the whole image, it has only moderate impact on the computational cost of the whole procedure (**Table 1**). As the segmentation algorithm accounts for the PSF of the microscope (Step 5 of the PROCEDURE), the influence of each pixel on its neighboring pixels is accounted for when segmenting apart individual objects. If the image is more likely to be the result of imaging two separate objects, they are split into two. If the image is better explained by assuming a single, connected object, the object is kept together. This always provides the segmentation that has the highest probability of explaining the observed image, given the knowledge of the PSF and the imaging noise model. Any further prior information, for example, about the expected shapes of objects, can be included in the postprocessing analysis.
- (S7) *The final segmentation* (**Fig. 1g**) is obtained by optimizing stage S6 to minimize the segmentation error, automatically done according to a rigorous optimality theory as previously described[13].

TABLE 1 | Computer time and memory requirements on a dual-core 2.3-GHz Intel Core i5 with 8 GB of RAM.

$x \times y \times z$ image size (pixels) and subpixel oversampling	Computer time (s)	Memory (MB)
512 × 512	17.2	275
512 × 512, 8× pixel oversampling	23.7	407
1024 × 1024	55.6	478
512 × 512 × 15	180.7	1,116
512 × 512 × 15, 4× pixel oversampling	201.3	1,536

Colocalization analysis. Object-based colocalization is computed after segmenting the objects by using information about the shapes and intensities of all objects in both channels. This allows straightforward calculation of the degree of overlap between objects from the different channels. We consider three different colocalization measures (see **Supplementary Note** for details). The first one, C_{number}, counts the number of objects that overlap between the two channels. Objects are considered overlapping if at least 50% of their volumes coincide. The second measure, C_{size}, quantifies the fraction of the total volume occupied by objects that overlap with objects from the other channel. The third measure, C_{signal}, quantifies colocalization in an intensity-dependent manner. It computes the sum of all pixel intensities in one channel in all regions where objects overlap with objects from the other channel. This signal-based definition has been described and used before[8], but here we use the estimated object intensities from the segmentation rather than the raw pixel values. This improves robustness against noise and optical blur from the microscope PSF, because the current segmentation method computes an optimally denoised and deconvolved estimate.

Cell masks. Often, analysis should be restricted to a single cell or a subset of the cells present in an image. Examples include working with transfected cells or with mixed cell populations. The Squassh software provides the option to compute cell masks by thresholding the original image and filling the holes of the resulting binary mask. Analysis is then restricted to the intersection of the image with the cell mask in order to ensure that only positive cells are considered in both channels.

Statistical analysis. Subcellular structures are often studied across a range of biological perturbations. A script for the R free open-source statistical software[21] is automatically generated by the plug-in, and it can be used to perform one-way ANOVA[27], followed by a Tukey-Kramer test[28] for the statistical significance of differences observed between different sets of data (**Supplementary Note**).

MATERIALS

EQUIPMENT

- Confocal or wide-field fluorescence microscope. When you are using the protocol for colocalization analysis, the microscope must be capable of two-channel acquisition
- Image data files in any format supported by ImageJ (see http://rsbweb.nih.gov/ij/features.html for supported formats)
- A computer running Linux, MacOS X or Microsoft Windows with at least 2 GB of RAM (4 GB of RAM for 3D images)
- The free ImageJ (http://rsbweb.nih.gov/ij/) or Fiji (http://www.fiji.sc) software installed on the computer
- The free MosaicSuite plug-in, available from http://mosaic.mpi-cbg.de/?q=downloads/imageJ. In addition to Squassh, MosaicSuite also provides software for single-particle tracking[29], image segmentation[19], spatial pattern analysis[9,22] and various utilities (for example, for estimating the microscope PSF from images)
- (Optional for statistical analysis) The free statistical analysis software R (http://www.r-project.org/) to use the plug-in-generated script
- (Optional for reading Leica .lif files) Leica .lif file extractor ImageJ/Fiji macro (provided in the **Supplementary Data**) to automatically extract individual images from .lif files. It uses BioFormats[30] and the channel rearrangement tools of ImageJ

EQUIPMENT SETUP

Installing Squassh in ImageJ Download the MosaicSuite plug-in JAR file (http://mosaic.mpi-cbg.de/?q=downloads/imageJ) and Install the plug-in by moving the file into the plug-in subfolder of ImageJ, or by dragging and dropping the file into the ImageJ main window (**Supplementary Video 1**).

Installing Squassh in Fiji Register the update site http://mosaic.mpi-cbg.de/Downloads/update/Fiji/MosaicToolsuite/ with your Fiji. In order to do so, open the Fiji Updater, then click on 'Advanced', and then on 'Manage update sites'. Click 'Add' and paste the URL to add the new update site. After adding the update site, run a Fiji update or select 'Mosaic ToolSuite.jar' from the list of files of the newly added update site, then restart Fiji. The Mosaic submenu should now appear under the 'Plug-ins' menu. The advantage of using the update site is that the Fiji installation will automatically upgrade MosaicSuite every time a new version of it becomes available in the future. If automatic updates are not required, use the manual installation procedure described for ImageJ (**Supplementary Video 1**).

Installing the .lif file extractor macro Follow the ImageJ/Fiji menu path 'Plug-ins' → 'Macros' → 'Install' to install the macro. Alternatively, paste the macro content into 'Plug-ins' → 'Macros' → 'Startup Macros' for a permanent installation.

PROCEDURE

Image data preparation

1| Export images from the microscope software. The plug-in works with any image format supported by ImageJ, but all stacks and channels of a single image have to be in the same .tif file. For Leica .lif files, use the .lif Extractor ImageJ macro to automatically extract dual-channel TIFF images. To do so, select 'Plug-ins' → 'Macros' → 'Lif Extractor' from the ImageJ/Fiji menu-bar and choose a folder containing one or more .lif files. A new folder containing the extracted TIFF images is then created for each .lif file.

▲ **CRITICAL STEP** Channel order needs to remain the same across all files. The Squassh software refers to 'Channel 1' and 'Channel 2' to indicate the location of segmented objects and for colocalization measures. It is possible to attribute other, clear-text names to the channels in the 'Visualization and output' settings (Step 15).

▲ **CRITICAL STEP** For colocalization analysis, care has to be taken to select spectrally separated fluorophores in order to avoid cross talk between the color channels, which would lead to spurious colocalization. Before computing colocalization between different color channels, it is important to correct for chromatic aberration as well as possible, as this biases the results[9].

? TROUBLESHOOTING

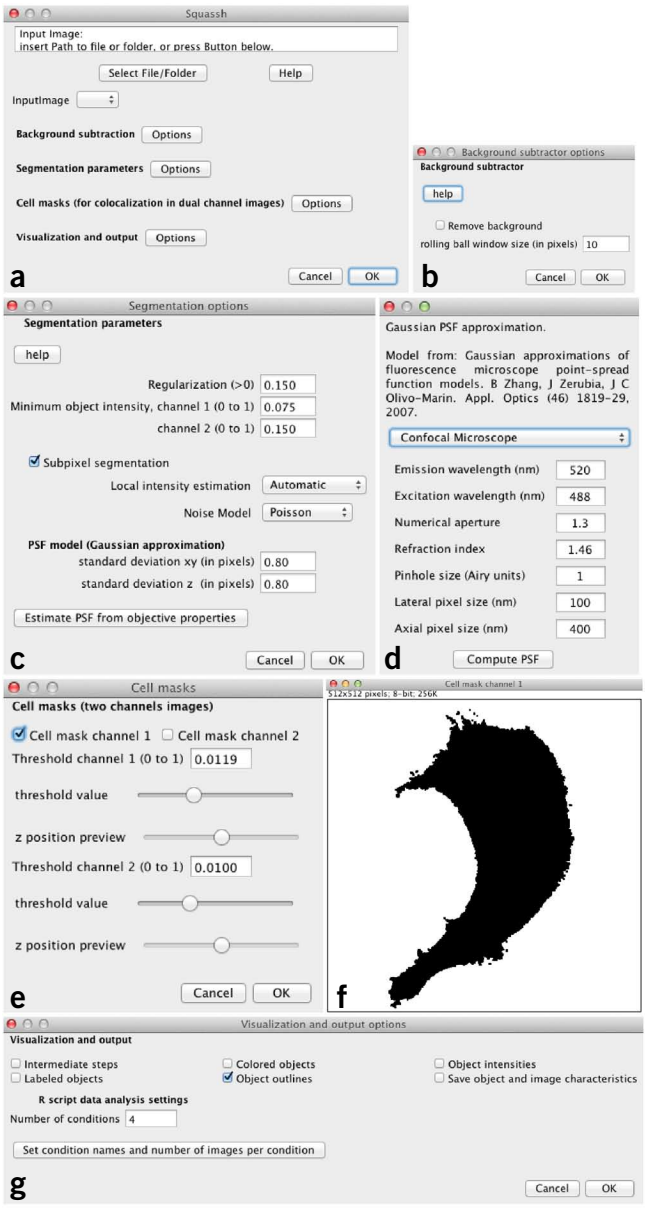


Figure 2 | Screenshots of the graphical user interface of the Squassh software. (**a**) Main window with file selection, help and links to background subtraction, segmentation, cell masks and visualization options. (**b**) Background-subtraction options. (**c**) Segmentation parameter options. (**d**) Window for specifying the microscope's point-spread function by using a Gaussian model. (**e**) Cell mask window. (**f**) Cell mask preview. (**g**) Visualization and output options.

Plug-in launch and input selection

2| Start the segmentation by choosing the ImageJ menu item 'Plug-ins' → 'Mosaic' → 'Segmentation' → 'Squassh'. This opens the main graphical user interface Squassh window shown in **Figure 2**. Protocol Steps 4–16 indicate how to navigate the four option windows that can be opened from this main window.

? TROUBLESHOOTING

3| Select the image file to be processed with the 'Select File/Folder' button (**Fig. 2a**). When selecting files containing two channels, both channels will be segmented and object-based colocalization analysis will be performed. For single-channel files, only segmentation is performed. For batch analysis, select a folder containing multiple images. It is recommended to first analyze a few individual images to determine suitable parameters that can then be applied across the whole set of images. Time-lapse images are not natively supported; they should be split into the different time points and analyzed separately. It is possible to batch-process a set of images by placing them in the same folder and selecting this folder as Squassh's input. The software is also compatible with ImageJ macros, enabling batch automation of larger analyses.

Background subtraction

4| Reduce the background fluorescence by using the rolling ball algorithm. Click on 'Background subtraction' options, select 'Remove background' and enter the window edge length in units of pixels (**Fig. 2b**). This length should be large enough so that a square with that edge length cannot fit inside the objects to be detected, but smaller than the length scale of background variations. This step corresponds to segmentation stage S1 (**Fig. 1b,c**).

? TROUBLESHOOTING

Segmentation parameters

5| Set the microscope PSF (**Fig. 2c**). To correct for diffraction blur, the software needs information about the PSF of the microscope. Either a theoretical PSF (option A) or a measured PSF (option B).

(A) Using a theoretical PSF

(i) Use a theoretical PSF model[31] by specifying the imaging condition parameters in the 'Segmentation parameters' options 'Estimate PSF from objective properties' subwindow (**Fig. 2d**). The software provides models for confocal and wide-field microscopes. Airy units only have to be entered for confocal microscopes.

(ii) Click on 'compute PSF'.

(B) Measuring the PSF

(i) Measure the microscope PSF from images of fluorescent subdiffraction beads. In this case, input the s.d. of the PSF separately for the lateral (x, y) and axial (z) directions (**Fig. 2c**). Use the menu item 'Plug-ins' → 'Mosaic' → 'PSF Tool' to measure these parameters from images of beads.

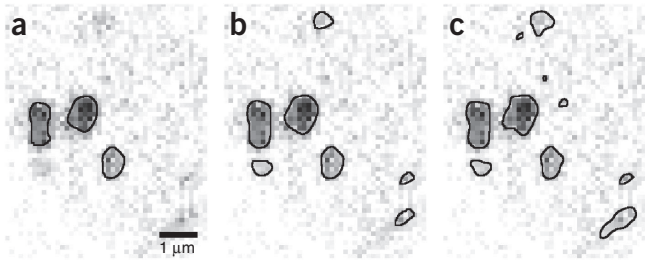


Figure 3 | Illustration of how the parameters affect segmentation results by using endosomes in a close-up view of a Cherry-RAB5-transfected HEK293 cell as an example. (**a**) Segmentation with the minimum intensity threshold set to 0.100 and the regularization weight set to 0.250. (**b**) Segmentation with the minimum intensity threshold decreased to 0.050, resulting in a larger number of dimmer objects being detected. (**c**) Segmentation with a minimum intensity threshold of 0.050 and the regularization weight decreased to 0.075, resulting in a closer fit of the segmentation with the image data, but an increased sensitivity to noise (e.g., the small islands segmented, which may correspond to noise pixels).

6| Set a regularization parameter for the segmentation (**Fig. 2c**). Use higher values to avoid segmenting noise-induced small intensity peaks (**Supplementary Note** and **Fig. 3**). Typical values are between 0.05 and 0.25. This parameter controls segmentation stages S2 and S6 (**Fig. 1d,f**).
▲ **CRITICAL STEP** Regularization is a key segmentation parameter. The effect of varying the regularization parameter is illustrated in **Figure 3**. Once determined, the same parameter must be used across all studied images in order to allow comparison of the results.
? TROUBLESHOOTING

7| Set the threshold for the minimum object intensity to be considered (**Fig. 2c**). Intensity values are normalized between 0 for the smallest value occurring in the image and 1 for the largest value. This parameter controls segmentation stages S2 and S6 (**Fig. 1d,f**). Use low values for increased sensitivity and high values to force object separation.
▲ **CRITICAL STEP** Minimum intensity is a key segmentation parameter. The effect of varying the minimum intensity parameter is illustrated in **Figure 3**. Once determined, the same parameter must be used across all studied images to allow comparison of the results.
? TROUBLESHOOTING

8| (Optional) Select 'subpixel segmentation' to compute segmentations with subpixel resolution in stage S6 (**Fig. 2c**). The resolution of the segmentation is increased by an oversampling factor of 8 for 2D images and a factor of 4 for 3D images. Note that subpixel segmentation requires more computer time and memory. See **Table 1** for typical requirements.

9| (Optional) When images contain objects with inhomogeneous internal intensity distribution, the automatic local intensity estimation performed in segmentation stage S5 may not always be suitable. The 'Local intensity estimation' parameter thus provides the ability to estimate local object intensities by clustering (**Fig. 2c**). Select 'Low', 'Medium' or 'High' to set the local object intensity to the low-, medium- or high-intensity clusters obtained by *k*-means clustering of the pixels within the object. All results and benchmarks given in ANTICIPATED RESULTS section are obtained with the 'Automatic' setting, which is the recommended default setting.

10| Set the noise model corresponding to the microscope and detector used (**Fig. 2c**). We recommend using the Poisson model for confocal microscopes and the Gaussian model for wide-field microscopes. However, your equipment may differ. Detailed information should be found in the manufacturer's data sheet.

Cell masks

11| A cell mask allows the analysis to be restricted to a certain region of an image, e.g., to a transfected cell. Select 'cell mask channel 1' and/or 'cell mask channel 2' to compute cell masks on the basis of the respective channel. Cell masks are computed by thresholding the respective channel and filling holes in the obtained binary mask. Open the 'Cell masks' option (**Fig. 2e**) and set a value between 0 and 1 for the threshold in 'threshold channel 1' and/or 'threshold channel 2'. Adjust the threshold slider for a live preview (**Fig. 2f**) of the resulting mask. Adjust the *z*-slider to simultaneously adjust the *z*-position in the mask preview and the original image.
▲ **CRITICAL STEP** Once they have been determined, the same cell mask parameters must be used across all studied images in order to allow comparison of the results. When set of images is analyzed, cell masks are saved in files ending in 'mask c1.zip' and 'mask c2.zip' for the respective channel. Open them with ImageJ in order to check that the cell masks are correct for all images.
? TROUBLESHOOTING

Visualization and output

12| Select one or several of the output visualization options 'Colored objects', 'Labeled objects', 'Object intensities' and 'Object outlines' (**Fig. 2g**). With 'Colored objects', each object is visualized in a different, random color. 'Labeled objects'

assigns to all pixels of an object a value that is identical to the label (index) of that object in the result file. 'Object intensities' displays objects in their estimated fluorescence intensities. 'Object outlines' shows an overlay of the original image with the segmented outlines in red (**Supplementary Note**). For two-channel images, an additional visualization is displayed, which attributes a distinct color to each channel in order to reveal colocalization.

13| (Optional) Select 'Intermediate steps' to visualize also the background subtraction result (step S1; **Fig. 1c**) and the initial segmentation approximation (step S2; **Fig. 1d**).

14| (Optional) Select 'Save object and image characteristics' to store the resulting object quantifications in .csv files, which can be opened with any spreadsheet software or with MATLAB. This option is always selected by default when applying the analysis in batch mode to a set of multiple images (**Fig. 2g**).

15| (Optional) In order to generate an R script for statistical analysis of the results, set the number of experimental conditions considered under 'R script data analysis settings'. Click 'Set condition names and number of images per condition' to set the number of images per condition, the condition names (you can freely choose them) and the object names (can also be freely chosen). Names are used to generate graphical output and plots in human-readable form. The number of conditions and the number of images per condition are used to divide the set of images into the biological conditions studied. The set of images is divided according to the lexicographical ordering of the image file names.

Run the analysis

16| Click 'OK' in the main window (**Fig. 2a**) to start the analysis. Depending on the speed of the computer, the analysis may take up to several minutes, and on older computers it may take even longer. See **Table 1** for typical runtimes on a typical modern computer.

? TROUBLESHOOTING

Quantitative results

17| Open the ImageJ log window for a preview of the analysis results. The log displays the number of objects found and the signal colocalization C_{signal} (**Supplementary Note**).

▲ **CRITICAL STEP** C_{signal} should only be used if the same microscope gain and offset settings were used to acquire all images.

18| Use any software that can read .csv files in order to open the file ending in '_ImagesData.csv'. This file contains the mean object features from each image. It stores the number of objects found in the image, the mean object size (in terms of area in 2D or volume in 3D), the mean object surface in 3D or perimeter in 2D, the mean length of objects (i.e., the maximum extension in the most extended direction) and the mean fluorescence intensity of the objects. For two-channel images, the file also stores the colocalization coefficients C_{signal}, C_{number} and C_{size}, as well as the Pearson correlation across the whole image and inside the cell mask. All .csv data files are contained in the folder in which the original images are located.

19| Open files ending in '_ObjectsData.csv' in order to obtain individual, per-object features. Each object is indexed by its segmentation label. For two-channel images, the file also stores the amount of overlap for each object with objects from the other channel, and the sizes and intensities of all colocalizing objects from the other channel. These features are used by the R script to statistically evaluate colocalization.

Graphical output and statistical analysis (optional)

20| (Optional) Edit lines 23 to 27 of the automatically generated statistical analysis script 'R_analysis.R' to set thresholds for minimum and maximum object sizes, minimum object intensity and minimum number of objects in an image. Objects and images violating the thresholds will be excluded from the analysis.

? TROUBLESHOOTING

21| (Optional) Edit line 32 of 'R_analysis.R' to change the order in which the conditions are displayed in the resulting plots and graphs.

22| Start the R statistical software. Change to the working folder where 'R_analysis.R' is located and type 'source('R_analysis.R')' in the R console. This generates bar plots and performs statistical analysis of the data with one-way ANOVA[27] followed by a Tukey-Kramer test[28] for statistical significance of differences between different conditions. 'Colocalization.pdf' (**Supplementary Note**) displays the colocalization measures defined in the **Supplementary Note** and the Pearson correlation coefficients in the whole image or inside cell masks. It also provides the *P* value from a one-way ANOVA. 'ColocalizationCI.pdf'

(**Supplementary Note**) gives 95% confidence intervals for the difference of colocalization between pairs of conditions, together with their *P* values. The confidence interval and the *P* values are computed by a Tukey-Kramer test. The Tukey-Kramer test is suited for multiple comparisons (e.g., if there are more than two conditions), and it provides *P* values to estimate the statistical significance of differences between pairs of conditions. ANOVA *P* values test the hypothesis that all conditions have identical means. The two last pdf files generated contain the mean object properties across all conditions. See the **Supplementary Note** for examples of the output generated by this script.

▲ **CRITICAL STEP** If the number of objects detected for any condition is small (fewer than about ten), or the statistical distribution of features significantly differs from Gaussian, the results of the statistical test are unreliable and should not be trusted.

? TROUBLESHOOTING

Troubleshooting advice can be found in **Table 2**.

TABLE 2 | Troubleshooting table.

Step	Problem	Possible reason	Solution
1	The 'Lif Extractor' menu item is not present	The .lif file extractor macro has not been installed	Install the macro as described in the Equipment section
2	The Squassh menu item is not present	The MosaicSuite has not been installed	Install the MosaicSuite as described in the Equipment section
4	Background removal removes objects of interest (check by selecting the visualization option 'intermediate steps')	Inappropriate parameter settings	Reduce the edge-length parameter
6,7,16	Segmentation tends to fuse neighboring objects	Image is saturated	Avoid saturation by adjusting microscope gain and offset. There should be no more than a few, isolated saturated pixels
		Inappropriate parameter settings	Decrease the regularization parameter
	Segmentation is too sensitive. Background perturbations are interpreted as objects	Inappropriate parameter settings	Increase the regularization parameter to remove small low-intensity objects. Increase the minimum intensity parameter to remove low-intensity objects
11	Masks from two positive cells fuse together and include parts of a negative cell	Cell spacing is too small	Use images in which cells are farther apart or manually edit/curate the mask to exclude the space between two positive cells
16	The plug-in hangs during computation	There is not enough memory available or the image is too large	Analyze a down-sampled version of the image. Image size can be lowered from ImageJ by selecting 'Image' → 'Adjust' → 'Size'. If this solves the problem, it is a memory issue. Increase ImageJ maximum memory by selecting 'Edit' → 'Options' → 'Memory'. Memory requirements are given in **Table 1**
		Computer is slow or data set is large	Wait longer. See **Table 1** for typical runtimes on a modern computer. On older computers, the software can take markedly longer to complete the task
20	Some segmented objects are not biologically relevant and should be removed from the analysis	Fluorescence is not specific to the studied structures	Remove objects from the analysis by setting filters on object size or intensity in Step 20.

● TIMING

The only protocol steps requiring computational time in Squassh are image segmentation (Step 16) and statistical analysis (Step 22). All other steps only require setting parameter values and incur no computational cost. As a guideline, we provide here typical computation times measured on a dual-core 2.3 GHz Intel Core i5 with 8 GB of RAM. Statistical analysis (Step 22) required <1 min for a data set of 100 3D images of sizes between 512 × 512 × 10 and 512 × 512 × 25. **Table 1** provides

typical times for segmentation of 2D and 3D images with and without subpixel refinement. The computational cost only depends on image size for stages S1–S3 of the segmentation procedure. All other stages do not depend on image size, but rather on the sizes and numbers of objects detected in the images. The data in **Table 1** was obtained for images containing ~100 objects each. Image sizes are indicated in the table. For other numbers of objects, the times scale proportionally. The overhead incurred by subpixel oversampling mainly depends on the number and the sizes of the objects, and it ranged from 10 to 40% for the images used here. All pixel-intensity values are normalized between 0 and 1 and stored as 64-bit Java double variables. This renders the computational performance of the software independent of the bit depth of the original images.

ANTICIPATED RESULTS

Figure 4 shows typical images and colocalization results for segmentation of a single dual-channel image. Squassh allows the user to visually confirm that the cell mask appropriately delineates the transfected cell, and that the objects are satisfactorily segmented (close-ups in **Fig. 4e**,**f**). The segmented objects (**Fig. 4f**) appear smaller and more compact than their images (**Fig. 4e**), because the segmentation algorithm corrects for the diffraction blur from the PSF of the microscope, which causes objects to appear more blurry and fused in the image. Dim objects (especially visible in the red channel) are cut off owing to the 'minimum intensity' parameter.

Figure 5 demonstrates a range of applications for Squassh. **Figure 5a** shows an analysis of subcellular localization of four RAB GTPases (RAB5, RAB4, RAB11 and RAB7) with known subcellular distribution and function, by using the following markers: early endosome antigen 1 (EEA1) as a marker for early endosomes; lysobisphosphatidic acid (LBPA) for late endosomes; lysosome-associated membrane protein 2 (LAMP-2) for lysosomes; and protein disulfide isomerase (PDI) for the endoplasmic reticulum. Two example images with close-ups of EEA1 (green) and RAB5 (red), and EEA1 (green) and RAB7 (red) are shown in **Figure 5b**,**c**, respectively. Example images for all other conditions are shown in **Supplementary Note**. Squassh analysis confirmed previously described localization data of RAB GTPases[32–35]. In addition, it allowed the maturation of endosomes to be monitored quantitatively, either by loss of colocalization with EEA1 or gain of colocalization with LBPA.

Figure 5d demonstrates that Squassh can be used in infection assays; Squassh successfully found a linear correlation between virus titer and the percentage of virus-infected cells in Sf21 cells infected with different amounts of YFP-expressing baculovirus[36].

Figure 5e demonstrates that Squassh can be used to study changes in cell morphology. In cells overexpressing GDAP1, a fission factor, which leads to fragmented mitochondria[37], Squassh was able to detect reduced mitochondria lengths and increased numbers of mitochondria (**Fig. 5e**). The total volume of mitochondria remained constant within the statistical noise (**Fig. 5e**). This is in line with previous manual analyses[37]. Example images from both conditions are shown in **Figure 5f**, with the outlines as segmented by Squassh overlaid in green. We observe the same effect that the segmented outlines seem to be smaller than the apparent objects in the image, because they correct for the diffraction blur from the PSF, hence deconvolving and segmenting at once.

We quantitatively compared the segmentation method used in Squassh with eleven other methods on a test set of noisy synthetic images of cells with fluorescently labeled nuclei and vesicles. All images contained out-of-focus blur and uneven background illumination. The ground truth of all vesicle outlines was available and was used to compute the F-score quality measure. With an F-score of 0.62, the method used in Squassh performs as well as the previous winner on this benchmark[38]. This is remarkable, because the previous winner was a spot detector, tuned to detect round objects such as those present in these images, whereas Squassh makes no assumptions about the shapes of the objects (see ref. 13 for examples of complex-shaped structures, which would not be possible with a spot detector).

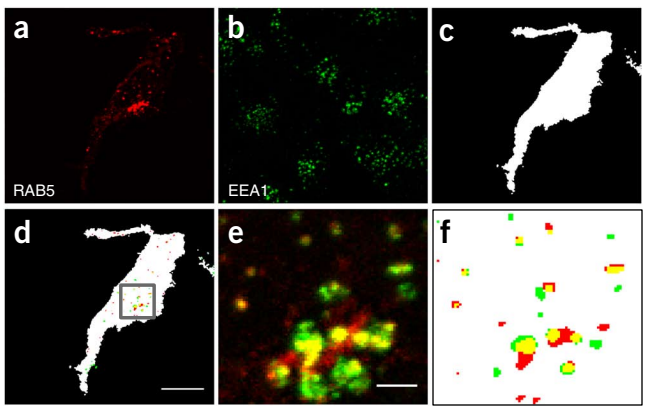


Figure 4 | Segmentation and colocalization of EEA1 and RAB5. HEK293 cells were transiently transfected with a Cherry-RAB5 expression construct. EEA1, a marker for early endosomes, was detected with a rabbit monoclonal antibody (Cell Signaling, cat. no. 3288). All images show maximum intensity projections along the optical axis. (**a**) Raw image of the Cherry-RAB5 channel. (**b**) Raw image in the EEA1 channel. (**c**) Mask of a transfected cell determined by using the RAB5 channel. (**d**) Segmentation results of the vesicles from both channels overlaid with the cell mask. EEA1-expressing vesicles are shown in green, and RAB5-expressing vesicles are shown in red. The colocalization coefficient computed for this image is C_{number} ($EEA1_{RAB5+}$/EEA1) = 0.36; it is C_{number} ($EEA1_{RAB5+}$/EEA1) = 0.02 when not using a cell mask. (**e**) Close-up view of the raw image data in the area highlighted by the black square in **d**. (**f**) Object segmentation and overlaps in this area. Scale bars, 10 µm (**a**–**d**) and 2 µm (**e**,**f**).

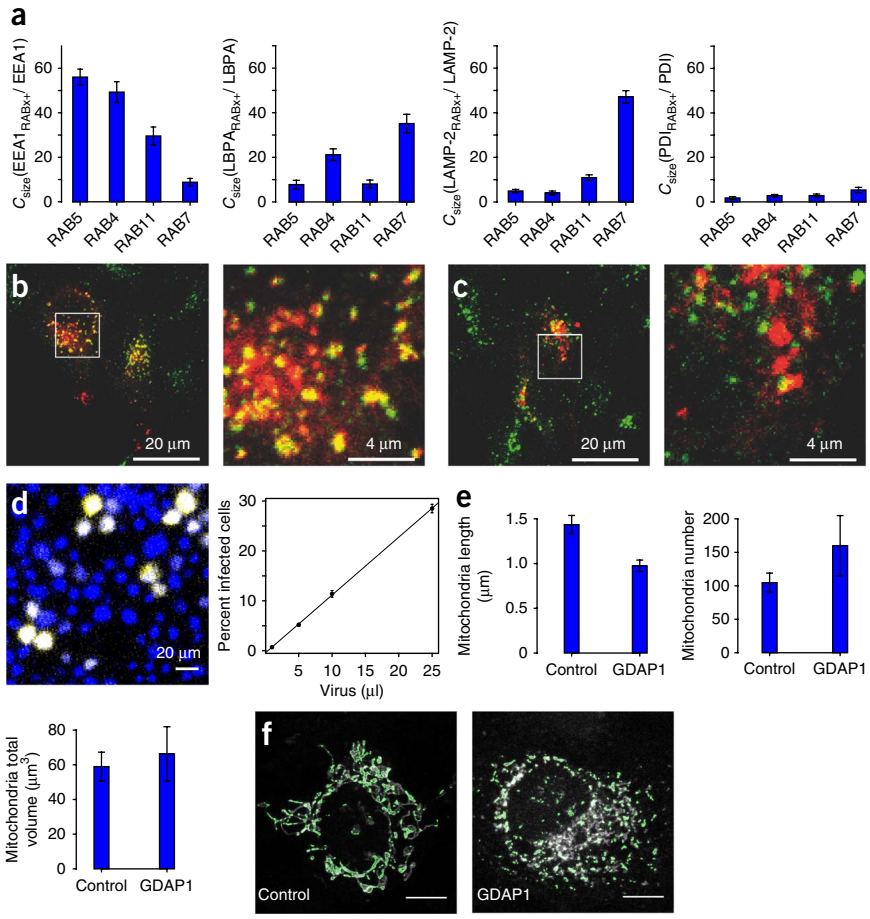


Figure 5 | Anticipated results of the Squassh protocol. HEK293 cells were transiently transfected with Cherry-tagged RAB5, RAB4, RAB11 and RAB7. Subcellular compartments were immunostained with the following antibodies: EEA1 for early endosomes, LBPA for late endosomes, LAMP-2 for lysosomes and PDI for the endoplasmic reticulum. (**a**) Colocalization results for subcellular markers with different RAB GTPases. Error bars show means and s.e.m. over 20 images per condition of the fraction of subcellular marker colocalizing with the RAB channel, as determined by Squassh. 'RABx' represents a RAB in RAB5, RAB4, RAB11 or RAB7. (**b**) Example image with close-up of EEA1 (green) and RAB5 (red). (**c**) Example image with close-up of EEA1 (green) and RAB7 (red). Examples images for all conditions, as well as statistical analyses, are provided in the **Supplementary Note**. (**d**) Sf21 cells were infected with YFP-expressing baculovirus and stained with Hoechst 33258. A linear correlation between the amount of virus and the fraction of infected cells is observed. (**e**) Morphological analysis of mitochondria without and with overexpression of the fission factor ganglioside-induced differentiation-associated protein 1 (GDAP1). Error bars show means and s.e.m. over eight images per condition. (**f**) Example images from both conditions with segmentation outlines in green; scale bars, 10 µm.

The Squassh object-based approach to colocalization analysis was compared with a pixel-based Pearson correlation analysis (**Fig. 6a**). The pixel-based method resulted in spurious colocalization of LAMP-2 with RAB4, RAB5 and RAB11. The object-based Squassh approach yields the best result. Two phenomena affect the pixel-based results: First, many objects in the LAMP-2 channel overlap with objects in the RAB7 channel, but not vice versa (compare the two C_{size} graphs in **Fig. 6b**). In such asymmetric situations, pixel-based colocalization analysis is not appropriate, as it lumps both sides together. Second, pixel-based analysis is more sensitive to imaging noise than object-based analysis. This is shown in **Figure 6b**, where the Pearson correlation score tends to zero with increasing noise level[3] (**Fig. 6c**).

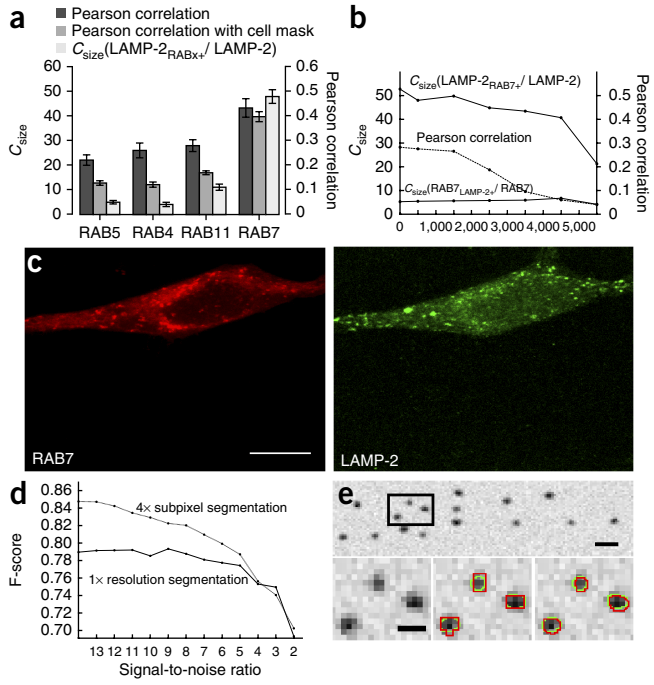


Figure 6 | Benchmarks of the Squassh protocol. HEK293 cells were transiently transfected with Cherry-tagged RAB7. Lysosomes were identified by LAMP-2 immunostaining. (**a**) Pixel-based Pearson correlation (range –1 to +1) and object-based colocalization (range 0–100%) results for LAMP-2 with RABx (where 'x' stands for any RAB). For object-based analyses, images were segmented with the Squassh software. LAMP-2 is known to colocalize mainly with RAB7. Error bars show means and s.e.m. over 20 images per condition. (**b**) Pearson correlation and object-based colocalization coefficients for the image shown in **c** corrupted with increasing amounts of Gaussian noise. (**c**) Example image with both channels shown; scale bar, 10 µm. (**d**) F-score of segmentation accuracy for fourfold subpixel–level and normal pixel-level segmentation. We show the mean values over five random images for each signal-to-noise ratio. (**e**) Example of a benchmark image with a signal-to-noise ratio of 12. The close-up views below show the ground truth outlines in green and those reconstructed by Squassh in red. The middle image uses normal pixel-level segmentation, whereas the right image uses fourfold subpixel oversampling; scale bars, 1 µm for the top image and 0.5 µm for the close-up images.

The accuracy of subpixel segmentation is illustrated on a synthetic image obtained by blurring a previously determined segmentation of RAB5 endosomes with the PSF of the microscope and adding modulatory Poisson noise of various magnitudes. We computed the F-score of the obtained segmentations with respect to ground truth. Subpixel segmentation improved the segmentation accuracy for signal-to-noise ratios (defined for Poisson noise according to ref. 29) above four (**Fig. 6d**). **Figure 6e** shows example results in comparison with standard pixel-level segmentation.

Note: Any Supplementary Information and Source Data files are available in the online version of the paper.

ACKNOWLEDGMENTS We thank J. Cardinale (MOSAIC Group) for help with the software implementation of the ImageJ plug-in and B. Cheeseman (MOSAIC Group) for the voiceover in the video tutorial (**Supplementary Video 1**). This work was supported by SystemsX.ch, the Swiss initiative in systems biology under grant IPP-2011-113, evaluated by the Swiss National Science Foundation. The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 290605 (PSI-FELLOW/COFUND) and from the German Federal Ministry of Research and Education under funding code 031A099.

AUTHOR CONTRIBUTIONS P.B., I.F.S. and U.Z. designed the project; G.P., I.F.S. and A.R. developed the image-processing algorithm; A.R., P.I. and I.F.S. were involved in software development and implementation; A.R., I.F.S., G.P. and P.B. were involved in benchmark design; experimental data were obtained by M.B., M.M., A.N. and P.B.; and A.R., I.F.S. and P.B. wrote the manuscript with input from G.P., A.N. and U.Z. Figures were prepared by A.R. and P.B., and the video tutorial was created by I.F.S.

COMPETING FINANCIAL INTERESTS The authors declare no competing financial interests.

1. Sbalzarini, I.F. Modeling and simulation of biological systems from image data. *Bioessays* **35**, 482–490 (2013).
2. Adler, J. & Parmryd, I. Quantifying colocalization by correlation: the Pearson correlation coefficient is superior to the Mander's overlap coefficient. *Cytometry A.* **77**, 733–742 (2010).
3. Bolte, S. & Cordelieres, F.P. A guided tour into subcellular colocalization analysis in light microscopy. *J. Microsc.* **224**, 213–232 (2006).
4. Manders, E.M.M., Verbeek, F.J. & Aten, J.A. Measurement of co-localization of objects in dual-colour confocal images. *J. Microsc.* **169**, 375–382 (1993).
5. Li, Q. *et al.* A syntaxin 1, $G\alpha_o$, and N-type calcium channel complex at a presynaptic nerve terminal: analysis by quantitative immunocolocalization. *J. Neurosci.* **24**, 4070–4081 (2004).
6. van Steensel, B. *et al.* Partial colocalization of glucocorticoid and mineralocorticoid receptors in discrete compartments in nuclei of rat hippocampus neurons. *J. Cell Sci.* **109** (Part 4): 787–792 (1996).
7. Costes, S.V. *et al.* Automatic and quantitative measurement of protein-protein colocalization in live cells. *J. Biophys.* **86**, 3993–4003 (2004).
8. Ramirez, O., Garcia, A., Rojas, R., Couve, A. & Hartel, S. Confined displacement algorithm determines true and random colocalization in fluorescence microscopy. *J. Microsc.* **239**, 173–183 (2010).
9. Helmuth, J.A., Paul, G. & Sbalzarini, I.F. Beyond co-localization: inferring spatial interactions between sub-cellular structures from microscopy images. *BMC Bioinformatics* **11**, 372 (2010).
10. Bolte, S. & Cordelières, F.P. Reply to letter to the editor. *J. Microsc.* **227**, 84–85 (2007).
11. Zinchuk, V., Wu, Y., Grossenbacher-Zinchuk, O. & Stefani, E. Quantifying spatial correlations of fluorescent markers using enhanced background reduction with protein proximity index and correlation coefficient estimations. *Nat. Protoc.* **6**, 1554–1567 (2011).
12. Danuser, G. Computer vision in cell biology. *Cell* **147**, 973–978 (2011).
13. Paul, G., Cardinale, J. & Sbalzarini, I.F. Coupling image restoration and segmentation: a generalized linear model/Bregman perspective. *Int. J. Comput. Vis.* **104**, 69–93 (2013).
14. Abramoff, M.D., Magalhães, P.J. & Ram, S.J. Image processing with ImageJ. *Biophoton. Int.* **11**, 36–42 (2004).
15. Schneider, C.A., Rasband, W.S. & Eliceiri, K.W. NIH Image to ImageJ: 25 years of image analysis. *Nat. Methods* **9**, 671–675 (2012).
16. Schindelin, J. *et al.* Fiji: an open-source platform for biological-image analysis. *Nat. Methods* **9**, 676–682 (2012).
17. Paul, G., Cardinale, J. & Sbalzarini, I.F. An alternating split Bregman algorithm for multi-region segmentation. in *Proc. 45th IEEE Asilomar Conf. Signals, Systems, and Computers* 426–430 (Asilomar, 2011).
18. Helmuth, J.A. & Sbalzarini, I.F. Deconvolving active contours for fluorescence microscopy images. in *Proc. Intl. Symp. Visual Computing (ISVC)*, vol. 5875 of *Lecture Notes in Computer Science*, 544–553 (Springer, 2009).
19. Cardinale, J., Paul, G. & Sbalzarini, I.F. Discrete region competition for unknown numbers of connected regions. *IEEE Trans. Image Process.* **21**, 3531–3545 (2012).
20. Helmuth, J.A., Burckhardt, C.J., Greber, U.F. & Sbalzarini, I.F. Shape reconstruction of subcellular structures from live cell fluorescence microscopy images. *J. Struct. Biol.* **167**, 1–10 (2009).
21. R Development Core Team. *R: A Language and Environment for Statistical Computing* (R Foundation for Statistical Computing, 2012).
22. Shivanandan, A., Radenovic, A. & Sbalzarini, I.F. MosaicIA: An ImageJ/Fiji plug-in for spatial pattern and interaction analysis. *BMC Bioinformatics* **14**, 349 (2013).
23. Eliceiri, K.W. *et al.* Biological imaging software tools. *Nat. Methods* **9**, 697–710 (2012).
24. Carpenter, A.E. *et al.* CellProfiler: image analysis software for identifying and quantifying cell phenotypes. *Genome Biol.* **7**, R100 (2006).
25. Sommer, C., Straehle, C., Koethe, U. & Hamprecht, F.A. Ilastik: Interactive learning and segmentation toolkit. *Genome Biomedical Imaging: From Nano to Macro, 2011 IEEE International Symposium on* doi:10.1109/ISBI.2011.5872394 (IEEE, 2012).
26. Sternberg, S.R. Biomedical image processing. *Computer* **16**, 22–34 (1983).
27. Chambers, J., Freeny, A. & Heiberger, R. Analysis of variance; designed experiments. in *Statistical Models in S* 145–193 (Wadsworth and Brooks/Cole Advanced Books and Software, 1992).
28. Miller, J.R. *Simultaneous Statistical Inference* (Springer-Verlag, 1981).
29. Sbalzarini, I.F. & Koumoutsakos, P. Feature point tracking and trajectory analysis for video imaging in cell biology. *J. Struct. Biol.* **151**, 182–195 (2005).
30. Linkert, M. *et al.* Metadata matters: access to image data in the real world. *J. Cell Biol.* **189**, 777–782 (2010).
31. Zhang, B., Zerubia, J. & Olivo-Marin, J.C. Gaussian approximations of fluorescence microscope point-spread function models. *Appl. Opt.* **46**, 1819–1829 (2007).
32. Sönnichsen, B., De Renzis, S., Nielsen, E., Rietdorf, J. & Zerial, M. Distinct membrane domains on endosomes in the recycling pathway visualized by multicolor imaging of Rab4, Rab5, and Rab11. *J. Cell Biol.* **149**, 901–914 (2000).
33. Ballmer-Hofer, K., Andersson, A.E., Ratcliffe, L.E. & Berger, P. Neuropilin-1 promotes VEGFR-2 trafficking through Rab11 vesicles thereby specifying signal output. *Blood* **118**, 816–826 (2011).
34. Kriz, A. *et al.* A plasmid-based multigene expression system for mammalian cells. *Nat. Commun.* **1**, 120 (2010).
35. Matsuo, H. *et al.* Role of LBPA and Alix in multivesicular liposome formation and endosome organization. *Science* **303**, 531–534 (2004).
36. Fitzgerald, D.J. *et al.* Protein complex expression by using multigene baculoviral vectors. *Nat. Methods* **3**, 1021–1032 (2006).
37. Niemann, A., Ruegg, M., La Padula, V., Schenone, A. & Suter, U. Ganglioside-induced differentiation associated protein 1 is a regulator of the mitochondrial network: new implications for Charcot-Marie-Tooth disease. *J. Cell Biol.* **170**, 1067–1078 (2005).
38. Ruusuvuori, P. *et al.* Evaluation of methods for detection of fluorescence labeled subcellular objects in microscope images. *BMC Bioinformatics* **11**, 248 (2010).

Subcellular object quantification with Squassh3C and SquasshAnalyst

Aurélien Rizk†, Maysam Mansouri, Kurt Ballmer-Hofer, and Philipp Berger
Paul Scherrer Institute, Biomolecular Research, Molecular Cell Biology, Villigen, Switzerland

† Present affiliation: Algorizk SARL, France

BioTechniques 59:309-312 (November 2015) doi 10.2144/000114352
Keywords: image analysis; segmentation; co-localization; live-cell imaging

Supplementary material for this article is available at www.BioTechniques.com/article/114352.

Quantitative image analysis plays an important role in contemporary biomedical research. Squassh is a method for automatic detection, segmentation, and quantification of subcellular structures and analysis of their colocalization. Here we present the applications Squassh3C and SquasshAnalyst. Squassh3C extends the functionality of Squassh to three fluorescence channels and live-cell movie analysis. SquasshAnalyst is an interactive web interface for the analysis of Squassh3C object data. It provides segmentation image overview and data exploration, figure generation, object and image filtering, and a statistical significance test in an easy-to-use interface. The overall procedure combines the Squassh3C plug-in for the free biological image processing program ImageJ and a web application working in conjunction with the free statistical environment R, and it is compatible with Linux, MacOS X, or Microsoft Windows. Squassh3C and SquasshAnalyst are available for download at www.psi.ch/lbr/SquasshAnalystEN/SquasshAnalyst.zip.

Many areas of biomedical research rely on cell imaging to obtain information about the phenotypic response of cells to different growth factors and drugs. A common task is the quantification of spatial colocalization between fluorescently labeled species since the proteins under study are usually related to known subcellular markers.

Squassh is a subcellular shape analysis tool based on a segmentation technique that corrects for microscope blur, detector noise, and uneven background (1,2). It enables the generation of large amounts of data on the features of subcellular objects. Once the objects of interest are distinguished from the background, object features such as size, intensity, or shape, as well as colocalization between fluorescence channels, are calculated and then written to comma-separated value (CSV) files.

Squassh comes with basic R functionalities for data analysis, figure generation, and statistical tests. However, the investigation of data generated for a particular problem often requires customized analysis, which then requires learning the R programming language, a prohibitive investment of time for many cellular biology researchers who lack training in software programming. Moreover, analysis of more than two fluorescence channels or of live-cell imaging is even more complex since specific data processing steps are necessary.

Here we present an updated version of Squassh called Squassh3C that works in combination with SquasshAnalyst, a new interactive web program written in the statistical programming environment R. This allows for the analysis of images (single images or movies obtained from fixed- or live-cell imaging) having up to three fluorescence channels that can be obtained, for example, with the co-expression of fluorescently tagged sensors (3) or with immunostaining. R is a free software and programming language (www.r-project.org) providing a wide variety of data processing, statistical, and graphical tools. SquasshAnalyst brings these tools to non-programmers by providing an easy-to-use graphical interface made with the R package Shiny (www.rstudio.com/shiny).

The workflow functions with image data files in any format supported by ImageJ. In contrast to Squassh, Squassh3C can analyze in batch mode different experimental conditions separated into distinct subfolders by selecting the parent folder. Conditions can then be compared and tested for statistical significance in the SquasshAnalyst user interface. It is also possible to compare conditions from separate Squassh3C computational runs.

Squassh3C is implemented as a plugin for the image processing program ImageJ (4,5). It provides a user interface for input selection, segmentation parameters, setting the microscope objective characteristics, and output options (Figure 1A). SquasshAnalyst is a web application with user interface tabs

METHOD SUMMARY

Here we describe the image analysis applications Squassh3C, a method for automatic segmentation and quantification of subcellular structures and analysis of their co-localization in up to three fluorescence channels, and SquasshAnalyst, an interactive web interface for the analysis of Squassh3C segmentation data.

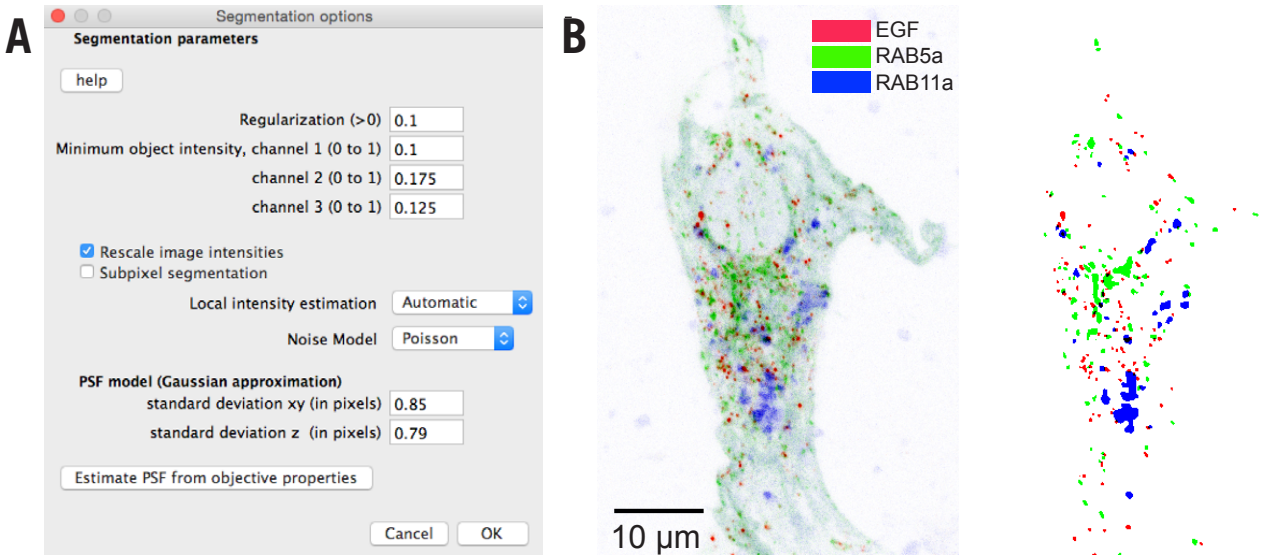


Figure 1. Squassh 3C. (A) Squassh3C interface showing segmentation parameters panel. A higher regularization parameter should be used for noisy images. Minimum object intensities are set separately for each channel. Further description of the parameters is available by clicking on the help button. (B) A COS cell transiently expressing EYFP-RAB5a (green) and mCherry-RAB11a (blue) is shown 30 min after stimulation with EGF-Cy5 (red). A z-axis projection of the fluorescence image is shown to the left of the Squassh segmentation (right). Lookup table values ranging from white (background) to color (maximum signal) produce a printer friendly colormap.

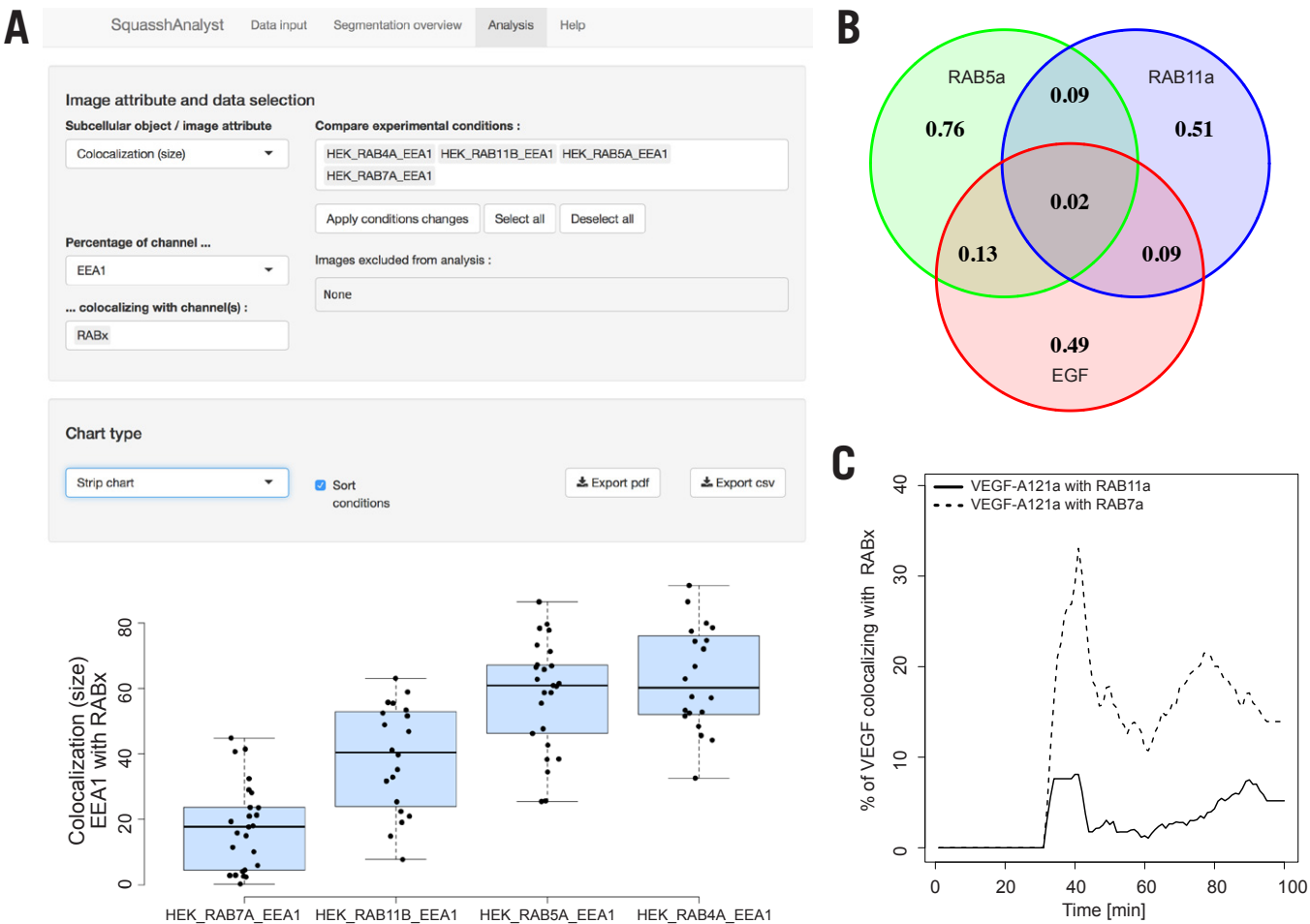


Figure 2. SquasshAnalyst. (A) SquasshAnalyst interface showing analysis tab with settings for a comparison of the percentages of early endosome marker EEA1 colocalizing with RAB7a, RAB11b, RAB5a, and RAB4a. (B) Venn diagram showing the average colocalization of five fluorescence images containing three fluorescence channels: RAB5a (green), RAB11a (blue), and EGF (red). (C) PAE cells expressing VEGFR-2 and NRP-1 receptors were transiently infected with a virus expressing mTFP1-RAB7a and mCherry-RAB11a and then stimulated with Cy5-labeled VEGF165a. The graph displays the change over time of the amount of VEGF colocalizing with RAB7a or RAB11a. We provide all microscopy images needed to reproduce these figures with SquasshAnalyst at tinyurl.com/squasshanalystdata.

for data input selection, an overview of Squassh3C generated segmentation, and data analysis (Figure 2A). Figures generated by SquasshAnalyst can be live-previewed while their graphical, processing, and analysis options are tuned, and can then be exported as PDF or PNG files.

Squassh3C and SquasshAnalyst form an automated workflow for the analysis of up to thousands of fluorescence images. However, for correct analysis, it is important to check that the intermediate steps, such as segmentation of individual images or filtering of outliers, are done properly. SquasshAnalyst therefore generates a contact sheet view containing thumbnails of the segmented images to facilitate inspection of Squassh3C segmentation analysis. Each thumbnail displays the original image and the segmented objects side-by-side. The original image is displayed with all channels merged, automatically enhanced contrast with 0.05% of the pixels saturated, and maximum z-axis projection in the case of 3-D images (Figure 1B). When a cell mask of positively transfected cells is computed by Squassh3C, the outline of the mask is overlaid in white on the image. It is possible from the contact sheet view to exclude selected images from further analysis.

The visual comparison of an image feature (colocalization, object count, cell size, mean object volume, mean object intensity, mean object sphericity) between image sets is achieved with bar plots, boxplots, or strip charts. Figure 2A compares colocalization of the EEA1 marker for early endosomes with four different RAB GTPases. The lesser colocalization of EEA1 with RAB11b and RAB7a compared with RAB4a and RAB5a reflects the maturation of early endosomes to recycling and degradation vesicles, respectively. Colocalization quantified after segmentation as done with Squassh3C was shown to yield better results compared with pixel-based approaches such as Pearson correlation (1,2). An option allows switching between colocalization-based on object count, overall overlap size, or overall signal colocalization.

Fast visualization of all colocalization channel subsets in a given image set is achieved by displaying them as a Venn diagram (Figure 2B).

Quantitative analysis of movies from live-cell imaging is done by separate analysis of the movie frames. Results are then combined to display the progression over time of the colocalization of selected channel subsets or of any mean image feature. A rolling average over time frames with a parameterizable time window can be used to smooth any noise. As an example, we quantified the localization of VEGF165a-Cy5 with mTFP1-RAB7a and mCherry-RAB11a in PAE cells stably expressing VEGFR-2 and NRP-1 receptors (Figure 2C) in order to monitor, respectively, the degradation and recycling of VEGF165a after VEFGR2 stimulation. (See Reference 6 for a description of the results.)

Statistical analysis to compare multiple conditions at a time and for pairwise multiple comparisons are done, respectively, with one-way ANOVA and Tukey-Kramer tests or with the nonparametric Kruskal-Wallis and Dunn tests. The most representative image of a condition is automatically detected and displayed. It is possible to remove objects from the analysis based on intensity or shape criteria thresholds. Removing objects with low intensity from the analysis is useful, for instance, to remove non-specific fluorescence in immunostaining experiments.

Squassh3C and SquasshAnalyst constitute a complete workflow from segmentation to data analysis that is based on the state-of-the-art segmentation method Squassh (1,2). A tutorial and installation instructions for the workflow are available in the SquasshAnalyst manual at www.psi.ch/lbr/SquasshAnalystEN/SquasshAnalyst_manual.pdf. An included toy example allows testing of the entire procedure in less than 15 minutes. Finally, the manual showcases further possible uses of SquasshAnalyst with the image data sets corresponding to Figure 2 containing more than 300 images in total.

Author contributions

P.B., K.B., and A.R. designed the project; A.R. developed the image-processing algorithm and programmed the software with input from P.B., K.B., and M.M.; experimental data were obtained by M.M.; A.R. and P.B. wrote the manuscript with input from K.B. and M.M; figures were prepared by A.R. and P.B.

Acknowledgments

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement no. 290605 (PSI-FELLOW/COFUND).

Competing interests

The authors declare no competing interests.

References

1. **Rizk, A., G. Paul, P. Incardona, M. Bugarski, M. Mansouri, A. Niemann, U. Ziegler, P. Berger, and I.F. Sbalzarini.** 2014. Segmentation and quantification of subcellular structures in fluorescence microscopy images using Squassh. Nat. Protoc. *9*:586-596.
2. **Paul, G., J. Cardinale, and I.F. Sbalzarini.** 2013. Coupling Image Restoration and Segmentation: A Generalized Linear Model/Bregman Perspective. Int. J. Comput. Vis. *104*:69-93.
3. **Kriz, A., K. Schmid, N. Baumgartner, U. Ziegler, I. Berger, K. Ballmer-Hofer, and P. Berger**. 2010. A plasmid-based multigene expression system for mammalian cells. Nat Commun. *1*:120.
4. **Collins, T.J.** 2007. ImageJ for microscopy. Biotechniques. *43*:25-30.
5. **Schneider, C.A., W.S. Rasband, and K.W. Eliceiri.** 2012. NIH Image to ImageJ: 25 years of image analysis. Nat. Methods *9*:671-675.
6. **Ballmer-Hofer, K., A.E. Andersson, L.E. Ratcliffe, and P. Berger.** 2011. Neuropilin-1 promotes VEGFR-2 trafficking through Rab11 vesicles thereby specifying signal output. Blood *118*:816-826.

Received 15 June 2015; accepted 05 August 2015.

Address correspondence to Philipp Berger, Paul Scherrer Institute, Molecular Cell Biology, Villigen, Switzerland. E-mail: Philipp.Berger@psi.ch

CHAPTER 5

CONCLUSION

Multigene delivery systems have gained significant attention in biological research in both academia and industry. Projects in biological research benefit directly from these systems notably in fields such as genome editing, regenerative medicine and biotechnological product development. Therefore, developing novel transduction systems with high delivery efficiency and simplicity of application is of high interest for biologists. Until recently, a multitude of strategies have been developed to genetically modify cells. A couple of years ago, our group developed MutiLabel, which is a plasmid-based multigene expression system for mammalian cells. In contrast to classical co-transfection procedures using several plasmids, which resulted in heterogeneous cell populations, MultiLabel allowed to simultaneously transduce and successfully express a large number of genes into transfected cells. Although the MultiLabel system works well with easy to transfect established cell lines such as HEK293, other, in particular primary cells are difficult to efficiently transfect with the current established procedures. In order to overcome this shortcoming of MultiLabel, I developed a highly efficient baculovirus-based multigene delivery system for eukaryotic cells. We termed this system “MultiPrime” because it is particularly well suited for multigene delivery to primary mammalian cells. MultiPrime relies on a set of customized DNA plasmid modules, called acceptor and donor plasmids that can be combined through in vitro Cre/loxP recombination. A fused acceptor-donor plasmid can then be inserted into a specially engineered baculoviral genome. During my thesis I have prepared a collection of MultiPrime vectors carrying specific regulatory sequences such as distinct promoters, or transcriptional terminators. Antibiotic resistance genes for subsequent cell selection or fluorescent marker genes for cell monitoring were introduced as additional tools to make the procedure generally applicable and easy to use.

To document the potential of MultiPrime, I delivered and simultaneously expressed up to five different expression cassettes in mammalian cells including both primary and stem cells. Our system thus opened the door for introducing a multitude of genes into hard-to-transfect cells. Our baculovirus constructs, especially the VSV-G protein pseudotyped versions, were shown to successfully transduce a wide range of eukaryotic cells both in vivo and in vitro.

I also validated a range of mammalian promoters that can be introduced into the MultiPrime system. This feature allows to fine-tune the level of transgene expression according to user requirment. In addition, we engineered two dual-host-specific promoters for multigene expression in both mammalian and insect cells using a single MultiPrime virus. The use of these promoters eliminates additional steps for re-cloning of the genes of interest from insect cell to mammalian vectors or vice versa. For example, I converted three scFvs to full length human antibodies and expressed them in mammalian and insect cells using a dedicated MultiPrime virus. I also used MultiPrime to convert mouse embryo fibroblasts to neural cells through forced expression of three redifferentiation-promoting transcription factors which are required for somatic cell reprogramming. MultiPrime baculovirus is thus an excellent vehicle for reprogramming of cells represent thus a transient and non-integrating expression system for safe application not only in cells but also in whole organisms. To document this I applied MultiPrime to transduce zebrafish with fluorescent protein markers. Finally, MultiPrime was applied for efficient CRISPR/Cas9-mediated genome editing in primary cells. MultiPrime thus allows not only RNA polymerase II dependent transcription but can also be used for U6-driven expression cassettes. When expressed together with a homology construct such a virus allows for genome editing through CRISPR/Cas9.

In all the applications mentioned above, reliable and reproducible production of functional MultiPrime virus stocks is crucial. I therefore developed a detailed step-by-step protocol for implementation of our technique in a standard laboratory environment.

Taken together, I developed and validated MultiPrime as powerful vehicle for multigene delivery, multi-protein expression and genome editing. I anticipate that numerous methodologies in contemporary research and drug discovery, which rely on efficient multigene delivery, will considerably benefit from MutiPrime.

CHAPTER 6

ACKNOWLEDGEMENTS

I would like to thank everyone who, by their willing support and help, however small, made this thesis possible. First and foremost, I would like to thank my supervisor, Dr. Philipp Berger for his great academic guidance and continuous support of my Ph.D study and related research, for his patience, motivation, and immense knowledge. I could not have imagined having a better advisor for my Ph.D study. Thank you Philipp for all things that you taught me.

I also sincerely appreciate my formal academic advisor, Prof. Dr. Kurt Ballmer-Hofer, for giving me this opportunity to work in his group. He is really a wonderful PhD father and guided me in a correct scientific way. I learned many things especially fluorescent microspore techniques and live cell imaging from you. Also, I will never forget meetings with Philipp and Kurt, where I learned how to think like a professional scientist.

I would like to extend my appreciation to my scientific committee member, Prof. Verdon Taylor for his helpful comments and discussion during my annual committee meeting and also Prof. Martin Spiess who kindly accepted to be a part of my dissertation examination committee.

I would like to thank our collaborators; Prof. Imre Berger, Prof. Taylor, Prof. Neuhauss and their groups, for the ideal collaboration we had.

I would like to thank my friends and colleagues in the Ballmer and Berger groups for being a friend, a motivator, and for the great time I had. I am extremely indebted to Aurélien Rizk as an omnipotent computational specialist. Thank you Aurélien for your constant help to me not only for the great time that we had analyzing images and movies but also for your help before my JLBRs. I also appreciate Kate Thieltges, Ye Xie, Mayanka Asthana, Sandra Markovic, Maria Misti, Conny Walther and Dragana Avramovic for their help during my thesis. I would like to thank Thomas Schleier for making me think about things I never thought I would think about. Additionally, I would like to acknowledge all the master or trainee students who worked with me

during these four years especially Milica, Nagjie, Greseldis, and Alexandra. Also, I would like to thank Ulla and Antonietta for their help in the lab and cell culture affairs.

Words fail me to express my appreciation to my parents, Alireza and Maryam, whose dedication, love and constant encouragement always guided me to pursue my goals. Their prayer for me was what sustained me thus far. I appreciate my brothers, Hamidreza, Majidreza, Mohammad and Mostafa for their continuous encouragement throughout my entire life. Additionally, I would like to thank my wife's family especially my parents-in-law for all their unselfish love, endless support, and unconditional encouragement.

Last but not least, my deepest gratitude goes to my wife, Zahra Ehsaei. I am forever indebted to Zahra who her endless love, priceless, perpetual, indispensable help and support, which made all this possible. Her love never fails and I look forward to the exciting journey ahead.

PERSONAL INFORMATION

Maysam Mansouri

Paul Scherrer Institute, OFLG/110, 5232 Villigen – PSI, Switzerland

+41 56 310 2870 +41 76 667 6324

maysam.mansouri@psi.ch
m.mansouri282@gmail.com
https://www.researchgate.net/profile/Maysam_Mansouri

Sex Male | Date of birth 23/03/1984 | Nationality Iranian

EDUCATION

01/2012 - present — PhD student

University of Basel – Paul Scherrer Institute
Molecular Genetics
Project thesis: Multigene delivery to mammalian cells and its applications in cell biology

Supervisors: Dr. Philipp Berger, Prof. Dr. Kurt Ballmer-Hofer

Projects/Experiments:

- Development of a baculovirus-based multigene delivery system (MultiPrime) for primary cells
- High resolution map of intracellular trafficking routes using Rab GTPases as markers
- Knock-in genome editing in mammalian cells using all-in-one CRISPR/Cas9 vector system
- VEGFR2 trafficking and signalling study through simultaneous expression of Rab GTPases as subcellular markers

11/2009 – 06/2010 — Internship (Military service)

BMSU, Tehran, Iran
Nanobiotechnology
Nasal vaccine against *Pseudomonas aeruginosa*

09/2007 – 09/2010 — M.Sc

IHU, Tehran, Iran
Cell and molecular biology
Project thesis: Production of CfaE, a tip-subunit of CFA/I fimbriae, in *E. coli* and evaluation of its immunogenicity for vaccine development against ETEC

09/2003 – 07/2007 — B.Sc

IHU, Tehran, Iran
Microbiology

SCOLARSHIPS AND AWARDS

01/2014 Best poster prize awarded by Swiss Society for Molecular and Cellular Biosciences, LS2 meeting, Lausanne, Switzerland

01/2011 Member of Iran National Elite Foundation

10/2010 1st rank amongst 15 undergraduate students of cell and molecular biology

07/2007 1st rank amongst 40 undergraduate students of microbiology

RESEARCH INTERESTS

- Genome engineering using the CRISPR/Cas9 system
- Cellular signalling and trafficking studies
- Assembly and delivery of multigene constructs
- Development of biosensors for mammalian cells

PUBLICATIONS

1. **Mansouri M**, Bellon-Echeverria I, Rizk A, Ehsaei Z, Cianciolo Cosentino C, Silva CS, Xie Y, Boyce FM, Davis MW, Neuhauss SC, Taylor V, Ballmer-Hofer K, Berger I, Berger P. "Highly efficient baculovirus-mediated multigene delivery in primary cells". **Nature Communications.** 2016. PMID: 271443231
2. Rizk A, **Mansouri M**, Ballmer-Hofer K, Berger P. 'Subcellular object quantification with Squassh3C and SquasshAnalyst". **Biotechniques**. 2015. PMID:26554508.
3. **Mansouri M**, Berger P. "Strategies for multigene expression in eukaryotic cells". **Plasmid**. 2014. PMID:25034976.
4. Rizk A, Paul G, Incardona P, Bugarski M, **Mansouri M**, Niemann A, Ziegler U, Berger P, Sbalzarini IF . "Segmentation and quantification of subcellular structures in fluorescence microscopy images using Squassh". **Nature Protocol**. 2013. PMID:24525752
5. **Mansouri M**, Mousavy SJ, Ehsaei Z, Nazarian S, Zali MR, Moazzeni SM."The codon-optimization of cfaE gene and evaluating its high expression capacity and conserved immunogenicity in Escherichia coli". **Biologicals**. 2013. PMID:23453276.
6. Taranejoo S, Janmaleki M, Rafienia M, Kamali M, **Mansouri M**. "Chitosan micro particles loaded with exotoxin A subunit antigen for intranasal vaccination against Pseudomonas aeruginosa: An in vitro study". **Carbohydrate Polymers**. 2011.

PRESENTATIONS

1. **Mansouri M**, et al. "Baculovirus-based multigene delivery to stem cells". ISSCR, Jun. 2016, San Francisco, USA.
2. **Mansouri M**, et al. "MultiPrime: a Baculovirus-based multigene expression system for mammalian cells". Genome engineering: The CRISPR/Cas9 revolution, Sep. 2015, Cold Spring Harbor Laboratory, New York, USA.
3. **Mansouri M**, et al. "MultiPrime: A Baculovirus-based multigene expression system for mammalian cells". Enable synthetic biology for mammalian cells, Jun 2015, Heidelberg, Germany.
4. **Mansouri M**, et al. "Bac-MultiLabel: A Baculovirus-based multigene expression system for mammalian cells". LS2 meeting, January 2015, Zurich, Switzerland.
5. **Mansouri M,** et al**. "**Baculovirus-based expression of multiple biosensors for quantitative single cell analysis". SystemX conference, October 2014, Lausanne, Switzerland.
6. **Mansouri M,** et al**.** "A baculovirus-based multigene expression system to study trafficking and signaling of receptor tyrosine kinases". Kinome targeting III conference, September 2014, Basel, Switzerland.
7. **Mansouri M**, et al. "Baculovirus-based multigene expression for studying membrane receptor trafficking ". D-BIOL ETH conference, Jun 2014, Davos, Switzerland.
8. **Mansouri M,** et al**.** "Baculovirus-based Multigene Expression in Mammalian Cells". USGEB conference, February 2014, Lausanne, Switzerland.

SKILLS

Molecular biology

Cloning, expression and purification in *E. coli,* insect cell and mammalian cells;

- Molecular Cloning: Conventional cloning methods (cut and paste),
  Overlapping PCR based cloning,
  Sticky end PCR- based cloning,
  Gibson assembly,
  In vitro cre/loxP recombination
- Genome editing: assembly, transfection/infection, analysis

Cell biology

- Working with insect cells (sf21, sf9, High five), cell lines (HEK293, HeLa, COS-7, PAE,…) and primary cells (HUVEC, REF, cortical Neuron)
- Transient transfection and generation of stable cell lines
- Generation, titration and infection of baculovirus, Lentivirus
- Trans-differentiation of MEF to induced-Neuron

Microscopy	Wield field microscopy, fluorescence microscopy, confocal microscopy, live cell imaging
Immunotechniques	ELISA, Western blotting, Immunostaining, mice immunization
Angiogenesis	Tube formation assay, migration assay, activity assay
Software	ImageJ, Ape, Office, Photoshop and Illustrator

REFERENCES

Dr. Philipp Berger
Group leader
Paul Scherrer Institute
+41 56 310 4728
philipp.berger@psi.ch

Prof. Kurt Ballmer-Hofer
Group Leader
Prof. at Univ. Basel
+41 56 310 4165
kurt.ballmer@unibas.ch

Prof. Gebhard Schertler
Head of department
Prof. at ETH (D-Biol)
+41 56 310 - 4265
gebhard.schertler@psi.ch