Universitat de les Illes Balears

DOCTORAL THESIS

ESTABLISHING A SAFE HARBOR SITE FOR THE INTRODUCTION OF GENETIC INFORMATION IN THE HUMAN CELLS BY THE RECOMBINATION SYSTEM attP/attB

Esther Palomino Lago

2018

Universitat de les Illes Balears

DOCTORAL THESIS

2018

Doctoral Programme of Translational Research in Public Health and High Prevalence Diseases (RD99)

ESTABLISHING A SAFE HARBOR SITE FOR THE INTRODUCTION OF GENETIC INFORMATION IN THE HUMAN CELLS BY THE RECOMBINATION SYSTEM attP/attB

Esther Palomino Lago

Thesis Supervisor: Dr. Daniel Bachiller

Thesis Co-supervisor: Dr. Aarne Fleischer

Thesis tutor: Dra. Julia García

Doctor by University of the Balearic Islands

This thesis has been financially supported by the Balearic Government FPI-CAIB 2013 grant co-financed by the European Union, the Spanish Ministry for Science and Innovation (grant numbers PLE2009-0091, RTC-2014-2207-1 and IPT-2011-1402-900000), and the ISCIII PI14/01073 grant, co-financed by ERDF/ESF.

The thesis titled: “Establishing a Safe Harbor Site for the Introduction of Genetic Information in the Human Cells by the Recombination System attP/attB”, has been written by

Esther Palomino Lago

PhD student

Universitat de les Illes Balears

Dr. Daniel Bachiller of Spanish National Research Council (CSIC)

I DECLARE:

That the thesis titled: “Establishing a Safe Harbor Site for the Introduction of Genetic Information in the Human Cells by the Recombination System attP/attB”, presented by Ms. Esther Palomino to obtain a doctoral degree, has been completed under my supervision and meets the requirements to opt for a Doctorate.

For all intents and purposes, I hereby sign this document.

Signature

Palma de Mallorca, October, 24th 2018

Universitat de les Illes Balears

Dr. Aarne Fleischer from Karuna GCT:

I DECLARE:

That the thesis titled: “Establishing a Safe Harbor Site for the Introduction of Genetic Information in the Human Cells by the Recombination System attP/attB”, presented by Ms. Esther Palomino to obtain a doctoral degree, has been completed under my supervision and meets the requirements to opt for a Doctorate.

For all intents and purposes, I hereby sign this document.

Signature

Palma de Mallorca, October, 24th 2018

Dedicated to my beloved parents

ACKNOWLEDGEMENTS

I would like to express my sincerest gratitude to my director Dr. Daniel Bachiller, for giving me the opportunity to work at the Department of Advanced Therapies in Islas Baleares. I wish to thank him for the immeasurable amount of support and guidance he has provided throughout this study. His extensive knowledge, vision, and creative thinking have been the source of inspiration for me throughout this project.

I am greatly indebted to my co-director Dr. Aarne Fleischer. I am honored to have a supportive and professional teacher like him. I am always thankful to his enthusiastic leadership, encouragement and positive thinking. I will never forget his support and care during the hard times. It has always been a pleasure to discuss our results and findings with him.

It is with great pleasure that I thank my co-worker Dr. Jose M. Martín. He was always ready to answer my questions and provide constructive guidance and criticisms. I am thankful for his support and encouragement. I would like also to thank Victor Gálvez for his support when I arrived to the laboratory.

I am also grateful to past and present members of our group: Patricia Tortosa, María Tortosa, Dr. Sara Vallejo, Dr. Almudena Sánchez, and many others that I do not have the space to mention here, thanks for being such great lab mates. It has been an honor to work with you and I will cherish our many discussions. Thanks guys for all the enjoyable moments in Mallorca, inside and outside the lab! Those breaks were quite important for finishing this work.

It is a pleasure to thank all of Dr. Calos colleagues at Stanford University for the experience of working at one of the most prestigious Universities. I would also like to thank Dr. Mital Bhakta, whose kind supervision turned into a friendship.

I am grateful for all my PhD colleagues and friends. I would like to express my heartfelt thanks to each and every one of you for supporting me in your own personal way. I will forever value the moments we shared. Your daily laughter and smiles have made my time as a PhD student extremely enjoyable and absolutely unforgettable. Thank you all for your friendship!

I would like to thank Amparo Díaz and her family, who always wished the best for me. Thanks for all the great conversations and shared moments during both good and bad times.

Most of all, I owe my deepest gratitude to my family from Cadiz and Malaga to whom I dedicate this thesis. My dearest thanks to my parents and sister for their everlasting support and for providing me with the necessary education that allowed me to pursue a scientific career. Thanks dad and mom, you are the best parents that a daughter can have. Without you this thesis would not have been possible.

PUBLICATIONS

Published articles from this thesis:

Fleischer, A., Lorenzo, I. M., Palomino, E., Aasen, T., Gómez, F., Servera, M., (...) & Bachiller, D. (2018). *Generation of two induced pluripotent stem cell (iPSC) lines from p. F508del Cystic Fibrosis patients*. *Stem Cell Research*.Vol 29, pp1-5.

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AA: Amino Acids

AAV: Adeno virus-associated

AAVS1: the adeno-associated virus site 1

att: attachment

bFGF: Basic Fibroblast Growth Factor

bp: base pair

CCR5: Chemokine (CC motif) receptor 5

CRISPR/Cas9: Clustered Regularly Interspaced Short Palindromic Repeats

DAXX: Domain-associated

ddH_2O: Double Distilled H_2O

DICE: Dual Integrase Cassette Exchange

DSBs: Double-Strand Breaks

EEF1A1: eukaryotic translation elongation factor 1 alpha 1

ESC: Embryonic Stem Cell

FACS: Fluorescence-activated cell sorting

FBS: Fetal Bovine Serum

GSHs: Genomic Safe Harbors

HaHyPBase: Hyperactive PiggyBac transposase

hiPS cells: Human Induced Pluripotent Stem cells

HIV: Human Immunodeficiency Virus

HR: homologous double repair

Int: Integrase

iPS: Induced Pluripotent Stem

ITRs: Inverted Terminal Repeats

Kb: kilo base

KOSR: knockout Serum Replacement

LB: Luria-Bertani

LINEs: Long Interspersed Elements

LTR: Long Terminal Repeat

MLV: Murine Leukemia Virus

NEAA: Non-Essential Amino Acids

NHEJ: Non-Homologous End Joined

O.N: Overnight

OPI: Overproduction Inhibition Phenomenon

PAM: Protospacer Adjacent Motif

PB: PiggyBac

PBS: Phosphate-buffered Saline

PSCs: Pluripotent Stem Cells

RDF: Recombination Directionality Factor

RT: Room Temperature

RTs: Recombination Target sites

RVDs: Repeat Variable Diresidues

SAPTA: Scoring Algorithm for Predicting TALEN Activity

SB: Sleeping Beauty

SE: Selection Element

SINEs: Short Interspersed Elements

SM: Selection Marker

SSRs: Site-Specific Recombinases

TALE: Transcription Activator-like Effector

TALEN: Transcription Activator-like Effector Nucleases

Tm: melting Temperature

ZD: Zin Ribbon Domain

ZF: Zinc Finger

ZFN: Zinc Finger Nucleases

ZFP: Zinc Finger Protein

ABSTRACT

The goal of this project is to generate a safe harbor site in the genome of human cells, where information could be specifically inserted without the action of nucleases. The mechanism that we present here is based on two different site-specific recombinases phiC31 and Bxb1, as well as on the piggyBack and Sleeping Beauty transposons. Recombinase phiC31 is an integrase used by phages to establish the lysogenic life cycle. During integration, phiC31 drives recombination between the attP and the attB attachment sites on the phage and host genome, respectively. In naturally occurring phage infestations, the end result is an integrated phage genome flanked by new attL and attR sites, generated by recombination of the original, attP and attB sites. In our system, the phage genome is substituted by the acceptor sites that will constitute the core of the safe harbor locus. Under inducing conditions, the phage genome is excised via integrase-mediated recombination between attL and attR regenerating the attP and attB attachment sites. This action is directed by phiC31 in the presence of an accessory protein (the recombination directionality factor, RDF). The alternative use of phiC31, alone or together with RDF, allows for the indefinite repetition of the cycle and the subsequent incorporation into the targeted locus of as many attachment sites as needed. The whole mechanism is made possible by the coordinated and alternative use of piggyBac and Sleeping Beauty transposons that, at each step, remove residual DNA fragments (plasmid sequences, selection elements, etc.). Once the final configuration of the safe harbor locus is reached, the Bxb1 recombinase is used to upload the desired genetic information: markers, therapeutic genes, inducible system or even complete regulatory routes.

RESUMEN

El objetivo principal de este proyecto es generar una plataforma de carga segura dentro del genoma de células humanas, donde la información pueda ser insertada de forma específica sin la necesidad de usar nucleasas. De este modo se reduciría la posibilidad de integraciones azarosas y se aumentaría la eficiencia de integración. Para este fin, el mecanismo propuesto en la presente Tesis Doctoral combina el uso de dos recombinasas diferentes, phiC31 y Bxb1, así como de los sistemas de transposición piggyBac y Sleeping Beauty. La recombinasa phiC31 es una integrasa encargada de la incorporación del ADN de fagos en el hospedador durante el ciclo lisogénico. La integración se realiza mediante recombinación entre dos sitios de unión específicos, attP presente en el fago y attB localizado en el genoma del hospedador. Durante el proceso de infección el genoma del fago queda integrado en el genoma del hospedador y flanqueado por dos nuevos sitios de reconocimiento específico, attL y attR, generados a partir de la recombinación de las secuencias originales, attP y attB. En nuestro sistema, el genoma del fago es sustituido por los sitios de carga que constituirán el núcleo de la plataforma de integración. Bajo condiciones de inducción el genoma del fago puede eliminarse del sitio de integración mediante recombinación entre los sitios attL y attR, regenerándose en el proceso los sitios attB y attP originales. Esta reacción inversa la cataliza phiC31 en presencia de una proteína accesoria denominada Factor de Direccionalidad de la Recombinación (RDF, por sus siglas en inglés). El uso alternativo de phiC31 sola o en combinación con RDF, hace posible la repetición indefinida del ciclo de carga y la consiguiente incorporación al locus genómico elegido de tantos sitios de recombinación como se consideren necesarios. El ensamblaje completo de la plataforma se consigue mediante el uso coordinado y alternativo de los transposones piggyBac y Sleeping Beauty que permite en cada paso la eliminación de los fragmentos de ADN no deseados (secuencias del plásmido, elementos de selección, etc.). Una vez constituida la plataforma de carga, la recombinasa Bxb1 es la responsable de mediar la carga de la información genética deseada: marcadores, genes terapéuticos, sistema inducible o incluso rutas regulatorias completas.

RESUM

L'objectiu principal d'aquest projecte és generar una plataforma de càrrega segura dins el genoma, en cèl·lules humanes, on la informació pugui ser inserida de manera específica i sense necessitat d'usar nucleases. D'aquesta manera es reduiria la possibilitat d'integracions degudes a l'atzar i s'augmentaria l'eficiència d'integració. Amb aquest finalitat, el mecanisme proposat en la present Tesi doctoral combina l'ús de dues recombinases diferents, phiC31 i Bxb1, així com dels transposons piggyBac i Sleeping Beauty. La recombinasa phiC31 és una integrasa encarregada de la incorporació de l'ADN de fags a l'hoste durant el cicle lisogènic. Aquesta integració es realitza mitjançant la recombinació entre dos llocs d'unió específics, attP present en el fag i attB localitzat en el genoma de l'hoste. Durant el procés d'infecció el genoma del fag queda integrat en el genoma de l'hoste i flanquejat per dos llocs de reconeixement específic nous, attL i attR, generats a partir de la recombinació de les seqüències originals, attP i attB. En el nostre sistema, el genoma del fag és substituït pels llocs de càrrega que constituiran el nucli de la plataforma d'integració. Sota condicions d'inducció el genoma del fag pot eliminar del lloc d'integració mitjançant recombinació entre els llocs attL i attr, regenerant-se en el procés els llocs attB i attP originals. Aquesta reacció inversa la catalitza phiC31 en presència d'una proteïna accessòria anomenada Factor de Direccionalitat de la Recombinació (RDF, per les sigles en anglès). L'ús alternatiu de phiC31 sola o en combinació amb RDF, fa possible la repetició indefinida del cicle de càrrega i la consegüent incorporació al locus genòmic triat de tants llocs de recombinació com es considerin necessaris. L'acoblament complet de la plataforma s'aconsegueix mitjançant l'ús coordinat i alternatiu dels transposons piggyBac i Sleeping Beauty que permet a cada pas l'eliminació dels fragments d'ADN no desitjats (seqüències del plàsmid, elements de selecció, etc.). Un cop constituïda la plataforma de càrrega, la recombinasa Bxb1 és emprada per a la càrrega d'informació genètica desitjada: marcadors, gens terapèutics, sistemes induïble o inclús rutes reguladores completes.

CONTENTS

1. INTRODUCTION

Genome editing is a type of genetic engineering in which DNA is modified in the genome of single cells or organisms. Recent advances in genome editing technologies have improved the ability to make specific and easy changes in the genomes of cells and, therefore, to create new applications into all areas of biotechnology, including biopharmaceutical production (1), environment (2), studies of genome structure, regulation, and function (3; 4; 5), and clinical applications (6; 7).

An ideal gene editing tool should present easy production, efficiency, targeting at multiple sites, high frequency of desired sequence changes in the target cell population and low toxicity. In addition, genome modifications should be heritable, and amenable to simple reading and testing the edition (8).

In eukaryotic cells, the most frequently used genome editing technologies are based on programmable nucleases. These enzymes enable precise genome editing by introducing DNA Double-Strand Breaks (DSBs) at specific genomic locations. The most important programmable nucleases are Zinc Finger Nuclease (ZFN) (9; 10), Transcription Activator-Like Effector Nucleases (TALENs) (11; 12; 13) and Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR/Cas9) (14; 15; 16; 17; 18; 19). Creating double-strand breaks may have undesired side effects, so an important limitation for the use of programmable nucleases is their potential off-target activity (20; 21; 22; 23), although in last few years the new generation of nucleases is showing low off-target cleavage frequency and more specificity (24; 25; 23; 26; 27; 28).

Other molecular tools that are able to perform specific modifications in the genome are Site-Specific Recombinases (SSRs) and Transposons. SSRs have been used by the research community to induce reproducible site-specific genomic integration with high efficiency (29). Some site-specific recombination methods relay on recombination of two non-identical DNA sequences called, attachment sites (att), catalyzed by a phage-encoded integrase protein (Int), that generates new, hybrid attachment sites as

products of the reaction (30; 31). In order to use this system in genome editing, an attachment site must already exist in the genome and a complementary attachment site in the DNA to be integrated into that locus (32; 33; 34).

Transposable elements are DNA sequences that move from one location on the genome to another. Transposon DNA has Inverted Terminal Repeats (ITRs) that are recognized by specific transposases and moved by cut and paste mechanism in which the transposon is excised from one location and reintegrated elsewhere (35; 36; 37; 38). Neither SSRs or transposases require DNA synthesis or degradation, nor any cofactors such as ATP.

Apart from the above-mentioned molecular tools, other critical parameters such as cell type, genome loci and delivery methods that profoundly affect the outcome of gene editing processes, have to be considered when dealing with genome editing.

Among the different cell lines currently used in genome engineering, induced Pluripotent Stem (iPS) cells hold the biggest promise for their use in disease modelling (39; 40), therapeutic screens (41; 42) or clinical applications (43; 44). The iPS cells derived from somatic cells and transformed to an embryonic-like state by the addition of reprogramming factors (45) show similar molecular and functional features as Embryonic Stem Cells (ESCs) (46). Pluripotent stem cells (iPS cells and ESCs) are capable of unlimited self–renewal and reproduction of all adult cell types in the course of their differentiation (47; 48). ESCs research is ethically and politically controversial because it involves the destruction of human embryos (49), but iPS cells are produced from somatic cells which are abundant and are not associated to ethical restrictions (50). In addition, iPSCs permit autologous transplantation in regenerative therapies thus eliminating the need for immunosuppressive treatment (51).

Stable and safe chromosomal integration of genetic payloads can help achieve long-term expression of transgenes (52). Genome Safe Harbors (GSHs) are secure locations in the genome where the newly integrated genetic material keeps their intended activity without adversely altering cellular functions (53; 54). Their location in selected

genes assumes that certain non-essential genes can be disrupted without pathological consequences (55). The most commonly GSHs used for targeted transgene addition in the human genome are the adeno-associated virus site 1 (AAVS1) (56), the chemokine (CC motif) receptor 5 gene (CCR5) (57), the human orthologue of the mouse ROSA26 locus (58) and the H11 locus (59; 60).

Exogenous genetic material or genome editing tools can be delivered into cells as DNA (61; 62), RNA (63) or proteins (9; 11; 16). These biomolecules must first avoid extracellular barriers and once inside the cell, the DNA, RNA or protein must escape the endosomes and localize to the nucleus to edit the target gene (8). There are different vehicles used to ferry theses biomolecules in cells, tissues, and whole organs. These delivery methods include viral vectors, chemical, or physical procedure. (64).

1.1. Genome editing methods

As mentioned above, genome editing can be achieved using engineered nucleases such as ZFN, TALENs or CRISPR/Cas9 (26; 27; 28), site-specific recombinases (65; 66; 67) and transposons (68; 69). This section is devoted to the description of the process, as well as to a brief overview of the molecular and functional characteristics of the genome editing tools that make it possible.

1.1.1. Double-strand DNA break repair

In living cells and tissues, DNA can be damaged both by man-made mutagenic substances and natural agents and processes. Natural agents and processes include exogenous agents, such as sunlight and dietary mutagens, as well as endogenous agents, such as reactive oxygen species (70; 71). The potential damage includes: deletions, deamination, alkylation, dimerization, oxidation, single-strand breaks (72; 73) and DSB (73). Cells have to be able to fix their damaged DNA, otherwise,

unrepaired alterations can result in the loss of genetic information, chromosomal aberrations, reduction in cell viability and, finally, in cell death.

Cells have two primary DBS repair pathways: (i) Non-Homologous End Joining (NHEJ) and (ii) Homologous Recombination (HR) (74; 75; 76; 77) (Figure 1).

(i) NHEJ modifies the broken DNA ends and ligates them together with no regard for generating deletions or insertions.

(ii) HR uses an undamaged DNA template to repair the break, leading to the reconstitution of the original sequence.

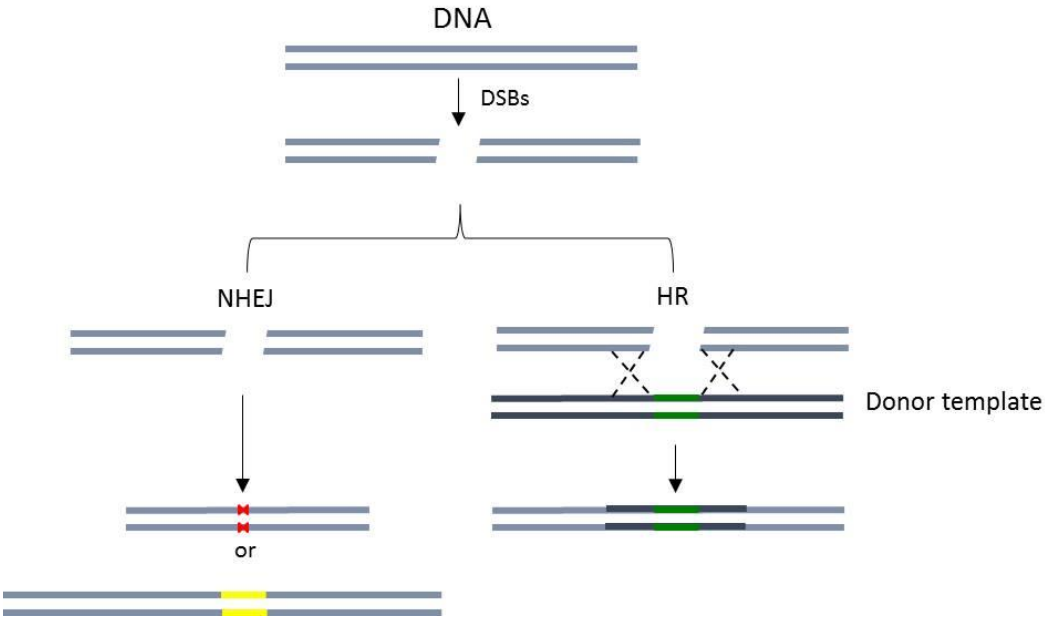

Figure 1| Cell mechanism involved in double strand break repair. (left) NHEJ generates gene disruption by deletion (red mark) or addition (yellow mark) of nucleotides. (right) HR generates gene correction (green mark) by the incorporation of a DNA fragment from a donor template.

Cellular repair mechanisms for DSBs are at the base of all techniques developed for the targeted introduction of genetic information into specific genomic locations (78; 79).

In the early years, the principal limitations for the application of HR to the modification

of genomes in the laboratory were its very low efficiency (80; 81) and the high rate of random (non-targeted) integration of vector DNA (82; 80). The integration frequency of the targeting vector correlates with vector configuration, the location of the target in the genome, the length of the homology arms (83; 84), the amount of repetitive DNA sequences present in the arms (83; 85), and the method of transfection (81). Efficient recombination requires at least 500 bp of uninterrupted homology (86; 87; 83), with recombination frequency improving as the length of the homology arms increases (86; 87; 83; 88; 89; 84). But, in spite of substantial increases in length, the frequency of successful gene targeting was usually in the range of 10^5 to 10^6 prior to selection (84).

Spontaneous recombination was inefficient, largely because it relayed on the spontaneous occurrence of DSBs in the region being targeted (82). Fortunately, the appearance of nucleases that perform double strand DNA breaks (90; 91; 92) at specific sites, like ZFNs, TALENs and CRISPR/Cas9, has significantly improved the frequency of homologous recombination, bringing it up to levels compatible even with clinical applications (93; 94; 95; 96; 97; 89).

1.1.2. Site specific nucleases

Nucleases are enzymes capable of recognizing and cleaving DNA sequences. Currently, three site-specific endonucleases are commonly used in the lab; ZFNs, TALENs and CRISPR/Cas9. These three types of nucleases share the same mechanism of action: they cleave chromosomal DNA in a site-specific manner, which triggers the endogenous DNA repair systems that results in targeted genome modification (98). ZFNs and TALENs are chimeric nucleases composed of programmable, sequence-specific DNA-binding modules linked to a nonspecific DNA cleavage domain, FokI. On the other hand, CRISPR/Cas9 utilizes a guide RNA to recognize the target (99).

Zinc Finger Nucleases

These programmable nucleases are derived from eukaryotic Cys2-His2 zinc finger proteins (ZFP) covalently linked to the nuclease domain derived from the type IIS restriction enzyme FokI (9). Two FokI molecules are involved in double-stranded DNA cleavage and dimerization occurs at the interface of two catalytic domains of FokI (100). Each zinc-finger (ZF) recognizes a 3 bp DNA sequence (101), and an array of 3-6 zinc-fingers are used to generate a single ZFN subunit that binds to DNA sequences of 4-30 bp in the genome, in plant and mammalian cells (102; 99). To cleave DNA sequences, two ZFN monomers must bind on opposite DNA strands in the appropriate orientation and at the correct distance, separated by spacers of 5-7 bp, from each other (103) (Figure 2).

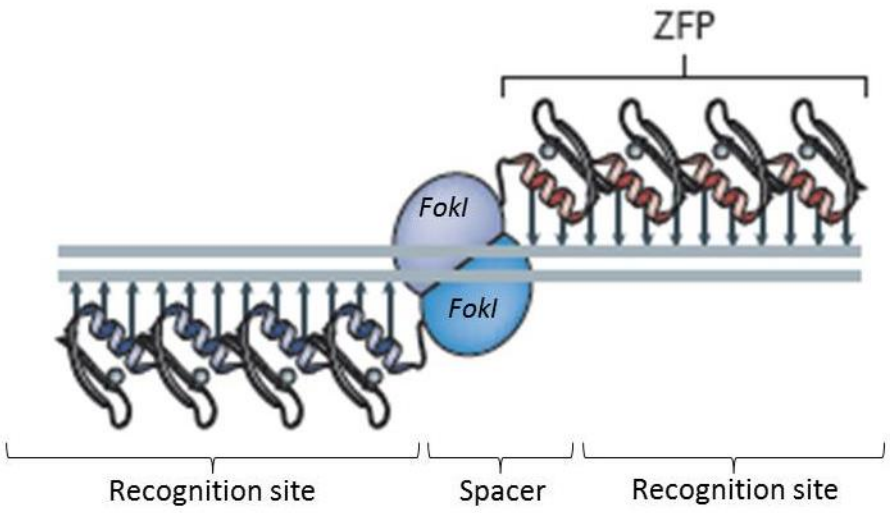

Figure 2| Diagram of ZFNs. Each ZFN consists of two functional domains: (i) Zinc Finger Protein (ZFP: blue and red) comprising, in this example, a chain of 4 zinc finger modules. (ii) DNA-cleavage domain of the FokI nuclease.

The production of new ZFNs requires a complex and labour-intensive engineering of the zinc-finger domain (104). ZF design remains difficult due to position effects and a lack of straightforward ZFP design principles. A number of amino acid sequences in a given finger can specify a given triplet, but the activity of any given zinc finger is

strongly dependent on its position in the ZFP and the nature of the neighbouring zinc fingers (105; 106). In addition, the use of ZFNs can lead to undesired cleavage at different genome locations, so it is essential to test in advance their potential "off-target" activity with bioinformatics tools (107). In general, off-target cleavage reduces efficiency of on-target modifications (108). Increasing the concentration of ZFNs and donor DNA could improve integration efficiency, but it increases the toxicity as well. ZFN toxicity can be lowered by selecting highly specific ZFPs, increasing the number of ZFPs within ZFNs (109) or by redesigning the FokI domain to avoid homodimerization (110; 111; 112).

Transcription Activator-Like Effector Nucleases

Transcription Activator-Like Effector (TALEs) are derived from bacterial plant pathogens from the genus *Xanthomas* (11). TALEs contain N-terminal signals for bacterial type III secretion, variable numbers of tandem repeats that specify the target nucleotide sequence, nuclear localization signals, and a C-terminal region that is required for transcriptional activation (113). Within TALE structure, the central repeat DNA-binding domain consists of tandem arrays 33-35 amino acids (AAs) repeats. Each repeat is largely identical except for hypervariable AAs on positions 12 and 13, referred to as the Repeat Variable Diresidues (RVDs) (13; 114; 115) (Figure 3). TALE target specificity is determined by its sequence of RVDs, where each RVD binds to one nucleotide of the target sequence (11; 115). Among the most common RVDs in TALE are NN (Asn-Asn), HD (His-Asp), NG (Asn-Gly) and NI (Asn-Ile), which specify the nucleotides G, C, T and A, respectively (11; 13). The HD, NG and NI RVDs can also show a limited affinity for other bases (116). The most common RVD that specifies for G is NN, which also interacts with A. The NK RVD is rarely found in nature but has been show to interact with G (13). TALE containing NK are more specific for G-containing targets and less active than their NN-containing counterparts (117). In order to design TALENs with alternative specificity and less off-target, new combinations of the two key amino acids (at position 12 and 13) (non-RVD sites) have been studied (118).

TALENs are composed of a modified TALE comprising of four elements - a DNA-binding domain with a variable number of repeats, a N-terminal domain lacking the first 152 residues of the original TALE, but equipped with a new nuclear localization signal, a C-terminal domain retaining just 28 or 63 of the original residues, and the wild type nuclease domain of the FokI restriction endonuclease. (13).

DNA cleavage by FokI proceeds via an enzyme dimerization step (100). Therefore, as in the case of ZFNs, TALENs can be further engineered to be active only as heterodimers using heterodimeric FokI variants (119). In this configuration, two distinct TALEN monomers are each designed to bind one target half-site resulting in cleavage within the DNA spacer sequence between the two half-sites (13; 120).

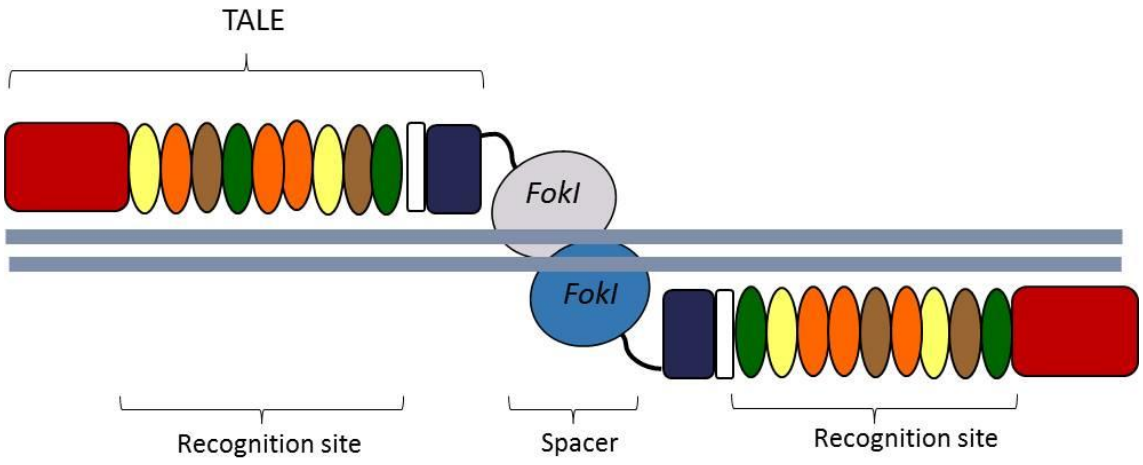

Figure 3| Diagram of TALENs. Each TALEN consists of two functional domains: (i) TALE consists of N-terminal domain (red), tandem arrays amino acids repeats (yellow, orange, brown and green), NLS (white) and C-terminal domain (dark blue). (ii) DNA-cleavage domain of the FokI nuclease.

Since their discovery, TALENs have been utilized to efficiently introduce targeted genetic modifications in a number of organisms such as *Drosophila melanogaster* (121; 122; 123), zebrafish (124; 125; 126), rats and mice (127; 128; 129; 130), non-human

primates (131), as well as in different cell types, such as T cells (132; 133), K562 cells (134), ESCs, and iPS cells (135; 136; 137; 120).

Although molecular cloning is required during their synthesis, TALENs can be rapidly designed and assembled with potentially high activity and specificity. TALENs are easier, cheaper and faster to generate than ZFNs (114; 138). Like with ZFNs there are bioinformatics programs that predict possible off-target localizations in the genome (104). E-TALEN (http://www.e-talen.org) was one of the first ones (139), but more recently, a new and improved web tool, SAPTA (Scoring Algorithm for Predicting TALEN Activity) (http://baolab.bme.gatech.edu/Research/BioinformaticTools/TAL_targeter.html) was published. SAPTA uses improved guidelines for TALEN design based on rules derived from experimentally testing 205 individual TALEN monomers, has been made available (140). Bioinformatics tools are also important to facilitate the screening of possible off-target cleavage sites (104). In addition to these algorithms, other lines of research have focused on TALENs structural modifications to improve their efficiency. For example, the generation of mutant TALENs in the C-terminal domain (consisting of K788Q, R792Q, and R801Q) produced equal on-target cleavage but a 10-fold lower average off-target activity in human cells (141). The introduction of periodic mutations to residues at positions 4 and 32 (non-RVD sites) in each repeat (VT-TALEN) exhibited increased genome editing efficacy compared with the same TALEN without such mutations (CT-TALE) (142; 143).

Clustered Regularly Interspaced Short Palindromic Repeats-Cas9

RNA-guided nucleases are based on the naturally occurring Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)-Cas9 system, which provides, in bacteria and archaea, an adaptive immunity against viruses and plasmids by using CRISPR RNAs (crRNA) to guide the silencing of foreign nucleic acids (16; 15; 18; 14; 144; 145). The complex CRISPR/Cas9 is an endonuclease that uses a single guide RNA

(sgRNA) sequence within an RNA duplex, tracrRNA:crRNA, to pair with DNA target sequences, thus introducing a site-specific DSBs in it (19; 145; 146). The Cas9/sgRNA complex binds and cleaves double-stranded DNA containing a sequence match to the first 20 nucleotides of the sgRNA, if the target sequence is immediately adjacent to a protospacer adjacent motif (PAM: 5'-NGG-3') (147) (Figure 4).

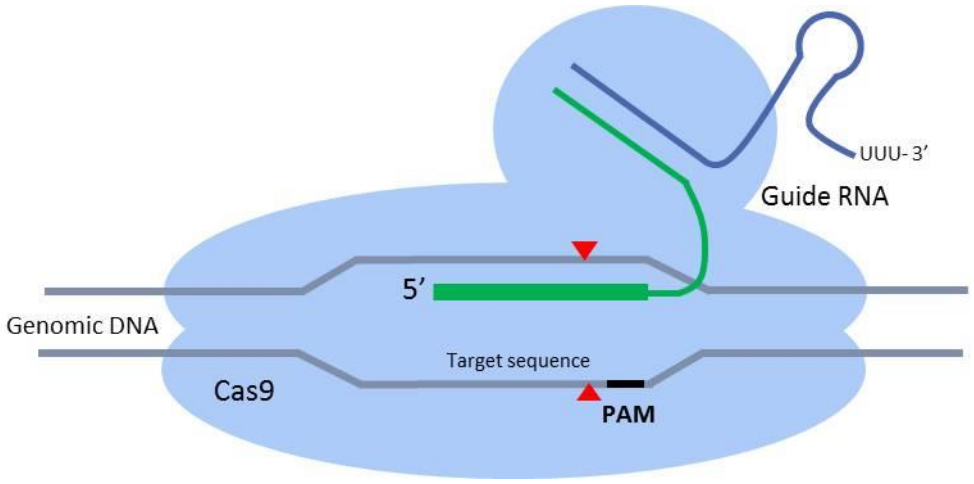

Figure 4| Diagram of CRISPR/Cas9. Cas9 nuclease (light blue cloud) and sgRNA (guide RNA) consistent of tracrRNA (dark blue) and crRNA (green) complex. Red triangle: double-strand break region, 5' to the protospacer-adjacent motif (PAM) (black).

The CRISPR/Cas system has been applied to modify single and/or multiple genes by either NHEJ or HR mediated repair in model organisms, like zebrafish (148; 149), *Drosophila* (150; 151), rats (152; 153), mice (152; 154; 155), non-human primates (156; 157), and is functional in human cells (18; 158; 137). CRISPR technology offers the possibility to generate large deletions in mammalian genomes (18; 137; 158) or achieve multiplex gene editing by simultaneous introduction of multiple sgRNAs (159).

One of the disadvantages of CRISPR/Cas9 is that this system tolerates single and double mismatches on certain positions along the sgRNA-DNA complementary region, leading to undesired mutagenesis in human cells (HEK293, U2OS and K562 cells) (160).

There are ongoing efforts to increase the efficiency of CRISPR/Cas9 systems and to decrease their off-target effects (161; 162; 163; 164). The off-target frequency can be reduced by changing the method used to deliver the Cas9 and gRNA, through transient expression of Cas9 and gRNA by mRNA delivery (165), or using Cas9 protein instead of Cas9 expression from a plasmid (166). Another approach consist of truncations (167) and extensions (168) at the 5' ends of gRNAs, generation of higher-fidelity version of Cas9 protein (25; 169) or incorporation of chemical modifications in gRNAs at specific sites in their DNA recognition sequence (162).

Large genome size and the large number of potential nuclease cleavage sites have made it difficult to determine potential off-target sites, especially as genomic context can greatly influence the cleavage of identical sites at different loci (161). In order to identity potential off-target sites, there have been developments in scripts that systematically scan genomes and offer web-based bioinformatics tools (170; 171; 172; 104).

Besides the problems associated with correct site recognition, observed in all types of nucleases, a very recent study identified a pre-existing adaptive immunity to Cas9 proteins in humans (173). This is critical point that must be taken into account before any clinical application of the CRISPR/Cas9 system is attempted in patients.

1.1.3. Site-specific recombinases (SSRs)

The superfamily of site-specific recombinases, also known as integrases, were originally found in phages, where they catalyze the site-specific recombination between phage attachment sites (attP) and bacterial attachment sites (attB). The attP and attB are relatively short, yet long enough to be specific on a genomic scale. SSRs can be grouped into two mechanistically distinct subfamilies: (31; 174): (i) tyrosine recombinases or λ integrase family, they use a conserved tyrosine catalytic residue to form a covalent bond between the recombinase and the DNA target (175), and (ii) serine recombinases, resolvase/invertase family. Serine recombinases use a conserved serine residue to establish the covalent link between the recombinase and its DNA target (176; 177). Tyrosine recombinases catalyze the recombination of two sequential single-strand cleavage-ligation steps at the recombination sites (178). Serine recombinases operate through a different mechanism than tyrosine recombinase. In a single step, they generate double-strand DNA cleavage, rotate the helices about a common axis and ligate at the recombination site (179; 180). SSRs recognize distinct sequence-specific motifs termed Recombination Target sites (RTs) (31; 181).

Different kinds of tyrosine recombinases, such as Cre-loxP (33), FLP-FRT (182) and Dre-rox (183) have been shown to function in mammalian systems. Cre and FLP are site-specific recombinases that specifically recognize the loxP and FRT DNA sequences, respectively. The reversible nature of Cre and FLP presents a limitation when their use in genome editing is considered. The target site for each of these recombinases is identical before and after the recombination reaction, which makes them potential targets for a reverse excision reaction when a second incision is attempted; furthermore, the excision is thermodynamically favoured. As a result, these enzymes are not efficient at inserting DNA, since the inserted DNA could be removed by the same enzyme (184). Contrary to that, serine integrase enzymes can be utilized to achieve site-specific insertion in which the same enzyme cannot reverse the insertion (179; 180).

It is estimated that there are 30 members belonging to the large serine recombinase subfamily (176), including phiC31 (176), Bxb1 (185; 186), φRv1, A118, and TP901-1. There are fifteen phage-encoded serine recombinases for which the attachment sites are known. Nine of these fifteen recombinases have been shown to mediate accurate site-specific integration into genomic DNA in mammalian and prokaryotic cells (phiC31, Bxb1, ϕBT1, ϕC1, MR11, TP901-1, R4, A118, and ϕ RV) while six (TG1, ϕ 370.1, Wβ, BL3, SPBc and K38) have not yet been shown to be active outside of their native hosts (187). Comparing the fifteen serine recombinases for DNA manipulations in vertebrate cells, both Bxb1 and phiC31 demonstrated better recombinase activity than the others integrases. In addition, Bxb1 integrase presents higher recombination frequency and less damage than the next best recombinase, the phiC31 integrase (187).

1.1.3.1 Mechanism of action of Serine Recombinases

Serine site-specific recombinases mediate the integration and excision of the phage genome into and out of the host DNA using the same molecular mechanisms for both processes (188) (Figure 5 A).

The att sites for serine integrases (Int) are small, 51 bp attB site and 50 bp attP attachment site are sufficient for recombination and to maintain site selectivity (34) (Figure 5 B). Although it has been reported that even smaller sites, 34 bp for attB and 39 bp for attP, are active in attP/attB recombination (189; 34). Serine integrase do not require accessory proteins to bind and rotate the attP and attB sites. After rotation, the result is an integrated phage genome flanked by the newly produced attL and attR sites (Figure 5 A), each containing half sites derived from the original attP and attB sites (Figure 5 B) (34; 190; 191).

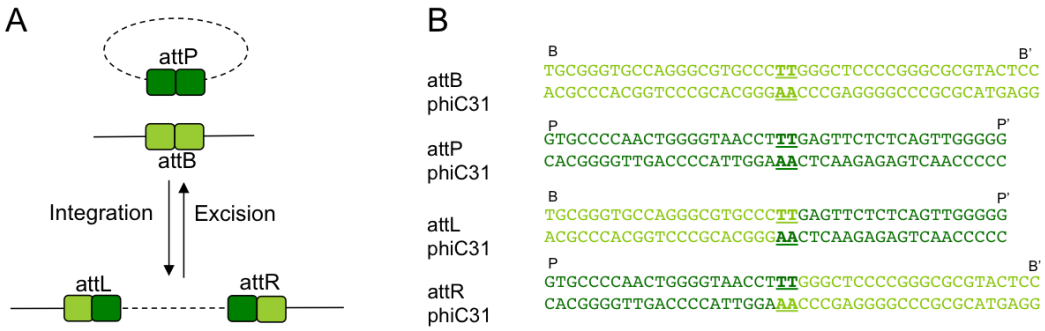

Figure 5| Site-specific recombination by serine integrases. **(A)** Integration: A serine integrase recombines an attP site (dark green) and an attB site (light green) to integrate a phage genome into a bacterial host genome. This reaction is directional and generates an attL (left) and an attR (right) which flank the integrated prophage. Excision: In the presence of a recombination directionality factor, the serine integrase can recombine attL and attR sites and restore the original attP and attB sites. **(B)** Attachment sites sequences of phiC31: An attB site (light green) consists of two half-sites, B and B', on either side of a central dinucleotide core sequence (bold), and an attP site (dark green) consists of P and P' half-sites, on either side of a central dinucleotide core sequence (bold). If the central dinucleotides in attP and attB sites match, they recombine to make attL and attR. Adapted from Merrick 2016.

The attP site contains P and P' half sites and the attB site contains B and B' half sites. Serine integrases bind their cognate attP and attB sites and bring them together to form a synapse containing an integrase tetramer (Figure 6). The Int/*attP* and Int/*attB* complexes are conformationally distinct due to the different positioning of a Zinc ribbon Domain (ZD) (191; 192). During recombination, a nucleophilic serine in each of the four subunits within a single synaptic complex attacks the DNA backbone at staggered positions on either side of the overlap sequence, forming a covalent 5' phosphoserine bond and freeing a 3' hydroxyl group (180; 179; 193). Afterwards, attP and attB sequences receive double stranded cuts on either side of the overlap sequence; half-sites B and P´ are joined to form attL sites, while P and B' 0 are joined to form attR sites through 180° rotation in relation to the others (188; 191; 194). This reaction is highly directional in that attL and attR sites do not recombine in the presence of integrase alone (187; 190; 191; 195) (Figure 6).

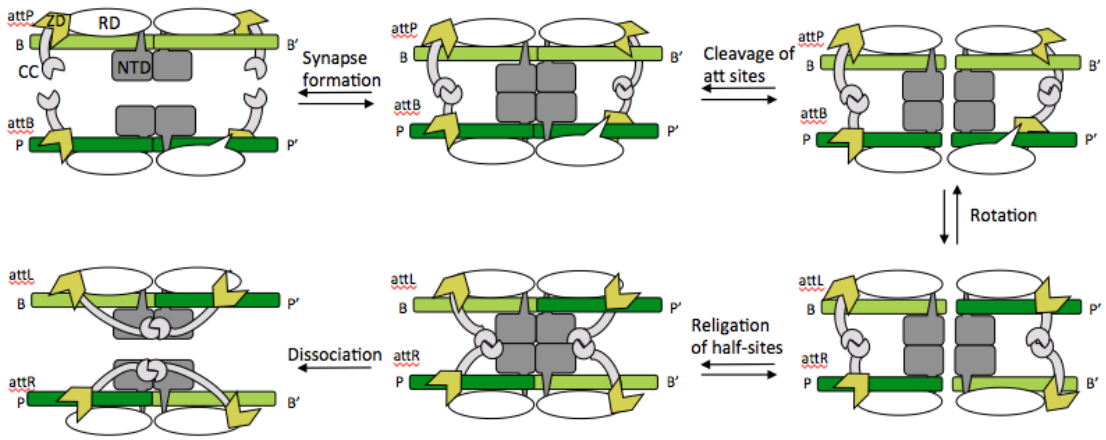

Figure 6| Mechanism of serine integrase recombination. *Synapse formation*: The serine integrase binds attB (light green) and attP (dark green) sites as a dimer, and brings the sites together to form a tetrameric protein-DNA synapse that is stabilized by interactions between coiled-coil motifs. *Cleavage of att sites*: nucleophilic serine residues in the integrase N-terminal domains (NTD) catalyze cleavage of all four DNA strands at the central dinucleotide, forming 5'-phosphoserine linkages between integrase subunits and DNA half-sites and generating 3'-dinucleotide overhangs. *Rearrangement of half-sites*: integrase subunits rotate 180° about a horizontal axis relative to the P and B-linked subunits. *Religation of half-sites*: If the 3' overhangs of the rearranged half-sites match, they are religated to form attL and attR sites. *Dissociation*: the unique arrangement of Zinc ribbon domains (ZDs) (yellow arrow) in the *attL* and *attR* sites allows the CC motifs to form intra-molecular interactions that prevents the reaction from running efficiently in the reverse direction. NTD: N-terminal catalytic domain (dark grey); RD: recombinase domain (white); CC: coiled-coil motif (pale grey). Adapted from Merrick 2016.

When the phage again enters the lytic cycle, its genome is excised by integrase-mediated recombination between the attL and attR sites, restoring in the process the attP and attB sites (Figure 5). Therefore, integration and excision of the phage genome need to be tightly regulated to prevent integration during the lytic cycle, to permit integration in a cell entering lysogeny and to enable excision during induction. Integrase is required for both integration and excision, but excision requires a phage-encoded accessory protein, named recombination directionality factor (RDF), to control the activity of integrase. The RDF binds to the C-terminal domains of integrase dimers on attL and attR and positions them to make the tetrameric synapse (196; 197;

198). In the absence of RDF, all phage integrases are unidirectional, mediating only integration. In the presence of the RDF the directionality of integrase is inverted, thus leading to activation of excision and blocking of integration (199; 200)

The phiC31 recombinase is one of integrases used by *Streptomyces* phages (189), whereas Bxb1 recombinase integrates into *Mycobacterium smegmatis groEL1* gene. Both are used to establish the lysogenic life cycle in bacteria and display the same recombination mechanism (185).

1.1.3.2 Serine integrases

PhiC31

PhiC31 is the most studied integrase. It is active in human cells, catalysing both intramolecular and intermolecular recombination (34; 201). In 2002, it was demonstrated that the phiC31 integrase system mediated genomic integration of coagulation factor (factor-IX) cDNA in mice at therapeutic levels (202). Subsequently, analogous experiments have been carried out with similar results in liver (203), hematopoietic stem cells (204), muscle (205), demonstrating the feasibility of phiC31 integrase system for gene therapy. Currently, phiC31 has been shown to catalyze the integration of up to 18 kb DNA fragments into mammalian cells (206).

PhiC31 integrases consist of a conserved N-terminal domain essential for DNA cleavage, strand exchange, and protein-protein interactions (207) and a large C-terminal domain, which is required for DNA binding (196). Though the wild-type phiC31 integrase functions adequately in mammalian cells, it is not very efficient due its low level of activity (208). In an attempt to further improve phiC31 recombination efficiency, distinct mutations have been introduced in the N-terminal (209; 210) and the C-terminal domains (196; 197).

In addition to the integrase itself, the attachment site is another element with an important effect on recombination efficiency. In the human genome there are native

sequences with partial identity to the wild attP site, named pseudo-attP sites, which could be used by phiC31 integrase to integrate exogenous DNA into the genome, although with low integration frequency (201; 211; 212). However, higher recombination efficiencies can be obtained by integrating canonical attP sites into the chromosomes of living mammalian cells, thus avoiding integration into the unfavoured pseudo attP sites (211; 213).

To ensure a safe and efficient use of phiC31 in gene therapy, potential interactions between the integrase and endogenous elements in mammalian host cells have been studied. Several research groups have identified patterns of cell type dependent phiC31 activity. These differences have been associated to variances in chromatin structure affecting the accessibility of attP sites by the recombinase. They have also hypothesized the existence of other, yet undiscovered, cell-specific cofactors of phiC31 that could influence integrase activity (187; 214; 215; 216). It has been shown that the death domain-associated protein (DAXX protein) is an inhibitor of phiC31 (215; 217), and that different expression levels of DAXX (215) might regulate phiC31 activity.

As mentioned above, RDFs are essential for the excision reaction of integrases. The phiC31 RDF is a 27.5 kDa protein encoded by the gp3 gene. Gp3 induces attL/attR recombination and inhibits attP/attB recombination in a stoichiometric rather than a catalytic manner. Kinetic analysis of the excision reaction suggests that gp3 modifies the interactions between Int, attL and attR to enable the formation of the excision synapse (195; 218). The attL/attR recombination by phiC31 plus gp3 had been proved in several mammalian cell systems with varying efficiencies (34; 219).

Bxb1

Like phiC31, Bxb1 recombinase has been used to insert a wide variety of genes, such as coagulation factor IX (203), vascular endothelial growth factor (220), ß-globin (204), α-L-iduronidase (221) and full-length human dystrophin, (206; 222) into mammalian cells. It has been reported that Bxb1 integrase can insert sequences as long as 27 kb

(187; 206; 223). The integration is catalyzed by a 500 aa serine integrase that is composed of two domains, an N-terminal 150 aa catalytic domain, and a 350 aa C-terminal domain that confers DNA binding activity (224). Contrary to phiC31, neither modifications of Bxb1 structure nor evidence suggesting that Bxb1 could mediate integration at pseudo att sites in the human genome have been reported so far (225).

The Bxb1 RDF, gp47, is a 30.6 kDa protein (255 aa) and contains a conserved domain of the protein phosphatase family. The biochemical mechanism of action of gp47 indicates that this protein does not directly bind to DNA, but instead controls integrase directionality through protein-protein interaction, similarly to gp3 function in the phiC31 system. Despite the weak interaction between gp47 and the attL/attR-bound integrase, gp47 can stimulate formation of the integrase:attL:attR synapse and the regeneration of attB and attP sites (218; 226).

When using phiC31 and Bxb1 in the same system, it is important to ensure that both recombinases cannot interact with each other. There are several studies showing that phiC31 does not recognize the Bxb1 attachment sites; conversely, Bxb1 does not interact with phiC31 attachment sites, regardless of whether they are isolated or present in the same integration plasmid (59; 187; 227).

1.1.4. Transposons

Transposable elements are natural, non-viral gene-delivery vehicles that have the ability to move and replicate within genomes (228). Depending on whether they require reverse transcription for transposition or not, transposable elements can be classified into two major classes:

Retrotransposons. Genetic information is transposed as RNA intermediates (35). Retroelements transpose through a replicative (copy-and-paste) process, in which the donor element itself does not move but, instead, a copy is produced and inserted elsewhere in the genome (229). This group includes the

Long Interspersed Elements (LINEs) (230), Short Interspersed Elements (SINEs) (231), and Long Terminal Repeat (LTR) retrotransposons (232).

DNA transposons. Genetic information is transposed as DNA by a conservative (cut-and-paste) mechanism of transposition, in which the element gets excised from the donor locus and is subsequently reinserted in a different genomic location (35). This group includes elements such as piggyBac (PB) isolated from the cabbage looper moth *Trichoplusia ni* (36) and Sleeping Beauty (SB), reconstructed from inactive Tc1 transposons from fish (233).

Transposons constitute a big proportion of the human genome. Almost half of it (45%) derives from retrotransposons, although there has been a marked decline in the overall activity of these elements in the hominid lineage. Around 2-3% of the genome comes from DNA transposons, which seems to be equally inactive, with no apparent activity in the past 50 million years (234; 235).

The natural capacity of transposons to move genomic material in the genome has been adapted to be used as a genome editing tool to deliver genes like ß-globin (236; 237), coagulation factor IX (238; 239), µdistrophin (240; 241), to targeted cell or organisms, or as a method to produce insertional mutagenesis (242).

Since the discovery of DNA transposition in maize (35) many different transposable elements have been described, but two of them, piggyBac and Sleeping Beauty, are the ones most commonly used for genome manipulation in vertebrate species (243).

The mechanism of action is the same in both types of DNA transposons. In their natural form, they consist of a single gene encoding the transposase polypeptide flanked by inverted terminal repeats (ITRs) which contain binding sites for the transposase (36; 38). However, DNA sequences can also be mobilized if the transposase is supplied in trans. For example, by its expression from a separate plasmid vector, and the new DNA sequence to be transposed is flanked by the ITRs in their original orientation (233; 244). Transposases are able to perform excision and insertion process. Both transposition activities are carried out by the same catalytic

domain of the transposase (244) (Figure 7-Figure 8). Beside all these similarities, there are differences between both transposases.

piggyBac

PB operates using a precise cut-and-paste mechanism, targeting and duplicating TTAA tetranucleotide sequences upon insertion, and reforming a single target site upon excision (36). Excision by PB does not leave behind nucleotides at the donor site (Figure 7). The piggyBac system has been reported to be highly active when introduced into different mammalian genomes, including the human genome (245; 246).

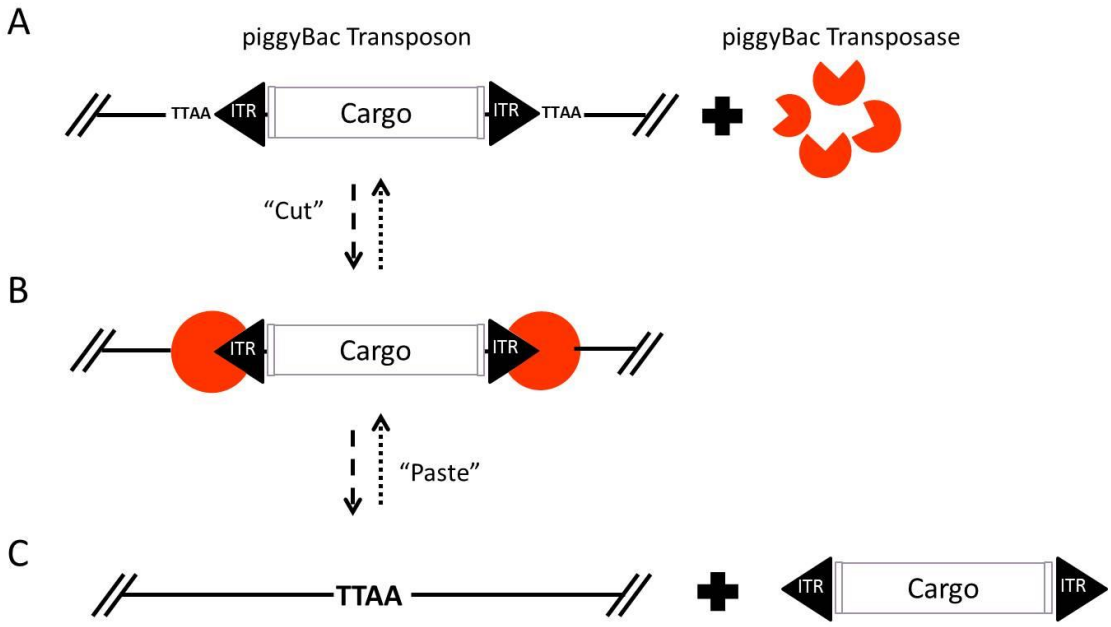

Figure 7| piggyBac transposition. (A) piggyBac transposon elements (ITR) flanking DNA sequence (cargo) and PB transposase (orange). **(B)** PB transposase (orange) recognizes their ITR elements. **(C)** Excision and mobilization of cassette flanked by ITRs elements and TTAA PB recognition region in DNA receptor.

In an attempt to increase its activity, a novel PB transposase has been generated. This new, hyperactive PB transposase (HAhyPBase) contains 7 aa substitutions with respect to the wild type version and shows more than a 10-fold higher rate of transposition than the mammalian codon-optimized piggyBac (247).

Sleeping Beauty

SB transposons can integrate into TA-dinucleotide (TATA) pairs at any location in the genome. As a result of the staggered cut at the TATA target sites, the transposons are flanked by TA dinucleotides on both sides after integration (233; 248). Excision by SB from the donor site involves staggered, double stranded DNA breaks at each side of the transposon, which result in a small number of nucleotides at the termini of the transposon being left behind (249). Each ITR element of SB transposon consists of two Direct Repeats (DRs) (249). The outer DRs, termed Lo and Ro, are located at the left and right termini of the transposon, respectively, and the inner DRs, Li and Ri, are located nearby the cargo. All DRs contain binding sites for SB transposase. Both the outer and inner DRs are required for efficient transposition, but they are not interchangeable (233; 242; 250; 251) (Figure 8). Sleeping Beauty system shows efficient transposition in cells of a wide range of vertebrates, including humans (38; 252; 247; 253; 254).

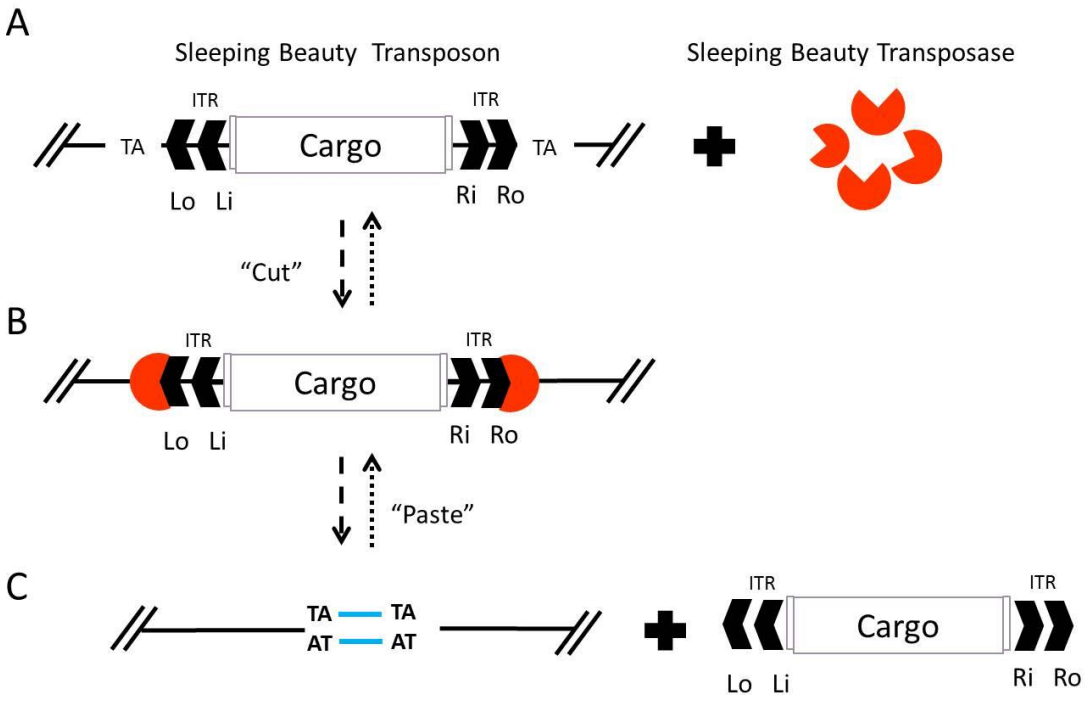

Figure 8| Sleeping Beauty transposition. **(A)** SB transposon elements (ITR) flanking DNA sequence (cargo) and SB transposase (orange). **(B)** SB transposase (orange) recognizes their ITR elements. **(C)** Excision and mobilization of cassette flanked by ITRs elements, leaving a TA motive at each end of the transposon's original location.

In order to improve SB transposition efficiency, new hyperactive mutants of SB transposase were developed by incorporating phylogenetically conserved amino acids from related transposases belonging to the Tc1 transposon family, e.g. the amino-acid at position 243 in SB varies in the Tc1 family of transposases between M, Q or H, being M the one present in the original SB. The Q does not increase the activity of the enzyme, but a version of containing H (SB100X) confers a 100 fold increase in activity (255).

1.2. Cellular systems

The era of efficient genome editing technologies has made possible, for the first time, the tailored modification of genetic information in living cells. The simplicity and adaptability of the new methods allows the creation of novel approaches in different research areas. Biopharmaceutical applications include the synthesis of stable cell lines with entire new genetic regulatory pathways, increased production of synthetic proteins, biosensors or even "intelligent networks" capable to detect endogenous stimuli and react to deficiencies in a host, upon transplantation (256). Another line of action is based on the ability of iPS cells to proliferate and give rise to practically any differentiated cell type in culture. Tissue specific cell derived from genetically modified iPS cells could be used to repopulate damaged organs or tissues in patients affected by hereditary diseases, cancer or age-related disorders.

While stable cell lines, have been routinely used in laboratories for the last decades, iPS cells have been recently discovered and will be therefore describe in some detail in the following chapter.

1.2.1. Cell reprogramming

In 1958, Gurdon reported the concept and first example of rejuvenation and cellular reprogramming through somatic cell nuclear transfer (SCNT) in frogs (257). Cellular reprogramming refers to the process whereby cells travel back up their maturation path, process during which the gene expression profile of a cell type is converted into the gene expression profile of the other cell type.

It took around 40 years before animal cloning could be attained in another vertebrate species, a sheep, showing that erasure of epigenetic memories that set somatic cell fate was possible even in mammals (258).

During this time, embryonic stem cells (ES cells) were identified in mice (259). ES have the inherent capacity to generate all cell types in the organism, a property known as

pluripotency (259; 260) (Figure 9). In 1998, human embryonic stem cells were isolated and grown in the lab (49). The rejuvenation of a cell to the pluripotent state has also been shown by fusing somatic cells with embryonic stem cells (261; 262). These two approaches suggest that fertilized eggs and pluripotent stem cells contain hidden reprogramming factors that are able to erase the somatic memories. Other research demonstrated that transcription factor(s) are crucial for both the maintenance of cellular identity, and the determination of cell fate (263; 264; 265; 266)

Based on those propositions, in 2006 Takahashi and Yamanaka announced the successful derivation of pluripotent stem cells from adult mouse fibroblast through the ectopic co-expression of only four genes (45). After screening a set of 24 gene candidates selected for being linked to ESC pluripotency, they proved that four factors were sufficient to reprogram adult fibroblasts into iPS cells. These reprogramming factors were octamer-binding transcription factor-3/4 (OCT3/4), SRY-related high-mobility-group (HMG)-box protein-2 (SOX2), c-MYC and Kruppel-like factor-4 (KLF4) (Figure 9). The discovery showed the importance of transcription factor networks in cell fate determination, and helped to understand the nature of cellular reprogramming.

iPS cells and ESCs show similar molecular and functional features (46), and they are cultured under the same conditions (46; 267; 268; 269; 270). Pluripotent stem cells (iPS cells and ESCs) are capable, when cultured, of unlimited self-renewal and differentiation into virtually all adult cell types (47; 48). These properties raised reasonable expectations that the cells could be useful tools for research on disease mechanisms, screenings for therapeutic drugs, and as therapeutic agents in regenerative medicine applications.

The procedure of deriving human embryonic stem cells implies embryo disaggregation (49), which itself poses ethical controversies that hindered research progress. In addition, it is not possible to generate retroactively patient-specific therapeutic cells from embryos. Unlike ESCs, iPS cells are produced from somatic cells which are not associated with ethical controversies (50), and permit autologous transplantations avoiding the host immune response (51).

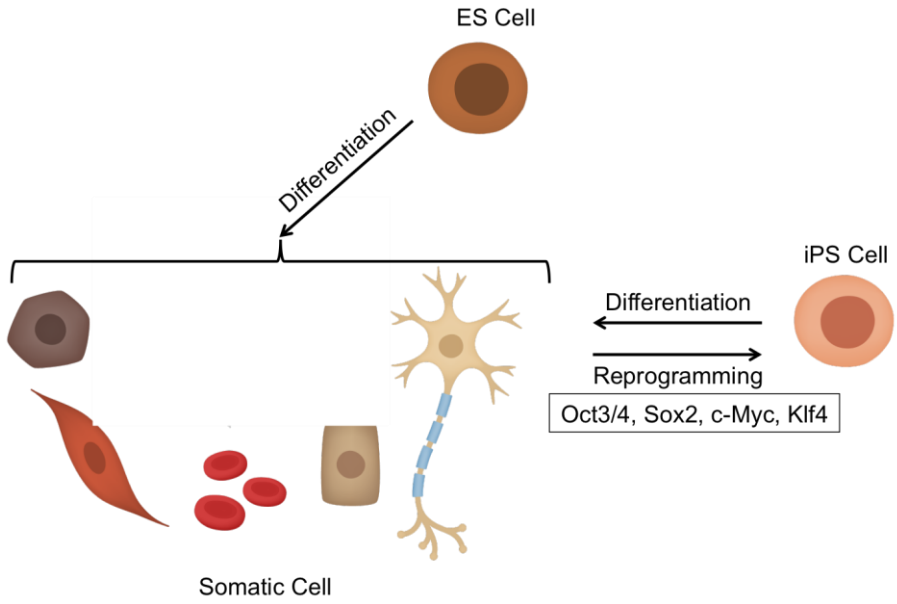

Figure 9| Differentiation and reprogramming processes. ESCs have the capacity to self-renew and to differentiate into cellular derivatives of the endodermal, ectodermal, and mesodermal lineages. iPS cells have equal capacities. Reprogramming allows us to turn any cell of the body into a pluripotent stem cell.

Since the seminal Yamanaka paper, a number of studies have demonstrated the feasibility of obtaining iPS cells from a wide variety of cell types, such as pancreatic β cells (271), neural stem cells (272), peripheral blood (273), fibroblast (274), etc. Recent years have also brought significant progress in increasing the efficiency and reducing the risks associated to this technique, making it more amenable to applications in the fields of regenerative medicine, disease modelling and drug discovery. Although the original gene set for reprogramming comprised four factors, follow-on studies have reduced, changed or increased the number of reprogramming factor required (275; 276; 277; 274). Additionally, modifications in the delivery of the reprogramming factor to the targeted cells have significantly improved the reliability of the process (278; 279; 280).

1.3. Delivery vectors

Choosing the optimal vector/delivery system is a delicate process that must take into account many factors, like the targeted cells and their characteristics, the duration of expression required, and the size of the genetic material to be incorporated into the vector (281). The mechanisms to deliver the DNA, RNA or proteins into the cells can be generally categorized in viral or non-viral vectors. Viral delivery vectors substitute the non-desired parts of the viral genome by the genetic information to be delivered into the cells. Non-viral vectors typically make use of chemical or physical means to deliver biomolecules (DNA, RNA or protein) into living cells or organisms (98).

1.3.1. Viral vectors

Viral vectors take advantage of the natural ability of viruses to enter into and deliver genetic material to cells (282). The most commonly used vectors are retroviruses, lentivirus, adenoviruses, and adeno-associated viruses (AAV).

Retroviral vectors. Replication-competent *retrovirus vectors* are based on *murine* leukaemia virus (MLV) (283). Retroviruses infect many types of cells in mammals. They need cells to divide in order to introduce their non-segmented RNA genome into the cell DNA and, therefore, will not infect many tissues where host cell growth and division have come to a standstill (284; 285). Retrovirus vectors have a relatively limited size capacity to carry genes, around 8 kb (286).

Lentiviral vectors. Lentiviral vectors derived from the human immunodeficiency virus (HIV). Lentivirus are able to integrate their genomic RNA into dividing and non-dividing cells of human and other mammalian species (287). They may carry up to 9 kb of genetic information (288).

Adenovirus vectors. Adenoviruses are non-enveloped (naked) viruses with an icosahedral nucleocapsid containing a double strand DNA genome. They have a broad range of vertebrate hosts; there are more than 100 serotypes characterized in

primates alone (289). Adenoviruses infect most cell types, so they are able to infect even non-dividing cells. This viral vector does not incorporate its genetic material into the host genome, so their DNA remains free in the cell nucleus. Their capacity to carry large DNA loads is not significantly better than other viral vectors, around 8 kb (64).

Adeno-associated viral vectors (AAVs). AAVs are small, non-pathogenic and single stranded DNA viruses from the Parvovirus family (290). They infect a broad range of cells including both dividing and non-dividing cells (290). The virus requires additional genes to replicate, which are provided by adenovirus or herpes virus that carry cytostatic and cytotoxic genes to the target cells (291; 292; 290; 293). AAVs present a limited packaging capacity of around 5 kb (294).

Although viral vectors are commonly used to transfer genes into cells and/or tissues (64), they have several limitations including immunogenicity, broad tropism, limited DNA packaging capacity, and difficulty of vector production and the unintended initiation of tumorigenic processes. Non-viral vector might resolve many of these limitations, particularly with respect to safety (98) and packaging capacity, a very important parameter when considering big payloads (68).

1.3.2. Non-viral vectors

Complex biomolecules like DNA, RNA and proteins must be able to reach specific cells (295; 296); and once on target, they have to cross several cellular barriers and avoid the cellular degradation machinery before reaching their destination inside the target (297; 298). All these requirements can be fulfilled by using non-viral vectors. Non-viral delivery tools are safe and specific, easy to produce in high quantities, and generally elicit low immunological responses (98). In addition, they are relatively inexpensive, can mobilize big molecules, and can be adapted to comply with specific needs to improve their efficiency (98).

Non-viral vectors include physical and chemical methods. The simplest physical method is direct injection (299). Other, more complex ones include electroporation (300), sonoporation (301; 302) or magnetofection (303; 304). They are used to temporarily disrupt cell membranes by electrical field, ultrasonic frequency or magnetic field, respectively (300; 301; 304). These physical methods are efficient at enhancing membrane penetration *in vitro*, although their efficiency *in vivo* is significantly lower (98; 305; 64). Chemical methods are generally synthetic/natural biodegradable liposomes (Lipoplexes), polymers (Polyplexes) or lipopolymers (Lipopolyplex) (306). Briefly, the liposomes consist of cationic lipids that encapsulate the hydrophilic molecules to be transferred and subsequently fuse with the cell membrane and release the cargo into the cytoplasm of the cell (307; 308). The polymers use a net-positive charge to from electrostatic complexes with anionic molecules (309; 310), and the complex liposomes/polymers use a polymer to condense biomolecule and a lipid coat to aid entry into the cell (306; 311; 312; 313).

1.4. Genomic Safe harbors

Genomic safe harbors (GSHs) are sites in the genome where newly inserted genetic elements function predictably and do not cause alterations of the host genome posing a risk to the host cell or organism. GSHs can support both addition of endogenous or exogenous genes and confer lost or new properties, respectively, and knock-out generation (314).

Currently, there are several loci that can be safely used for transgene insertion in human cells, such as adeno-associated virus site 1 (AAVS1), H11, human orthologue of the mouse ROS26 or the chemokine (C-C motif) receptor 5 (CCR5) gene, all of which are in exons or introns of endogenous genes and are differentially expressed in different tissues.

AAVS1. The AAVS1 locus on chromosome 19 (position 19q13.42) is a well characterized site for transgene expression in human pluripotent stem cells. Although AAVS1 supports transgene expression in multiple cells types, there are some cell lineages in which transgenes inserted at the AAVS1 locus can be silenced (315; 316; 317; 318; 55). Overall, gene targeting at AAVS1 supports stable transgene expression without significant silencing during *in vitro* and *in vivo* differentiation (319).

H11 locus. The H11 locus is a safe, intergenic, transcriptionally active region on chromosome 22 which can support expression of multiple transgenes in hiPS cells and is a reliable option for genomic editing (59). For example, iPS cells from limb girdle muscular dystrophy patients were corrected for dysferlin expression by targeting the wild-type dysferlin cDNA to the H11 locus (60)

Human ROSA26. The human ROSA26 locus is on chromosome 3 (position 3p25.3). The human ROSA26 locus is also located near genes that can potentially be dysregulated by transgene targeting into this location. This safe harbor locus has thus far not been routinely used to express transgenes in PSCs (58). Hence, the efficacy and safety profile of this site is uncertain.

CCR5. CCR5 is a co-receptor involved in HIV-1 infection of macrophages and T cells. Individuals homozygous for the CCR5Δ32 mutation are resistant to HIV-1 infection and do not manifest any major pathology, showing normal cellular differentiation and function (57). Preliminary studies have shown that the homozygous deletion of the CCR5 gene, located on chromosome 3 (3p21.31 position), does not have any effect in humans (320). For this reason, modifying the CCR5 gene by genetic engineering methods will not have any damaging effect for the cell (99; 132).

1.5. Aims of this study

The aim of this project is to generate a stable docking platform for secure and easy loading of genetic information into the genome of living cells. To build this safe harbor system, a complex strategy combining different molecular tools has been designed. The strategy provides a serial mechanism by which multiple docking modules, each comprised of four docking sites, can be assembled into the CCR5 locus.

The objectives of this study are:

1. To design a novel docking platform that does not require the use of nucleases for adding genetic material into cells.
2. To design a novel docking platform that allows serial addition of new docking sites.
3. To assemble the docking platform in human cells.
4. To demonstrate the functionality of the platform in a stable human cell line.

In order to achieve our objectives, we have designed a protocol based on several technologies presented in the introduction. Our design uses two recombination systems, TALEN-assisted homologous recombination and site-specific recombination mediated by the phiC31 and Bxb1 integrases. PhiC31 is used during the construction of the platform and Bxb1 is used in a posterior phase for the uploading of relevant genetic elements into the platform. Finally, the two transposon systems, Sleeping-Beauty and piggyBac are used to eliminate selection markers, vector fragments and any other left-over DNA residue that could interfere with the efficiency of the cargo loading steps or the posterior expression of the newly incorporated genes (Figure 10 and 11). As a proof of concept, the platform has been assembled in HeLa cells and later demonstrated to work also in human iPS cells.

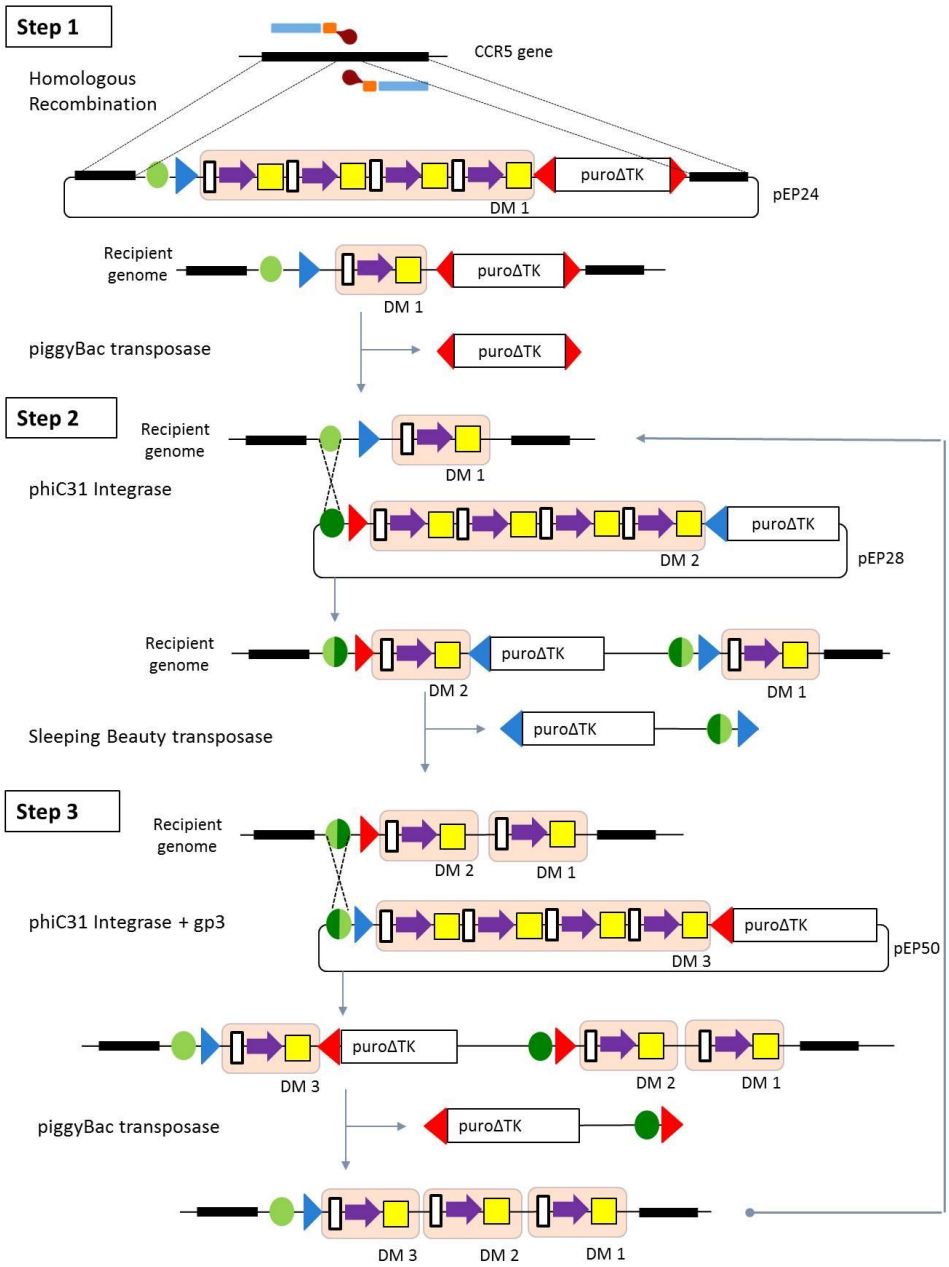

Figure 10| Diagram of the building strategy. **Step 1**, integration of the vector carrying a docking module (DM) by TALEN-assisted homologous recombination into the CCR5 locus, followed by excision of the selection element (puroΔtk) by PB transposition. **Step 2,** integration of a second docking module by phiC31-assisted site-specific recombination. **Step 3**, integration of a third docking module by phiC31/gp3-assisted site-specific recombination. Bold black line: CCR5 recombination arms. Light green circles: phiC31-specific attB sites. Dark green circles: phiC31-specific attP sites. Dark/light green circles: phiC31-specific attR sites. Light/dark green circles: phiC31-specific attL sites. Blue triangles: SB-specific ITR elements. Red triangles: PB-specific ITR elements. DM composition: white rectangle: cHS4 insulator, purple arrows: SV40 promoters and yellow squares: Bxb1-specific attB sites.

The sequential implementation of the three steps described above constitutes a complete cycle of integration. After the cycle, the targeted locus has the same molecular structure than at step 2, which makes it possible to continue adding as many new docking elements as needed by simply repeating the last two steps of the protocol.

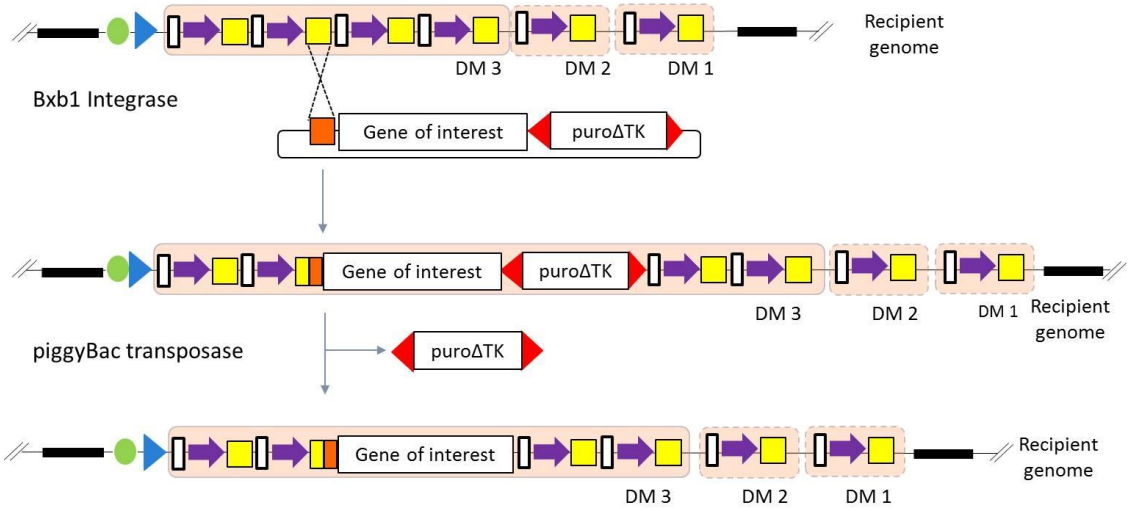

Figure 11| Diagram of the loading process. Integration Bxb1-assited site-specific recombination of a typical loading module carrying a promoterless gene. Bold black line: CCR5 recombination arms. Light green circle: phiC31-specific attB sites. Blue triangle: SB-specific ITR elements. Red triangle: PB-specific ITR elements. White rectangle: cHS4 insulator. Purple arrow: SV40 promoter. Yellow square: Bxb1-specific attB site. Orange square: Bxb1-specific attP sites. Half yellow/orange square: Bxb1-specific attL sites.

The design allows also for the incorporation of new docking elements after some of them have already been loaded (Figure 11).

2. MATERIALS AND METHODS

2.1. Materials

2.1.1. Laboratory equipment

Autoclave	Presoclave 75L (P-Selecta, Barcelona, Spain)
Centrifuges	BR4i multifunctions-S20 (Thermo Fisher Scientific, Massachusetts, USA).
	Biofuge fresco (rotor #3328) (Heraeus Instruments, Hanau, Germany)
	FiberLite F14-6x250LE, legend XTR centrifuge (Thermo Fisher Scientific, Massachusetts, USA)
	B4i multifunction. Thermo electron corporation (Thermo Fisher Scientific, Massachusetts, USA)
Thermo regulate devices	Modell 100-800 (Memmert, Schwabach, Germany)
	Flask. Infors HT-Ecotron (Bottmingen, Switzerland).
	Water bath PLS200 (Grant Instruments, Shepreth, UK)
	Thermomixes comfort (Eppendorf, Hamburg, Germany)
PCR machine	DNA engine- Peltier Thermal Cycler (Biorad Laboratories, California, USA)
	PTC-100 Programmable Thermal Controller (Biorad Laboratories, California, USA)
qPCR machine	iQ™5 Multicolor Real-Time PCR Cycler (Biorad Laboratories, California, USA)
Agarose Electrophoresis Chamber	(Biorad Laboratories, California, USA)

Spectrophotometer	Nanodrop (Thermo Fisher Scientific, Massachusetts, USA)
Nucleofector	4D-Nucleofector (Lonza, Basel, Switzerland)
Incubator	Stericycle 160 CO_2 incubator (Thermo Fisher Scientific, Massachusetts, USA)
Biosafety Cabinet	CLASS II-Cabinet. Telstar bio-II-A (Azbil Telstar, Sl, Terrassa, Spain)
Microscopy	Leica DMIL (Leica, Wetzlar, Germany) Leica DMI 6000B inverted fluorescence microscope (Leica, Wetzlar, Germany)
Flowcytometry	Cytopedia BD INFLUX cell soter (BD Biociences, NYSE, USA)

2.1.2. Plasmids

pKOscrambler V920	Cloning vector with ampicillin resistance (Stratagene, California, USA)
pmCherry-C1/C3	Mammalian expression vector designed to express a protein of interest that is fused to mCherry at the C-terminus mCherry is, a mutant fluorescent protein derived from the tetrameric *Discosoma sp*. red fluorescent protein, DsRed. (Clontech Laboratories, California, USA)
pEP24-SV40	Expression vector construct containing CCR5-5' and CCR5-3' homology arms, attB recombination site of phiC31 recombinase, SB transposon, cHS4 (insulator derived from the chicken β-globin locus), SV40 (simian virus

	promoter), attB recombination site of Bxb1 recombinase, PB, PGK-puroΔtk (double selection cassette), hybrid puromycin/thymidine kinase under the control of the mouse phosphoglycerate kinase 1 promoter flanked by PB transposon.
pEP28-SV40	Expression vector construct containing attP phiC31 attachment site, left-ITR of PB transposon. It harbors four attb Bxb1 elements plus other structures that constitute the docking module. PGK-puroΔtk (double selection cassette): hybrid puromycin/thymidine kinase protein under the control of the mouse phosphoglycerate kinase 1 promoter flanked by right-ITR of SB.
pEP50-SV40	Expression vector containing the attR phiC31 attachment site, left-ITR of the Sleeping Beauty transposon. It harbors four attB Bxb1 elements plus other structures that constitute the docking system. PGK-puroΔtk (double selection cassette): puromycin/thymidine kinase under the control of the mouse phosphoglycerate kinase 1 promoter flanked by right-ITR of PB.
pEP46-mcherry	pmCherryC1 vector containing the attP Bxb1 attachment site at 5’ of mCherry coding sequence. Promotorless.
pCMV_int	Expression vector containing the phiC31 recombinase for use in mammalian system. It has been kindly provided by Dr. Michele Calos, Stanford University (USA).
pCS_KI	Expression vector containing the fusion protein gp3-phiC31 recombinase for use in mammalian system. It has been kindly provided by Dr. Michele Calos, Stanford University (USA).

pCMV-Bx	Mammalian expression vector containing the Bxb1 recombinase under the control of the CMV promoter. Addgene #51552 (USA).
pCMV-HAhyPBase	pcDNA3.1 expression vector of the HA-tagged hyperactive piggyBac transposase.
SB100X	pCMV(CAT)T7-SB100 containing Sleeping Beauty (SB100X) transposase. Addgene #34879 (USA).
PB(R/L)	pUC57 cloning vector with the PiggyBac transposon. Genewiz (USA).
SB(R/L)	pUC57 cloning vector with the Sleeping Beauty transposon. Genewiz (USA).
SV40-attB-Bxb1	pUC57 cloning vector with four cHS4 elements (insulator sequences derived from the chicken β-globin locus), four SV40 promoters and the attB attachment site for Bxb1. Genewiz (USA).

2.1.3. Primers

The following oligonucleotides were purchased from Thermo Fisher Scientific (Massachusetts, USA). Their use is described in the text.

Primers for molecular cloning

Name	Sequence (5' → 3')
FORWARD-Golden GATE	
GG1-F	GCGCGTCTAGAGGATCCCGAGCTCACG
GG2-F	GCGTCTAGACGTCTCA**CATG**CGAGCTCACG
GG3-F	GCATCTAGACGTCTCA**GGAC**CGAGCTCACG
GG4-F	GCTTCTAGACGTCTCA**CCAG**CGAGCTCACG
REVERSE-Golden GATE	
GG1-R	CGACGCGT**CATG**TGAGACGGCCCGGATGA
GG2-R	TTTACGCGT**GTCC**TGAGACGGCCCGGATGA
GG3-R	TTTACGCGT**CTGG**TGAGACGGCCCGGATGA
GG4-R	TTACGCGTGGCGCGCCTATTGCTAGCGCCCGGATGA
Recombination arms	
480up	GACTGAGTTGCAGCCGGGCATG
480down	ACCAACCAGGATCTCCCTGCTCAG
CCR5-5homUP	TATGCGGCCGCACGCGTTCCAGGCTGCAGTGAGCCATG
CCR5-5homDN	CCTGCTAGCGAGACGTCATTAAAACACAGCCACCACCCAAGTG
CCR5-3homUP	TATCCTGCAGGTGCTTGTCATGGTCATCTGCTACTCG
CCR5-3homDN	TTTACGCGTGCTTCCCCAGCTCTCCCAGG
oligonucleotides	
attP-Bxb1-up	CATGTGTCGTGGTTTGTCTGGTCAACCACCGCGGTCTCAGTGGTGTACGGTACAAACCCAG
attP-Bxb1-down	CTAGCTGGGTTTGTACCGTACACCACTGAGACCGCGGTGGTTGACCAGACAAACCACGACA

Primers for molecular characterization

Name	Sequence (5'→3')
CCR5-5for2	CAGGCTTCCCGCATTCAAAAT
attBphiC31	GATGGGTGAGGTGGAGTACG
CCR5-3rev2	TGTCTCCTTCTACAGCCAAGC
PURO/tk	GGTAATGACAAGCGCCCAGA
PB_HR reint fwd	CGCATGTGTTTTATCGGTTGA
attB-Bxb1-1	GCGCTAGCGAATTCGTATGTT
attLphiC31	CAACCCCTTGTGTCATGTCG
attRphiC31	TTTTCCCAGGTCAGAAGCGG
attxphiC31	AGTGTGATCACTTGGGTGGTG
3'_SB	CATCACATTCCCAGTGGGTCA
cHS4_R	GTAATTACATCCCTGGGGGCTT
PB_R	TGACGCATGTGTTTTATCGGT
SB_R	TCCCTGTCTTAGGTCAGTTAGGA
CCR5_Cat_980_UP	AAGATGGATTATCAAGTGTCAAGTCC
CCR5_Sang_490_20UP	TTAAAAGCCAGGACGGTCAC
Bxb1.BQ-F	GGACAGGTATCCGGTAAGCG
Bxb1.BQ1-R	CCCGTGAGCTCGCATGT
Bxb1.BQ2-R	GTGAGCTCGGTCCTGAGAC
Bxb1.BQ3-R	GTGAGCTCGCTGGTGAGAC
Bxb1.BQ4-R	AGCTGCAGGTTTAAACAGTCG

2.1.4. Bacteria strains and mammalian cell lines

2.1.4.1. Bacterial strains

Supercompetent XL1-blue cells were used for all cloning and transformation procedures. XL1-Blue Genotype: *recA1 endA1 gyrA96 thi-1 hsdR17 supE44 relA1 lac* [F´ proAB lacIqZΔM15 Tn10 (Tetr)].

2.1.4.2. Mammalian cell lines

HeLa cells. Human cell line derived from cervical cancer cells (321).

HAFi-W3R feeder cells. Immortalized Human Adult Fibroblast (HAFi) with resistance to three antibiotics (G418, puromycin and hygromycin) (322; 323; 324). In addition, over-expression of the Wnt3a transgene enhanced the ability of HAFi-W3R feeder cells to support the undifferentiated growth of hiPS cells (325).

IMEDEAi003-A. human induced pluripotent stem (hiPS) cells reprogrammed from skin biopsies (fibroblast) of healthy individual by MKOS retroviral infection (326).

2.1.5. Software

Electrophoresis	GeneSnap (Syngene, NJ, USA)
Spectrophotometer	ND-100 V3.8.1 (Thermo Fisher Scientific, Massachusetts, USA)
Sequence analysis and Design	Vector NTI Express Designer (Thermo Fisher Scientific, Massachusetts, USA), SerialCloner2-1 (Serial Basics, France)
Online applications	BLAST data analysis (NCBI, Bethesda MD, USA)

qRT PCR	qPCR Analysis Software (Biorad Laboratories, California, USA)
Fluorescence microscopy	Microscope Imaging Software: Leica Microsystems (Leica, Wetzlar, Germany)
Flowcytometry	Sortware (BD Biociences, NYSE, USA)

2.2. Methods

2.2.1. Molecular biology methods

2.2.1.1. RNA isolation and quantitative RT-PCR

Total RNA was isolated using the RNeasy Mini Kit (Qiagen, Hilden, Germany) following manufacturer's instructions. 50 ng of RNA was used to synthesize cDNA with the Sensiscript® Reverse Transcriptase (Qiagen, Hilden, Germany). Master mix reaction and thermocycler program are showed below:

Master mix-Sensiscript	
10xbuffer RT	2 µl
25xdNTPs	0.8 µl
oligo dt (20 µM)	1 µl
RNAase inhibitor	1 µl
(10U/ µl)	1 µl
RNA	1 µl
H_2O	4.2 µl

cDNA synthesis program		
Step 1	25˚C	10 min
Step 2	37˚C	120 min
Step 3	85˚C	5 min
Step 4	4˚C	∞

qPCR analyses were performed on a iQ™5 Multicolor Real-Time PCR Cycler (Biorad Laboratories, California, USA) with the KAPA SYBR FAST Universal qPCR Kit (KAPA Biosystems, Woburn/MA, KK4600). The primers for either endogenous or transgenic genes have been previously described (269).

	Forward (5'→3')	Reverse (5'→3')
NANOG	ACAACTGGCCGAAGAATAGCA	GGTTCCCAGTCGGGTTCAC
CRIPTO	CGGAACTGTGAGCACGATGT	GGGCAGCCAGGTGTCATG
REX1	CCTGCAGGCGGAAATAGAAC	GCACACATAGCCATCACATAAGG
GAPDH	GCACCGTCAAGGCTGAGAAC	AGGGATCTCGCTCCTGGAA
Trans-OCT4	TGGACTACAAGGACGACGATGA	CAGGTGTCCCGCCATGA
Trans-SOX2	GCTCGAGGTTAACGAATTCATGT	GCCCGGCGGCTTCA
Trans-KLF4	TGGACTACAAGGACGACGATGA	CGTCGCTGACAGCCATGA
Trans-c-MYC	TGGACTACAAGGACGACGATGA	GTTCCTGTTGGTGAAGCTAACGT
Endo-OCT4	GGGTTTTTGGGATTAAGTTCTTCA	GCCCCCACCCTTTGTGTT
Endo-SOX2	CAAAAATGGCCATGCAGGTT	AGTTGGGATCGAACAAAAGCTATT
Endo-KLF4	AGCCTAAATGATGGTGCTTGGT	TTGAAAACTTTGGCTTCCTTGTT
Endo-c-MYC	CGGGCGGGCACTTTG	GGAGAGTCGCGTCCTTGCT

Master mix reaction and thermocycler program for qPCR analysis are showed below:

Master mix-Sensiscript®

SYBR	12.5µl
Primer UP	0.5µl
Primer DOWN	0.5µl
Template	1µl
H_2O	10.5µl

Program thermocycler

Step 1	95˚C	3 min
Step 2	95˚C	10 sec
Step 3	55˚C	30sec
Step 4	Go to Step 2 (40 cycles)	

2.2.1.2. Cloning procedures

All cloning steps involved standard techniques such as quantitative restriction enzyme analysis, separation of DNA fragments by agarose gel electrophoresis, DNA extraction and purification of the desired DNA fragments, ligation, bacterial transformation, plasmid DNA miniprep purification and, finally, analytical plasmid digestion to identify the correctly cloned construct.

2.2.1.2.1. Enzymatic digestion

All enzymes used in digestion analyses belong to the type IIS restriction enzyme family. Digestions were incubated for 2-4 h at 37-55˚C according to manufacturer's instructions. Conditions for standard reaction depending on number of enzymes used are showed below:

Standard unique enzyme digestion	Standard double enzymes digestion
2 µl specific enzymatic buffer (10x)	3 µl specific enzymatic buffer (10x)
0.5 µl enzyme	0.5µl enzyme A
2 µg plasmid DNA	0.5 µl enzyme B
x to 20 µl ddH_2O	2 µg DNA
	x to 30 µl ddH_2O

To verify the newly cloned constructs, plasmid DNA was digested with different restriction enzymes. The unique electrophoretic pattern of the fragments indicated if the right insert had been correctly cloned into the vector. Two control digestions were usually carried out. First, plasmids were digested with enzymes that cut inside and outside the insert. Then, to confirm correct cloning, the resulting positive clones were digested with another enzyme resulting in a different restriction pattern.

2.2.1.2.2. DNA purification from agarose gels

Analytical and preparative gel electrophoresis of double-stranded DNA fragments was performed in 0.8% for 0.8-4 kb or 2% for 0.05-0.8 kb agarose gels supplemented with ethidium bromide (0.1 mg/ml) (327). DNA fragment sizes were determined by using DNA markers (50 bp or 1 kb depending on the length of the expected DNA fragments). Bands were visualized using a UV transilluminator at 302 nm.

DNA fragments were excised from agarose gel with a scalpel and purified on columns according to the manufacturer's instructions, E.Z.N.A Gel extraction kit (Omega, Georgia, USA).

2.2.1.2.3. Ligation reactions

All ligation reactions were performed using the Mighty Mix DNA ligation Kit (Takara Bio, Kioto, Japan) at a 1:3 vector:insert molar ratio. Standard ligation mixes generally contained 100 ng of total DNA. Standard ligation reaction was prepared as follows:

8 µl ligation mix
10 µl DNA fragments
2 µl ddH_2O
V_T = 20 µl for 5 min at 25˚C

2.2.1.2.4. Bacterial transformation

XL1-blue supercompetent cells were transformed following standard protocols. The complete ligation mixture (20 µl) was added to 50 µl of competent cells ($1x10^5$), mixing gently by pipetting up and down. The mixture was incubated for 30 min on ice, exposed to heat shock for 40 sec at 42°C, and chilled on ice for 40 sec. Finally, 1000 µl of Luria-Bertani (LB) media was added for the recovery phase and incubated for 60 min at 37°C with vigorous shaking (180 rpm). Transformed bacteria were centrifuged at

13,000 rpm and the supernatant discarded, leaving around 30-60 µl which were seeded on pre-warmed agar plates with the appropriate selection antibiotic. Agar plates were incubated O.N. at 37◦C. On the following day, 6-24 colonies were picked and individually grown in 2 ml LB plus antibiotic on a shaker at 180 rpm O.N. at 37◦C. Next day, plasmid DNA was extracted using the miniprep kit (mi-Plasmid Miniprep kit-Metabion, Germany).

2.2.2. Karyotyping

A standard optimized G-banding technique with slight modifications was used to karyotype hiPS cell lines. Briefly, actively proliferating cells were treated with 10 µg/ml colcemid (Sigma, Spain) for 1 to 3 h, trypsinized, washed in phosphate-buffered saline (PBS) and incubated in hypotonic PBS (74.85 mOsm/kg H_2O) for 20 min, before immersing them in Carnoy's fixative (cold methanol/acetic acid 3:1, both from BDH) for 30 min. Nuclei were then centrifuged at 500 xg for 2 min and resuspended in fresh Carnoy's to wash residual PBS. Fixed nuclei were spread and G-banded by Prenatal Genetics, Barcelona. At least 20 selected spreads were banded to produce a result in each line.

2.2.3. Immunofluorescence and histochemistry

Undifferentiated hiPS cells were washed with PBS, fixed for 20 min with 4% paraformaldehyde, and washed again with PBS. Immunocytochemistry was performed for NANOG (1:150, rabbit IgG polyclonal, Abcam, Spain), OCT4 (1:100, mouse IgG monoclonal, Santa Cruz Biotechnology, USA), SSEA3 (1:100, rat IgM, Millipore, USA), SSEA4 (1:100, mouse IgG, Millipore, USA), TRA-1-60 (1:100, mouse IgM, Millipore, USA), TRA-1-81 (1:100, mouse IgM, Millipore, USA), α-AFP (1:100, rabbit IgG, Dako, Denmark), NESTIN (1:500, rabbit IgG, Sigma, Spain), TUJ1 (1:500, rabbit IgG, Covance, UK), SOX17 (1:100, goat IgG, R&D Systems, UK), α-ACTININ (1:200, mouse igM, Sigma,

Spain) and GATA4 (1:300, mouse IgG, Santa Cruz Biotechonology, USA). Permeabilization was carried out with 0.2% Triton X-100 (Sigma, Spain) and 100 mM glycine in PBS for 30 min at room temperature. Blocking was performed by incubating the cells with 5% BSA in PBS for 30 min at room temperature. Cells were incubated with primary antibodies O.N. at 4ºC in 2% BSA in PBS, followed by three washing steps with PBS. Alexa Fluor 555 (1:500, donkey anti-mouse IgG or donkey anti-rabbit IgG Invitrogen, USA), Alexa Fluor 546 (1:500, donkey anti-goat IgG, Invitrogen, USA), Alexa Fluor 488 (1:500, donkey anti-mouse IgG, donkey anti-rabbit IgG or donkey anti-rat, Invitrogen, USA) were used as secondary antibodies, incubated 1 h at room temperature with 2% BSA in PBS. After washing with PBS, cells were stained with DAPI (5 min, 1 µg/ml), washed three times and mounted with Dako fluorescent mounting medium. Pictures were taken with a Leica DMI 6000B microscope system.

2.2.4. Fluorescence activated cell analysis and sorting.

72 h after transfection, cells were dissociated using Trypsin/EDTA (Invitrogen, USA) or Tryple Express (Thermo Fisher, USA), washed with PBS (Invitrogen, USA) and resuspended in sorting buffer (PBS with 2% FBS (Biowest, France) or KRS (Biowest, France), 0.2% penicillin/streptomycin and DAPI (1 µg/ml). To avoid clumps, cells were filtered through a 40 µm cell strainer. Flow cytometry analysis and sorting was performed with a Cytopedia BD INFLUX cell sorter (BD Biociences, NYSE, USA). DAPI positive cells were excluded from sorting. Cells were collected into 15 ml Falcon tubes with DMEM complete media supplemented with 20% FBS and 0.2% penicillin/streptomycin. Based on fluorescence intensity, cells were sorted and separated into low and high mCherry expressing cells. Subsequently sorted cells were centrifuged for 5 min at 320 xg and seeded in a p6 well.

2.2.5. Genomic DNA extraction

Cells were harvested and centrifuged in PBS. The supernatant was discarded and genomic DNA extracted using 300 µl lysis buffer (20 mM Tris, 100 mM NaCl, 1 mM EDTA, 1mg/ml proteinase K). The mix was incubated O.N. at 55˚C. Following centrifugation at 3200 rpm, supernatant was discarded and 700 µl NaCl/ETOH 100% (at -20˚C) was added. The reaction mix was incubated for 30 min at -20˚C, centrifuged and rinsed twice with 700 µl 70% ETOH. After last rinse, pellets were air-dried for 15-20 min. Samples were resuspended in 50 µl of ddH_2O.

2.2.6. PCR strategy for characterization of the genomic loci following recombination.

To evaluate the correct integration of the donor DNA during HR and site-specific recombination, 5' and 3' integration sites were checked by Polymerase Chain Reaction (PCR) with specific pairs of primers. In both cases, one primer pair attaches to a region outside and the other inside the recombination cassette.

phC31 dependant site-specific recombination events were characterized by detecting the newly generated att-sites. Primer sequences and their localization in the genome are presented in chapter 2.1.3. Length and location of PCR amplicons as well as melting temperature (Tm) of primers are presented Figure 12 and Table 1, respectively.

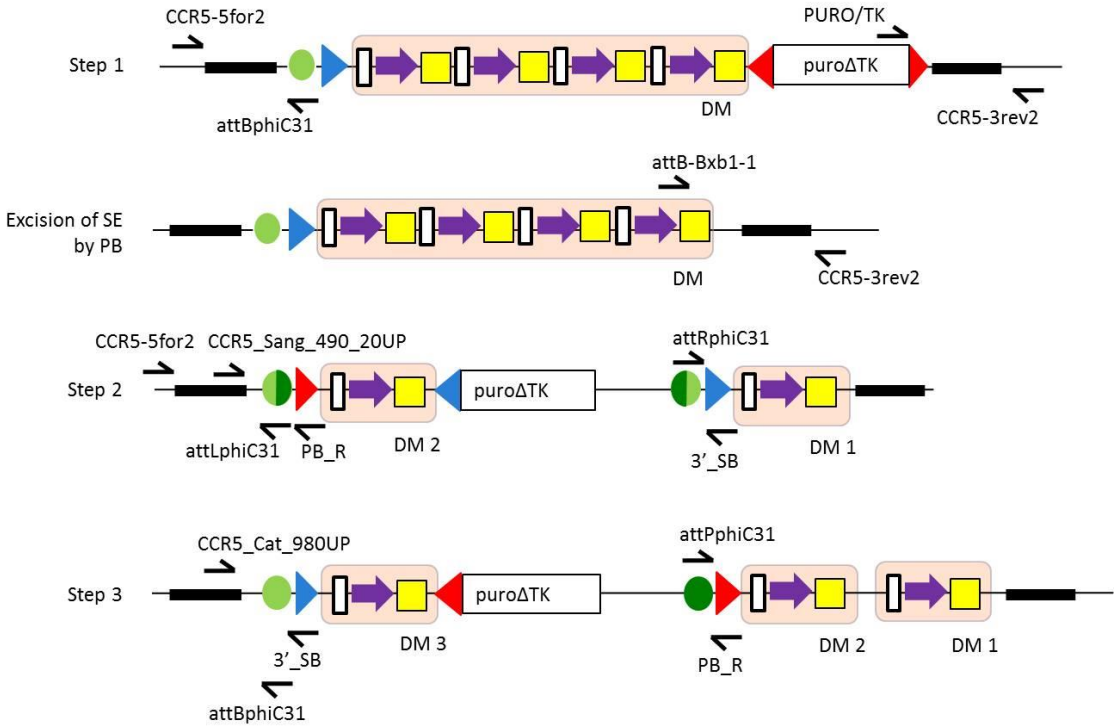

Figure 12| Position of primers used to verify the integration of the different modules used during the process. *Step 1,* CCR5 locus structure of after HR. CCR5-5for2 (forward) and attBphiC31 (reverse) primers to check 5' recombination, and PURO/tk (forward) and CCR5-3rev2 primers to check 3' recombination. *Excision of selection element* (*SE) by PB transposase*: attB-Bxb1-1 (forward) and CCR5-3rev2 (reverse) primers to screen for SE-free clones. *Step 2*, locus structure after site-specific phiC31 attB/attP recombination, CCR5-5for2 (forward) or CCR5_Sang_490_20UP (forward) and attLphiC31 or PB_R (reverse) primers to check 5' recombination. attRphiC31 (forward) and 3'_SB (reverse) primers to check 3'recombination. *Step 3*, locus structure after site-specific attL/attR recombination by phiC31 plus gp3, CCR5_Cat_980UP (forward) and attBphiC31 or 3'_SB (reverse) primers to check 5' recombination. attPphiC31 (forward) and PB_R (reverse) primers to check 3' recombination. Bold black line: CCR5 recombination arms. Light green circles: phiC31-specific attB sites. Dark green circles: phiC31-specific attP sites. Dark/light green circles: phiC31-specific attR sites. Light/dark green circles: phiC31-specific attL sites. Blue triangles: SB-specific ITR elements. Red triangles: PB-specific ITR elements. Docking module (DM) composition: white rectangle: cHS4 insulator, purple arrows: SV40 promoters and yellow squares: Bxb1-specific attB sites.

Bxb1-dependant site-specific recombination events were characterized by detecting the unique BsmBI site present in each one of the four docking sites present in the docking module (Figure 13). Primer sequences and their localization in the genome are

presented in Table 1 and Figure 13, respectively. PCR product lengths and melting temperature (Tm) of the different pairs of primers are also presented in Table 1.

Each docking element was assembled by Golden Gate cloning (see Results, 3.2), using the type IIS endonuclease BsmBI, whose recognition sequence includes 4 variable bases (CGTCTCN^NNNN). By modifying the bases, the Golden Gate system allows simultaneous assembly of the four modules in a single cloning step. At the same time, it provides specific PCR targets to check the docking status of the different units.

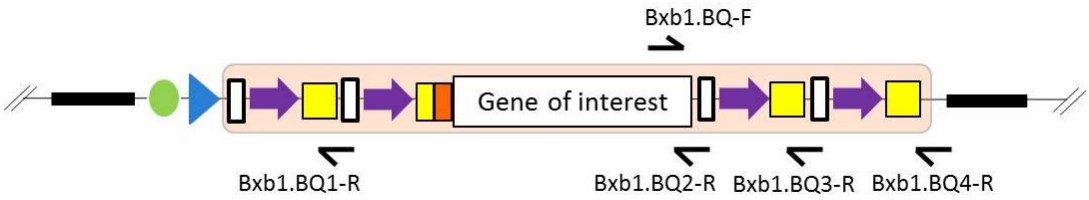

Figure 13| Position of primers used to verify loading status in each one of the different docking units. Docking module structure after site-specific integration by Bxb1. Bxb1.BQ-F and Bxb1.BQ-R (1-4) are the docking unit-specific primers. Bold black line: CCR5 recombination arms. Light green circle: phiC31-specific attB site. Blue triangle: SB-specific ITR element. DM composition: white rectangle: cHS4 insulator, purple arrows: SV40 promoters and yellow squares: Bxb1-specific attB sites. Half yellow/orange rectangle: Bxb1-specific attL sites.

Primer pairs		Length (bp)	Tm (°C)
Forward	**Reverse**		
CCR5-5for2	attBphiC31	1053	65
PURO/tk	CCR5-3rev2	2500	65
PB HR reint fwd	CCR5-3rev2	1800	65
CCR5-5for2	attLphiC31	1500	65
attB-Bxb1-1	CCR5-3rev2	1500	65
attRphiC31	3'_SB	173	63
attPphiC31	PB_R	343	63
CCR5_Sang_490_20UP	attLphiC31	170	63
CCR5_Cat_980_UP	attBphiC31	581	63
attxphiC31	cHS4_R	581	63
attxphiC31	SB_R	296	63
attxphiC31	PB_R	456	63
Bxb1.BQ-F	Bxb1.BQ1-R	306	60
Bxb1.BQ-F	Bxb1.BQ2-R	303	60
Bxb1.BQ-F	Bxb1.BQ3-R	303	60
Bxb1.BQ-F	Bxb1.BQ4-R	338	60

Table 1| Primer pairs used to verify each step of the synthesis of the CCR5 docking platform. The name of the primers, the length of the PCR product and the melting temperature (Tm) are shown**.**

PCR reactions were carried out using the WonderTaq polymerase (Euroclone, via Figino, Italy) according to the following program:

PCR program			Reaction mix
Step	Temperature	Time	5 µl buffer
Step 1	95°C	1 min	2 µl primer UP
Step 2	95°C	15 sec	2 µl primer DOWN
Step 3	X*C	30 sec	0.5 µl taq polymerase
Step 4	72°C	Y**	150-300 ng template (DNA)
Step 5	Go to step 2 (33 cycles)		x to 25 µl ddH_2O
Step 6	72°C	1 min	
Step 7	4 °C	∞	

*X: annealing temperature depends on PCR primer pairs.
**Y: elongation time depends on amplified PCR product (1 min for 1 kb length).

2.2.7. Sequencing

Genomic DNA from individual hiPS and HeLa cell clones was PCR amplified and Sanger sequenced using the primers attxphiC31/cHS4_R; attxphiC31/SB_R; and attxphiC31/PB_R. A 5 µl aliquot of the PCR reaction was run on a 2% agarose gel to check the size of PCR products. Following confirmation, PCR products were purified and sent to sequencing at SECUGEN (Madrid, Spain).

PCR program (ExTaq)

Step	Temperature	Time
Step 1	95°C	1 min
Step 2	98°C	10 sec
Step 3	x°C*	30 sec
Step 4	72°C	Y**
Step 5	Go to step 2 x 30 cycles	
Step 6	72°C	2 min
Step 7	4°C	∞

Reaction mix

5 µl buffer
4 µl dNTPs
2.5 µl primer UP
2.5 µl primer DOWN
0.25 µl taq polymerase
150-300 ng template (DNA)
x to 50 µl ddH_2O

*X: annealing temperature depends on the particular pair of PCR used.
**Y: elongation time depends on the length of the amplified PCR product (1 min for 1 kb length).

2.2.8. Cellular Biology methods

2.2.8.1. Cell culture

HeLa cells and subclones were cultured in complete medium (DMEM High glucose (Biowest, France) supplemented with 10% fetal bovine serum (FBS) (Biowest, France), 2 mM Glutamax and 1% non-essential aminoacids (NEAA) (Invitrogen, USA), 50 U/ml penicillin/streptomycin (Thermo Fisher Scientific, USA). HeLa cells were maintained in a humidified atmosphere at 37°C, 5% CO_2, Cells were passaged at 80-90% confluency (one or twice weekly) and never diluted more than 1:10. For passaging, culture

medium was removed and cells were washed twice with PBS. 1 ml of trypsin/EDTA (Invitrogen, USA) was added per 100 mm culture dish and incubated for 5 min at 37°C. Complete medium was then added and pipetted up and down to break up cell clumps. Trypsinized cells were collected and centrifuged for 4 min at 320 xg. Cells were resuspended, diluted at 1:10 ratio in 10 ml of complete medium and seeded on p100 plates.

HAFi-W3R feeder cells, inactivated by *γ*-irradiation, support the undifferentiated growth of hiPS cells. For this purpose, HAFi-W3R feeder cells were grown in 150 mm plates at 37°C, 5% CO_2, in a humidified atmosphere. Once cells reached 80-90% confluency, they were washed twice with PBS and trypsinized for 5 min at 37°C. Finally, 7 ml of DMEM medium was added to stop trypsinization and gently pipetted 10-15 times up and down to break up cells clumps into single cells. Totally, $2x10^8$ cells were harvested and distributed into four 50 ml conical tubes. To induce replicative arrest, cells were *γ*-irradiated at 40 Gy. Viable cells were counted using trypan blue-dye exclusion staining and centrifugated for 5 min at 270 xg. Inactivated cells were resuspended in cold freezing medium (90% FBS + 10% DMSO) adjusting cell density to $2x10^6$ viable cells/ml. 1 ml aliquots of the cell suspension were distributed into as many cryovials as needed and subsequently stored at -80°C. Next day, vials were transferred to -150°C container. When it required, inactivated frozen cells were thawed, centrifuged for 5 min at 240 xg. The pellet was resuspended in 10 ml of DMEM medium, seeded on gelatine-coated p100 plates and incubated O.N. at 37°C, 5% CO_2, in a humidified atmosphere. Next day, hiPS cells could be passaged onto the HAFi-W3R feeder layer.

hiPS cells, and the *subclones* derived from them, were cultured on irradiated HAFi-W3R feeder cells. The hiPS cells were cultured in hiPS medium consisting of Knockout (KO) DMEM (Invitrogen, USA) supplemented with 20% Knockout Serum Replacement (KOSR) (Invitrogen, USA), 2 mM Glutamax, 1% non-essential aminoacids (NEAA), 0.1 mM β-Mercaptoethanol (Invitrogen, USA) and 8 ng/ml basic fibroblast growth factor (bFGF) (Peprotech, UK). (325). The addition of ROCK inhibitor (10 μM Y-27632) to hiPS

media was necessary in the first 24 h after passaging, nucleofection and sorting. For nucleofection experiments, hiPS cells were transferred to matrigel-coated dishes (10 µg/cm^2) allowing growth in feeder-free conditions After culturing on matrigel for 2 day, cells were used for nucleofection experiments as described in section 2.2.8.3.

2.2.8.2. hiPS cell derivation

Primary cells were obtained from 2 mm^3 skin samples. Epidermal keratinocytes and dermal fibroblasts were isolated from the biopsies and cultured following a previously published protocol (328). Primary cultures were transduced with three (KLF4, OCT4, SOX2) or four (c-MYC KLF4, OCT4, SOX2) reprogramming factors by two rounds of retroviral infection. A detailed protocol has been already published (325). Reprogrammed cells were mechanically split. Pluripotent colonies were selected by morphology and seeded onto murine embryonic fibroblast feeder cells (325) to produce stable hiPS cell lines. Feeder cells were cultured in DMEM High glucose (Biowest, France) supplemented with 10% Fetal Bovine Serum (FBS) (Biowest, France), 2 mM GlutaMAX and 1% NEAA (Invitrogen, USA) and 50 U/ml penicillin/streptomycin (Thermo Fisher Scientific, USA). The hiPS cell lines were cultured in KO-DMEM (Invitrogen, USA) supplemented with 20% KOSR (Invitrogen, USA), 2 mM GlutaMAX, 1% NEAA, 0.1 mM β-Mercaptoethanol (Invitrogen, USA) and 8 ng/ml basic bFGF (Peprotech, UK). After generating a master bank of frozen cell lines at early passages, lines selected on the basis of colony morphology and replication efficiency were further amplified and cultured over antibiotic-resistant, Wnt expressing, human foreskin fibroblasts (323). Mechanical splitting was used to generate the frozen master bank, whereas enzymatic splitting using recombinant trypsin (TrypLE Select, Invitrogen, USA) was performed thereafter with the addition of 10µM ROCK inhibitor Y-27632 (Sigma, Spain) to enhance survival of dissociated single hiPS cells (329; 267).

2.2.8.3. Transfection methods

2.2.8.3.1. Lipofection: HeLa cells

HeLa cells were lipofected in p6 wells plates with jetPEI, a cationic linear PEI transfection reagent (Polyplus transfection, New York, USA), according to manufacturer's instructions. At the time of transfection, HeLa cells were 70 % confluents which roughly correspond to 2-4x10^5 cells growing in a p6 well. For each lipofection reaction two reaction mixtures were prepared. The first one consisted of 3 μg DNA diluted in 150 mM NaCl to a final volume of 100 μl and the second one, of 6 μl jetPEI reagent in 150 mM NaCl to a final volume of 100 μl. Solution were mixed gently, vortexed for 15 sec and spun down briefly. 100 μl jetPEI solution were added to 100 μl DNA solution, vortexed, spun down briefly, and then incubated for 30 min at room temperature. The 200 μl jetPEI/DNA mix was then drop-wise added to the cells maintained in 2 ml serum-containing complete medium. Cells were homogenously distributed on the plate by gently swirling. After culturing cells for 8 h at 37°C, growth media was replaced. Puromycin [2μg/ml], G418 [1.2mg/ml]) or FIAU ((1-(2-deoxy-2-fluoro-β-d-arabinofuranosyl)-5-iodouracil) [1μM]) were generally added 72 h after lipofection. Following 3-4 weeks of selection, individual cell clones were manually picked and expanded for further experimentation.

The docking platform was integrated into the genome in three recombination steps. Step 1 consisted in the incorporation of the first docking vector, pEP24, by TALEN-assisted homologous recombination followed by excision of the selection element by PB transposase. In Step 2, pEP28 was integrated through site-specific attB/attP recombination catalyzed by phiC31 integrase, followed by excision of the new selection element by SB transposase. Finally, in step 3 the pEP50-SV40 vector was incorporated by phiC31/gp3 site-specific recombination between the attR and attL sites. The selection element was then excised by PB transposase (see Results 3.1).

In step 1, the first transfection included three plasmids: pEP24, CCR5 FW (TALEN-F) and CCR5 REV (TALEN-R). Correct recombination events were selected with puromycin. To excise the selection marker (puroΔtk), 3 μg of the plasmid carrying PB transposase

were transfected, and negative selection with FIAU was employed. In step 2, two plasmids were transfected: pEP28 and phiC31 integrase. Correct recombination events were selected by puromycin. To excise the selection element (selection marker: puroΔtk) and remaining vector elements, 3 μg of the plasmid carrying SB transposase were transfected, and negative selection with FIAU was employed. In step 3, the transfection included two plasmids: pEP50, and phiC31 integrase plus gp3 (both encoded in the same plasmid). Correct recombination events were selected with puromycin. To excise the selection element, cells were transfected with the plasmid carrying PB transposase and negative selection with FIAU was administrated. The total amount of added plasmids was 3 μg in each transfection assay.

In order to evaluate the functionality of the Bxb1-specific attB attachment sites present in the docking modules, cells containing the 4 module docking platform were transfected with a loading plasmid, pEP46, carrying a promoterless mCherry fluorescence gene and the plasmid encoding Bxb1 integrase (pCMV_Bx) at a 1:1 ratio (total DNA: 3 μg) (Table 2).

Transfection	**Plasmids**			**μg DNA**	**Selection**
	Vector		**Enzyme**		
HR	pEP24	:	TALENs forward and reverse	1.5:0.75:0.75	Puromycin
phC31-mediated integration	pEP28	:	phiC31	1.5:1.5	Puromycin
phC31/gp3-mediated integration	pEP50	:	phiC31+gp3	1.5:1.5	Puromycin
SE Excision by PB	-		PB transposase	3	FIAU
SE Excision by SB	-		SB transposase	3	FIAU
Bxb1 loading assay	pEP46	:	Bxb1	1.5:1.5	G418-mCherry

Table 2| Transfection conditions in HeLa cells. For each transfection, donor plasmids and enzymes are indicated. SE excisions only required transfection of the plasmid encoding the corresponding transposase. Total volume of DNA was 3 μg per transfection.

2.2.8.3.2. Nucleofection: hiPS cells

hiPS cells were transfected by nucleofection with the P4 Primary Cell 4D-Nucleofector_Kit (Lonza, Walkersville, MD, USA) according to the manufacturer's instructions. To avoid feeder cells contamination, hiPS cells were expanded in p100 plates coated with matrigel solution (10 μg/cm2) prior to nucleofection. The cultures were dissociated with TrypleE (Thermofisher, USA) into single cells. Then, $1x10^6$ cells were nucleofected in a 100 μl cuvette. Totally, six nucleofections were carried out per experiment ($6x10^6$ cells).

In order to incorporate the first module of the docking platform by TALEN-assisted HR (step 1), hiPS cells were transfected with 4 μg of each plasmid. The first transfection included three plasmids: pEP24, CCR5 FW (TALEN-F) and CCR5 REV (TALEN-R). Correct recombination events were selected with puromycin. To excise the selection marker (puroΔtk), 8 μg of the plasmid encoding PB transposase were transfected, and negative selection with FIAU was applied (Table 3).

To evaluate the functionality of the Bxb1-specific attB attachment sites present in the docking module, cells containing the 4-module docking platform were transfected with a loading plasmid, pEP46, carrying a promoterless mCherry fluorescence gene, and the plasmid encoding Bxb1 integrase. Total amount of plasmid transfected was 8 μg (Table 3).

Transfection	Plasmid			μg DNA	Selection
	Vector		Enzyme		
HR	pEP24	:	TALENs forward and reverse	4:4:4	Puromycin
SE Excision by PB	-		PB transposase	8	FIAU
Bxb1 loading assay	pEP46	:	Bxb1	4:4	G418-mCherry

Table 3| Transfection conditions in hiPS cells. For each loading reaction, donor plasmids and enzymes are indicated. SE excisions only required the transfection of the plasmid encoding the corresponding transposase. Total volume of DNA was 12 μg to HR transfection and 8 μg for the others two transfection

After nucleofection cells were quickly reseeded onto multi-antibiotic resistant HAFi-W3R feeders. To enhance cell survival, the Rho kinase inhibitor Y27632 (Tocris Bioscience, Bristol, UK) was used for the first 24 h after nucleofection. Transfected hiPS cells were selected with 800 ng/ml puromycin (Invitrogen) during 2-3 weeks, starting 3 days after nucleofection. Surviving clones were picked manually and expanded for DNA analysis and further experimentation.

3. RESULTS

3.1. Design of building strategy

The aim of this project is to generate a stable docking platform for secure and easy loading of genetic information into the genome of mammalian cells. To build this safe harbor system, a complex strategy combining TALEN-assisted HR and site-specific recombination methods was designed, and constitutes the first result of the project. The strategy provides a serial mechanism by which multiple docking modules, each comprised of four docking sites, can be assembled into the CCR5 locus.

The initial step in the creation of the docking platform was the integration of the first docking module consisting of four docking sites at the CCR5 locus. Each docking site is composed of one cHS4 insulator element (330), the SV40 promoter (331; 332) and a Bxb1-specific attB attachment site (185; 186). In addition to the docking module, the donor plasmid, pEP24, also includes a recombination cassette flanked by 5' and 3' CCR5 homology arms. The recombination cassette is composed of a phiC31-specific attB attachment site, a Sleeping Beauty right-ITRs and a puroΔtk selection cassette flanked by piggyBac ITR elements. After a correct homologous recombination event, the puroΔtk gene is removed by piggyBac transposition (Figure 14).

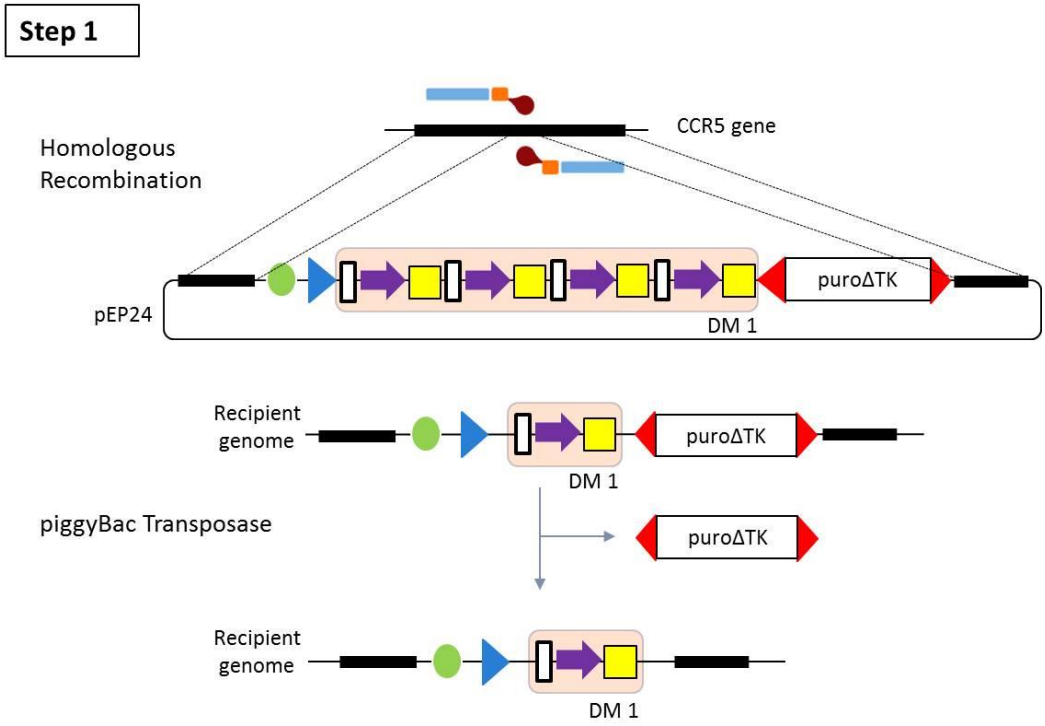

Figure 14|Diagram of the building strategy. Step 1. Integration of first docking vector carrying a docking module (DM) by TALEN-assisted homologous recombination into the CCR5 locus followed by excision of the selection module (puroΔtk) by PB transposase. Bold black line: CCR5 recombination arms. Light green circles: phiC31-specific attB sites. Light/dark green circles: phiC31-specific attL sites. Blue triangles: SB-specific ITR elements. Red triangles: PB-specific ITR elements. DM composition: white rectangle: cHS4 insulator, purple arrows: SV40 promoters and yellow squares: Bxb1-specific attB sites.

The next step is the incorporation, by site-specific recombination catalyzed by phiC31, of a second docking module identical to the first one. In this case the recombination vector, pEP28, contains the docking module, a phiC31-specific attP site, the right-ITR of PB, and a puroΔtk selection element preceded by the left ITR of SB. Site-specific integration results in two new attachment sites: attL and attR. After integration, left and right-ITRs SB appear flanking a section of the newly incorporated DNA fragment that contains the puroΔtk selection module, the attR and a portion of the recombination vector. This configuration facilitates the SB transposase to remove all the unnecessary elements after the right clones have been selected. The removal of

the attR attachment site is critical to avoid interferences during the third recombination step (Figure 15).

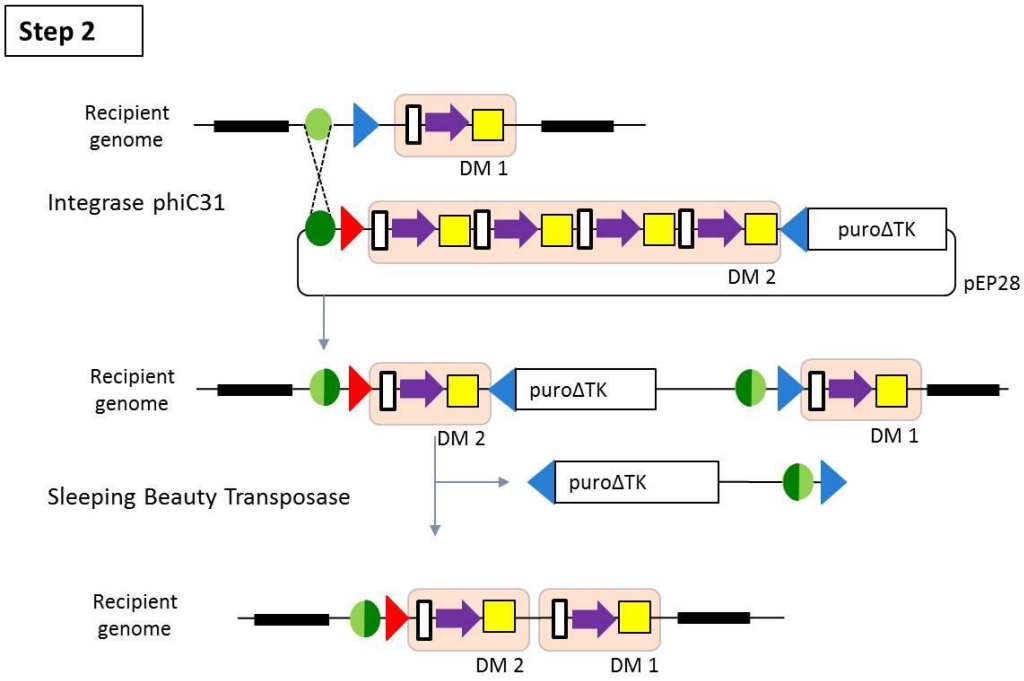

Figure 15| Diagram of the building strategy. Step 2. Integration of the second docking vector carrying a DM by site-specific phiC31-assisted recombination. Bold black line: CCR5 recombination arms. Light green circles: phiC31-specific attB sites. Dark green circles: phiC31-specific attP sites. Dark/light green circles: phiC31-specific attR sites. Light/dark green circles: phiC31-specific attL sites. Blue triangles: SB-specific ITR elements. Red triangles: PB-specific ITR elements. DM composition: white rectangle: cHS4 insulator, purple arrows: SV40 promoters and yellow squares: Bxb1-specific attB sites.

The final step in the creation of a docking platform involves integration of a third docking module consisting of another 4 docking sites. The new module is incorporated by site-specific recombination between phiC31-specific attL and attR sites. As explained previously, phiC31 by itself is not able to catalyze the reaction efficiently and has to be aided by the recombination directionality factor gp3. Besides the docking module, the recombination vector, pEP50, contains a phiC31-specific attR site, a right-

ITR SB element and a puroΔtk selection element preceded by a left-ITR of PB. Site-specific integration resulted in recombination between the attL and attR sites, thus regenerating the attB and attP sites. After integration, the right-ITR PB is located at the 3´end of the puroΔtk selection cassette in the correct orientation to allow the activity of PB transposase. PB activity leads to the removal of the attP site as well as vector fragments and the puroΔtk selection element (Figure 16).

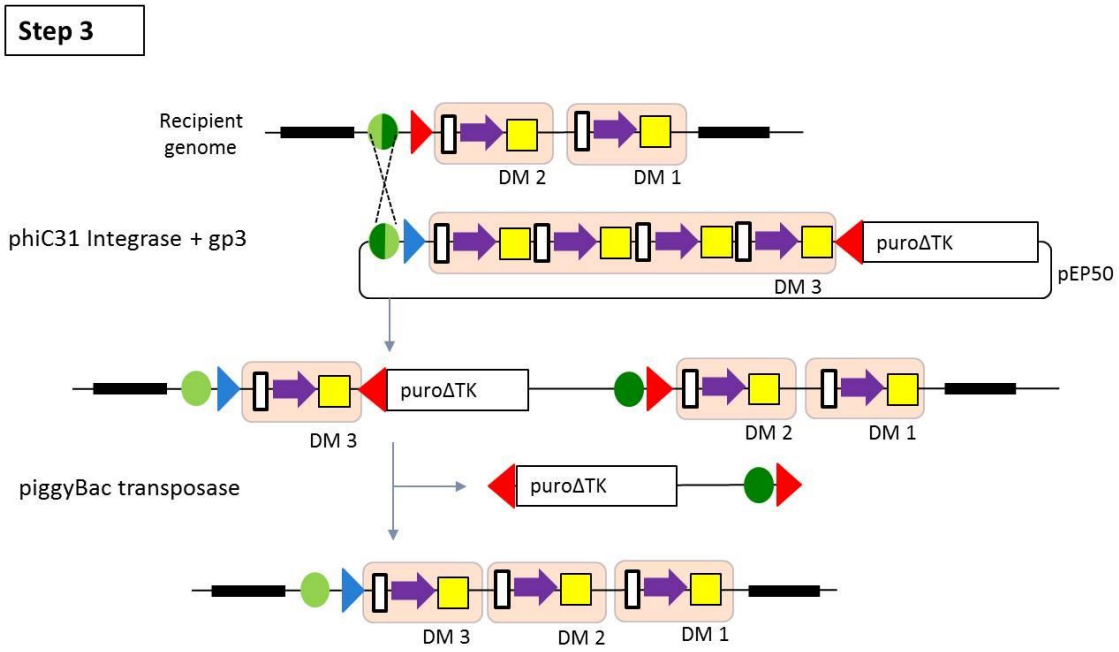

Figure 16| Diagram of the building strategy. Step 3. Integration of the third docking vector carrying a DM by site-specific recombination mediated by phiC31 and its RDF (gp3). Bold black line: CCR5 recombination arms. Light green circles: phiC31-specific attB sites. Dark green circles: phiC31-specific attP sites. Dark/light green circles: phiC31-specific attR sites. Light/dark green circles: phiC31-specific attL sites. Blue triangles: SB-specific ITR elements. Red triangles: PB-specific ITR elements. DM composition: white rectangle: cHS4 insulator, purple arrows: SV40 promoters and yellow squares: Bxb1-specific attB sites.

3.2. Plasmid construction.

To build the platform it was necessary to start by synthesising a series of plasmids carrying the different elements to be assembled in the final structure.

3.2.1. Recombination plasmids

Three plasmids were built: pEP24, pEP28 and pEP50 (Figure 17). pEP24 was used in the homologous recombination step. pEP28 and pEP50 are the plasmids used in the second and third recombination steps, respectively. All of them contain one docking module consisting of four docking sites.

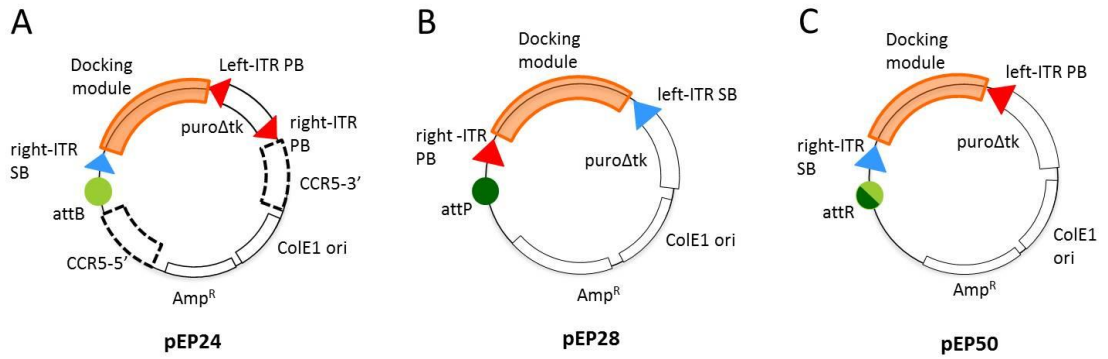

Figure 17| Schematic representation of the recombination vectors. **(A)** pEP24 contains two CCR5 homology arms, attB phiC31 site, right-ITR element of SB, a docking module and selection marker (puroΔtk) flanked by ITR element of PB. **(B)** pEP28 consists of an attP phiC31 site, right-ITR element of PB, a docking module and selection marker (puroΔtk) with an ITR element of SB at 5'. **(C)** pEP50 consists of an attR phiC31 site, right-ITR element of SB, a docking module and a selection marker (puroΔtk) with an ITR element of PB at 5'. Bold discontinue square: CCR5 recombination arms. Light green circles: phiC31-specific attB sites. Dark green circles: phiC31-specific attP sites. Dark/light green circles: phiC31-specific attR sites. Blue triangles: SB-specific ITR elements. Red triangles: PB-specific ITR elements. Orange rectangle: docking module.

The construction of the recombination plasmids has several common cloning steps that simplified their synthesis. The plasmid used as a common backbone was pKOscrambleV920 (see Materials and Methods, section 2.1.2). This plasmid is 1968 bp in length and confers ampicillin resistance. In the first step of cloning, specific phiC31 recognition sequences (attB, attP or attR), synthesized by Genewiz (USA), were incorporated into the vector between the AgeI and KpnI restriction sites (Figure 18).

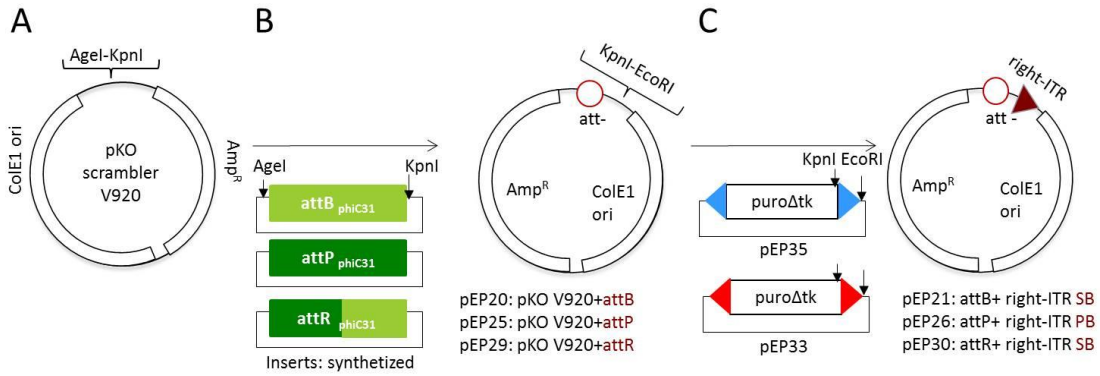

Figure 18| First and second steps of phiC31 recombination vectors construction. (A) pkOscramblerV920 carries the ColE1 origin of replication and the ampicillin resistance gene. **(B)** phiC31 attachment sites were synthetized flanked by specific (AgeI and KpnI) restriction sites. Three new plasmids were obtained after cloning each of the attachment phiC31 sites into pKOscramblerV920: pEP20 (attB), pEP25 (attP) and pEP29 (attR). **(C)** pEP35 harboring the selection marker (puroΔtk) flanked by ITR elements of SB. pEP33 harboring selection marker (puroΔtk) flanked by ITR elements of PB. Only the right-ITR element of each transposon was integrated in the final vectors generating three new plasmids: pEP21 (attB+right-ITR SB), pEP26 (attP+right-ITR PB) and pEP30 (attR+right-ITR SB). Light green rectangle: phiC31-specific attB sites. Dark green rectangle: phiC31-specific attP sites. Dark/light green rectangle: phiC31-specific attR sites. Blue triangles: SB-specific ITR elements. Red triangles: PB-specific ITR elements. Bordeaux circle: phiC31-specific attP, attB and attR sites. Bordeaux triangle: transposon-specific ITR elements.

As a result, three new plasmids were obtained: pEP20 with an attB phiC31 site, pEP25 with an attP phiC31 site and pEP29 with an attR phiC31 site (Figure 18). Restriction pattern analysis revealed positives clones for each plasmid. Six clones were analysed

for pEP20 and pEP25, and five clones for pEP29 (Figure 19). All colonies obtained from pEP25 and pEP29 were positives showing the expected restriction digestion pattern (Figure 19 B and C). pEP25 and pEP29 were digested by BglII and BsaI, respectively, obtaining two restriction bands with slightly different sizes: 1044 and 1135 bp for pEP25; 989 and 1131 bp for pEP29 (Figure 19 B and C). Regarding pEP20, digestion by SalI, of the only positive clone, pEP20.3, showed the correct restriction pattern (209 and 1853 bp) (Figure 19 A).

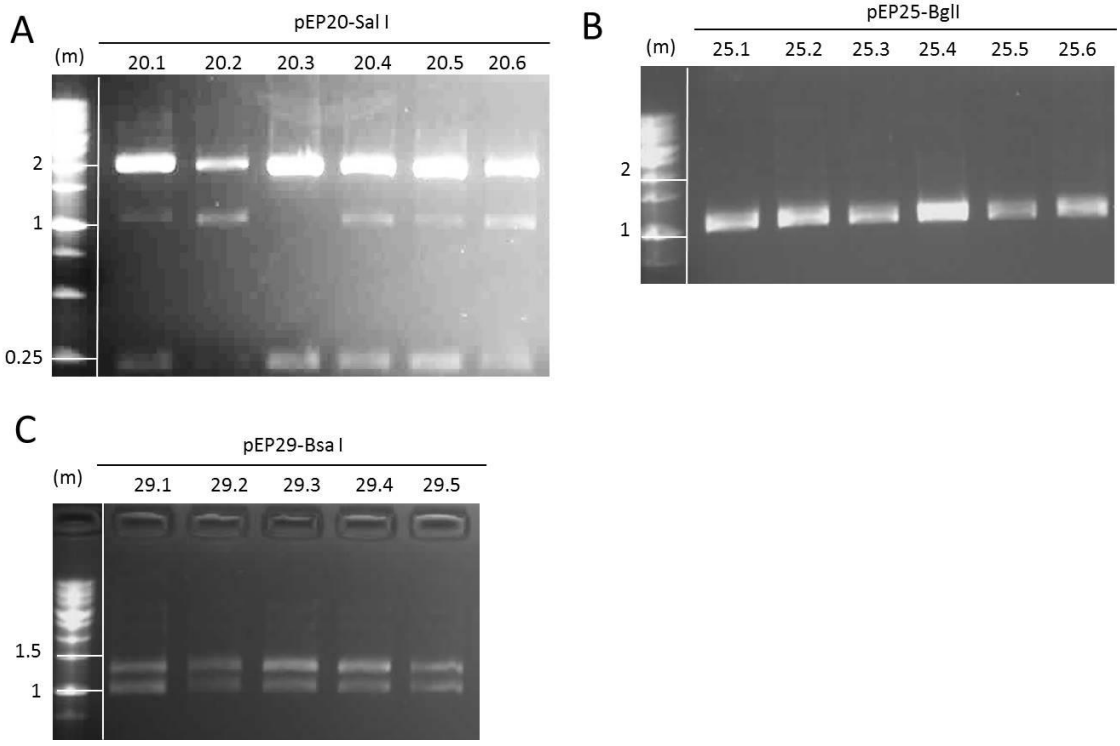

Figure 19| Restriction fragment analysis of pEP20, pEP25 and pEP29. Agarose gel (0.8%) electrophoresis with ethidium bromide staining. **(A)** SalI digestion of pEP20. Expected bands: 209 and 1853 bp. **(B)** BglII digestion of pEP25. Expected bands: 1044 and 1135 bp. **(C)** BsaI digestion of pEP29. Expected bands: 989 and 1131 bp. Molecular size marker, m= 1 kb DNA ladder.

In the next step, the right-ITR element of PB or SB, depending on the plasmid, were incorporated (Figure 18) into the newly generated plasmids pEP20, pEP25 and pEP29.

For this purpose, two new plasmids, pEP33 and pEP35 had been previously constructed. They consist of the puroΔtk selection element flanked by ITR elements of PB and SB, respectively. To incorporate the ITRs, pEP20, pEP25 and pEP29, were cleaved by KpnI and EcoRI. The right-ITRs were extracted, purified and incorporated into the corresponding pEP33 or pEP35 previously digested with the same enzymes. The right-ITR of SB were cloned into pEP20 and pEP29 and the right-ITR of PB into pEP25 giving rise to three new plasmids named pEP21, pEP26 and pEP30 (Figure 18).

Eight bacterial colonies were grown from pEP21, six from pEP26 and five from pEP30. pEP21 colonies were digested with SalI, showing the correct restriction pattern of two band of 435 and 1853 bp (Figure 20 A). The pEP26 clones were digested by XmnI and those from pEP30 by SalI. Three pEP26 clones (26.1-26.4-26.6) showed the expected restrictions bands of 509 and 1910 bp (Figure 20Figure 20 B). Three pEP30 clones (30.1-30.2-30.3) were also positive, producing bands of 493 and 1853 bp (Figure 20 C)

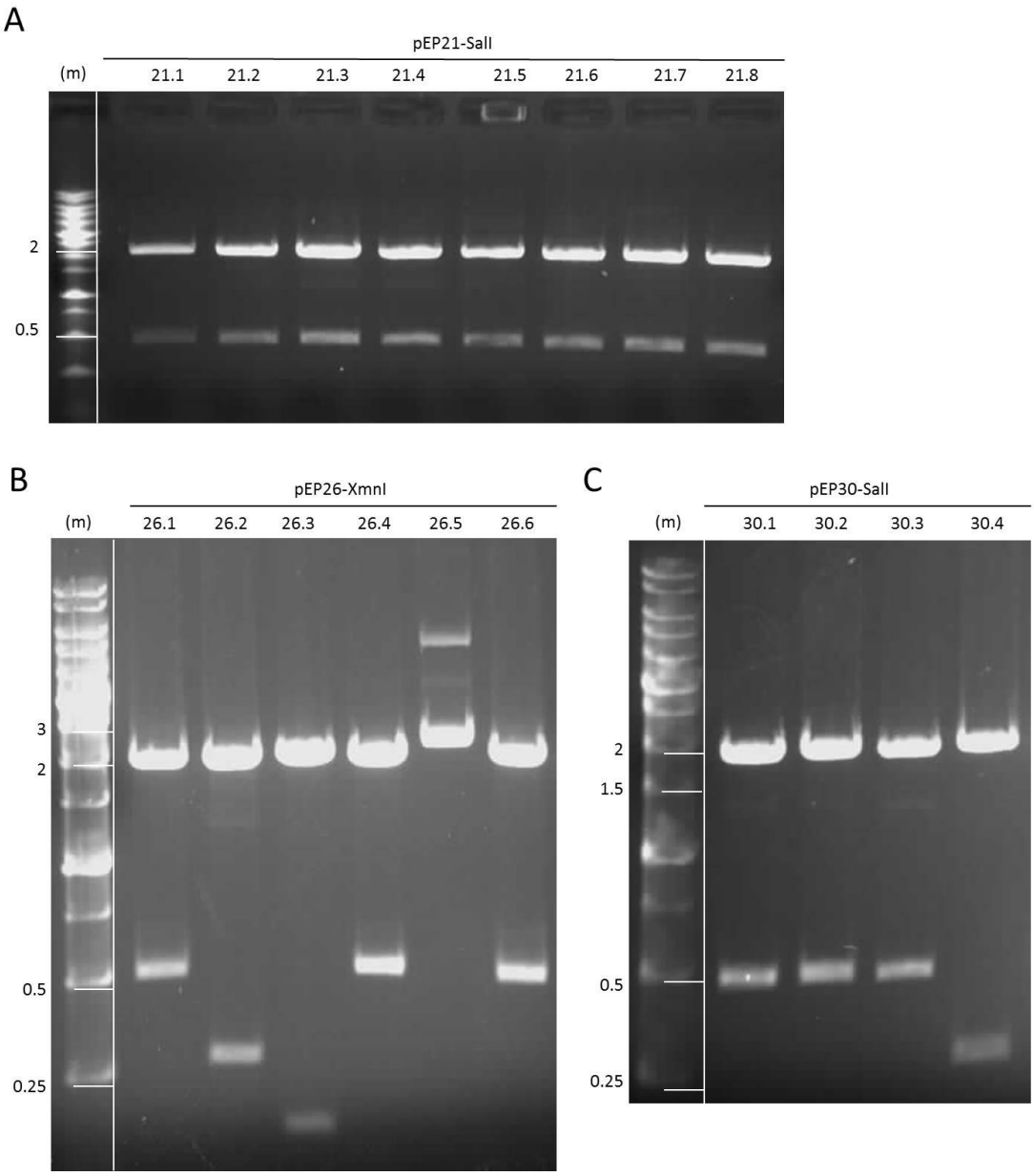

Figure 20| Restriction fragment analysis of pEP21, pEP26 and pEP30. Agarose gel (0.8%) electrophoresis with ethidium bromide staining. **(A)** SalI digestion of pEP21. Expected bands: 435 and 1853 bp. **(B)** XmnI digestion of pEP26. Expected bands: 509 and 1910 bp. **(C)** SalI digestion of pEP30. Expected bands: 493 and 1853 bp. Molecular size marker, m= 1 kb DNA ladder.

After incorporating the specific phiC31 attachment site and right-ITR elements in the vectors, the next step consisted in the insertion of the docking modules. One docking module comprises four docking sites, each one consisting of a cHS4 insulator, the SV40

promoter and one Bxb1-specific attB site. A core docking module was synthetized by Genewiz (USA) (≈1.6 kb length) (Figure 21 A) and then used as a template for PCR synthesis of four docking elements with different but compatible restrictions sites for Golden Gate assembling (BamHI-BsmBI(A), BsmBI(A)-BsmBI(B), BsmBI(B)-BsmBI(C) and BsmBI(C)-AscI (see material and method, section 2.1.3) (Figure 21 A). BsmBI is a Type IIS enzyme used in Golden Gate cloning to assemble multiple modules in a single reaction (333). Type IIS restriction enzymes bind to their recognition sites but cut the DNA downstream from that site at a positional, not sequence-specific site. Thus, a single Type IIS restriction enzyme can be used to generate docking units with unique but compatible overhangs. PCR products from the four reactions were cloned into the pGEM-Teasy vector and sequenced (data not shown) to confirm their identities. pGEM-Teasy based plasmids containing the different PCR products were digested with the following enzymes: docking element 1 with BamHI and BsmBI; docking element 2 and 3 with BsmBI and docking element 4 with BsmBI and AscI (Figure 21 B). The isolated PCR fragments were assembled using their complementary overhangs to create a complete docking module with four docking elements each (Figure 21 B-C).

A second ligation reaction was then set to clone the full docking module into the BglII/AscI sites of pEP21, pEP26 and pEP30. Three new plasmids were recovered: pEP22, pEP27 and pEP31 (Figure 21 C). Beside the docking module pEP22 carried a phiC31-specific attB site and a SB right-ITR element, pEP37 a phiC31-specific attP site and a PB right-ITR element, and pEP31 a phiC31-specific attR site and a SB right-ITR element (Figure 21 C).

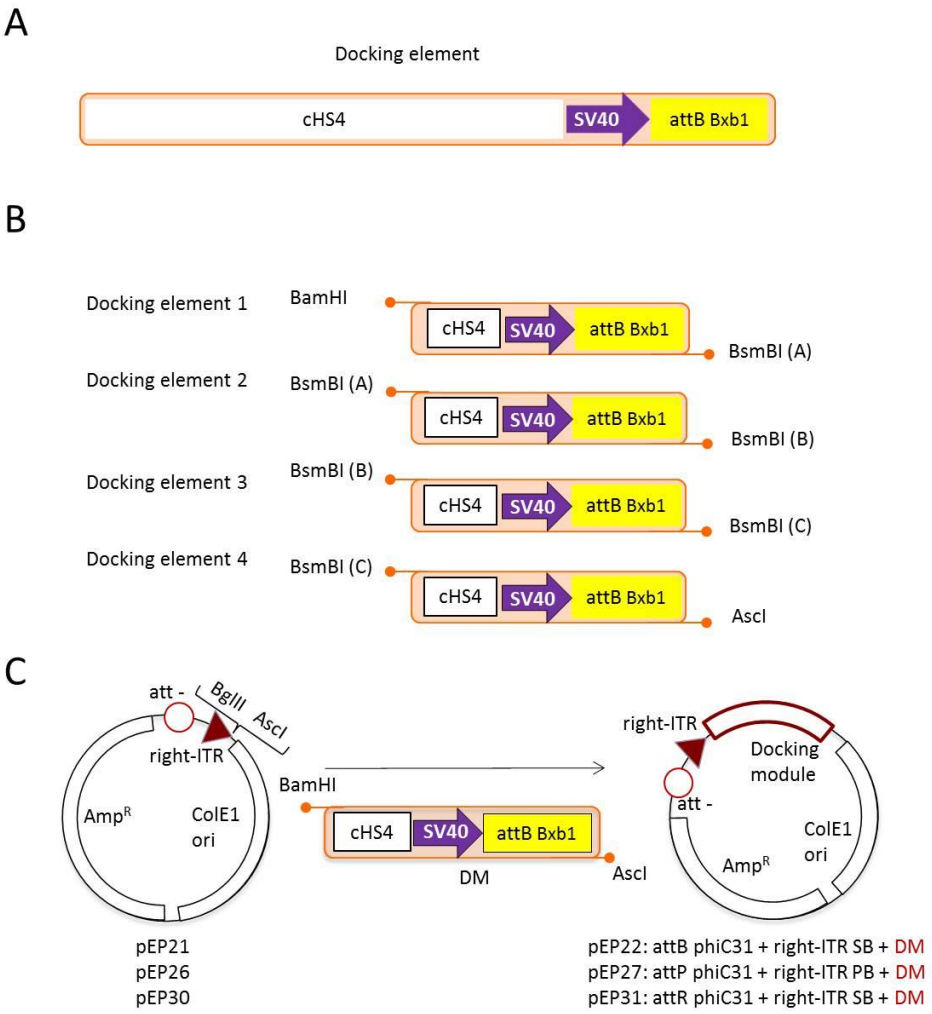

Figure 21| Docking module assembly. (A) Each Docking element consists of one cHS4 insulator, the SV40 promoter and a Bxb1-specific attB site. **(B)** PCR amplification of the docking element with four different sets of primers containing specific Type IIS restriction endonuclease sites at their overhangs. **(C)** Four docking elements assembled into a single docking module (DM) were cloned into pEP21, pEP26 and pEP30. The resulting plasmids, pEP22, pEP27 and pEP31 contain a phiC31 attachment site, a transposon right-ITR element and docking module. White rectangle: cHS4 insulator. Purple arrow: SV40 promoter. Yellow rectangle: Bxb1-specific attB site. Bordeaux circle: phiC31-specific attP, attB and attR sites. Bordeaux triangle: transposon-specific ITR elements. Bordeaux rectangle: docking module.

Twelve clones from pEP22 were picked and analysed by AgeI restriction. Only clone 22.9 was positive for the expected band of 9047 bp (Figure 22 A). Twenty clones from pEP27 and pEP31 plasmids were analysed. In a first step, plasmids were digested with

NheI (data not shown), and then, all possible positives clones digested again with EcoRI. Restriction pattern analysis confirmed the presence of two positive clones for pEP27: 27.6 and 27.18 (Figure 22 B-left), and one positive clone for pEP31: 31.17 (Figure 22 B-right).

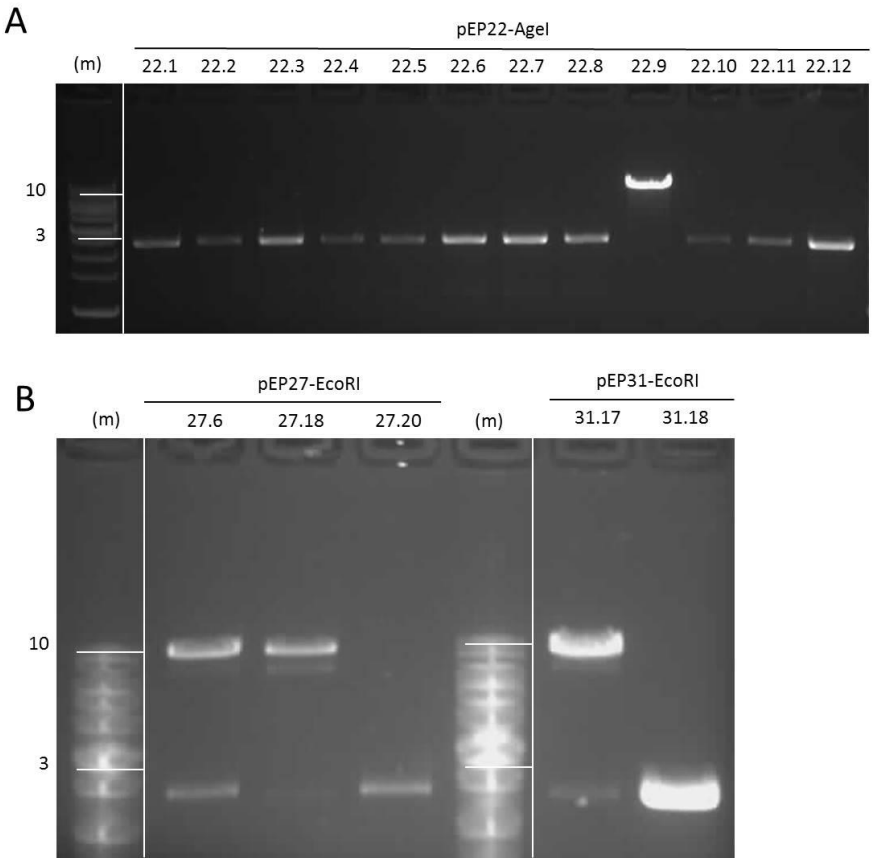

Figure 22| **Restriction fragment analysis of pEP22, pEP27 and pEP31.** Agarose gel (0.8%) electrophoresis with ethidium bromide staining. **(A)** AgeI digestion of pEP22. Expected band: 9047 bp. **(B)** EcoRI digestion of pEP27 and pEP31. Expected bands 9178 and 9105 bp, respectively. Molecular size marker, m= 1 kb DNA ladder.

The next cloning step consisted in the incorporation of the selection marker (SM) flanked by transposon elements. From this point on, the plasmid generated from pEP22 (pEP23) followed a different cloning strategy than the plasmids generated from pEP27 (pEP28) and pEP31 (pEP50). pEP23 carries the SM flanked by two ITR elements,

whereas pEP28 and pEP50 contains only one ITR located 5' to the SM. The generation of pEP28 and pEP50 is described in page 76 and page 77, respectively.

To construct pEP23, the puroΔtk cassette, flanked by ITR elements of PB, was isolated from pEP33 by NheI/AscI digestion and cloned into pEP22 previously linearized with the same enzymes (Figure 23 A). The resulting pEP23, included a phiC31-specific attB site, the right-ITR element of SB, the docking module and puroΔtk preceded the by ITR element of PB. One (23.7) out of 12 clones analyzed showed the expected AgeI restriction pattern of 4653 and 7603 bp (Figure 23 B).

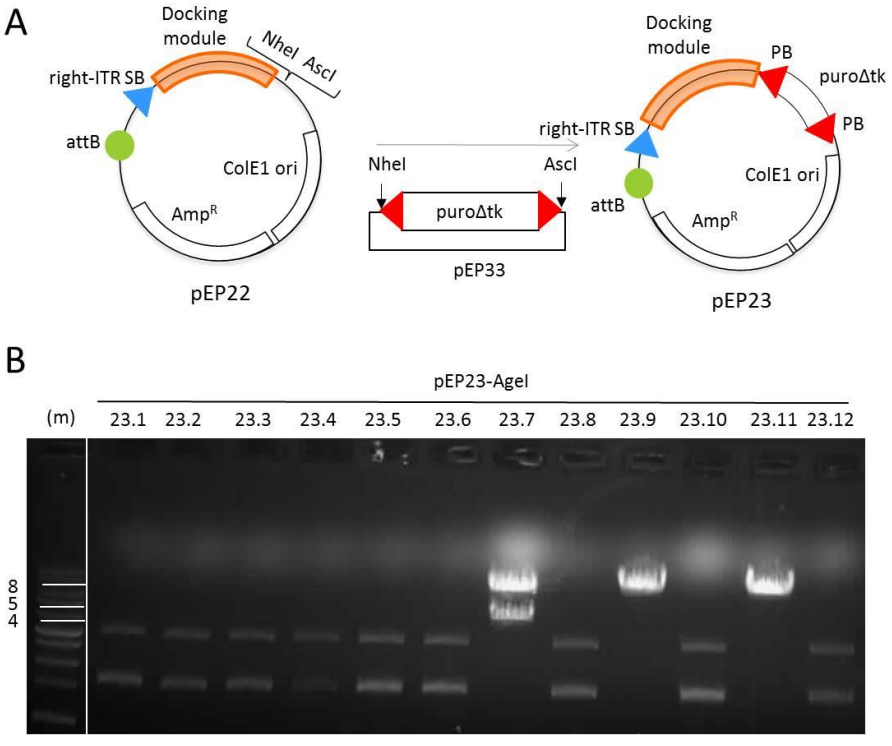

Figure 23| **Construction of vector pEP23. (A)** Diagram of the cloning of the selection marker (puroΔtk), flanked by PB ITR elements, into pEP22. **(B)** AgeI digestion of pEP23. Expected bands: 4563 and 7603 bp. Agarose gel (0.8%) electrophoresis with ethidium bromide staining. Molecular size marker, m= 1 kb DNA ladder. Light green rectangle: phiC31-specific attB sites. Blue triangles: SB-specific ITR elements. Red triangles: PB-specific ITR elements. Orange rectangle: docking module.

The last step to generate the homologous recombination vector, pEP24 (Figure 17), consisted in the cloning of the CCR5 5' and 3' recombination arms. To this intend, the pEP34 plasmid was constructed (Figure 24). The exon 3 region of the human CCR5 gene was amplified using the 480up/down specific pair of primers (see material and methods, section 2.1.3) using as template DNA from the IMEDEAi003-A hiPS cell line (326). This newly synthesized PCR product served as a template for a second round of PCR reactions in which the CCR5-5' and 3' recombination arms were amplified (Figure 24 A). Both final PCR products were sequenced to confirm the correctness of DNA sequence (data not shown). Finally, double-digested SbfI/MluI 3' recombination arm and the NotI/BsmBI 5' recombination arm were sequentially cloned into the linearized pKOscramblerV920 (Figure 24 B).

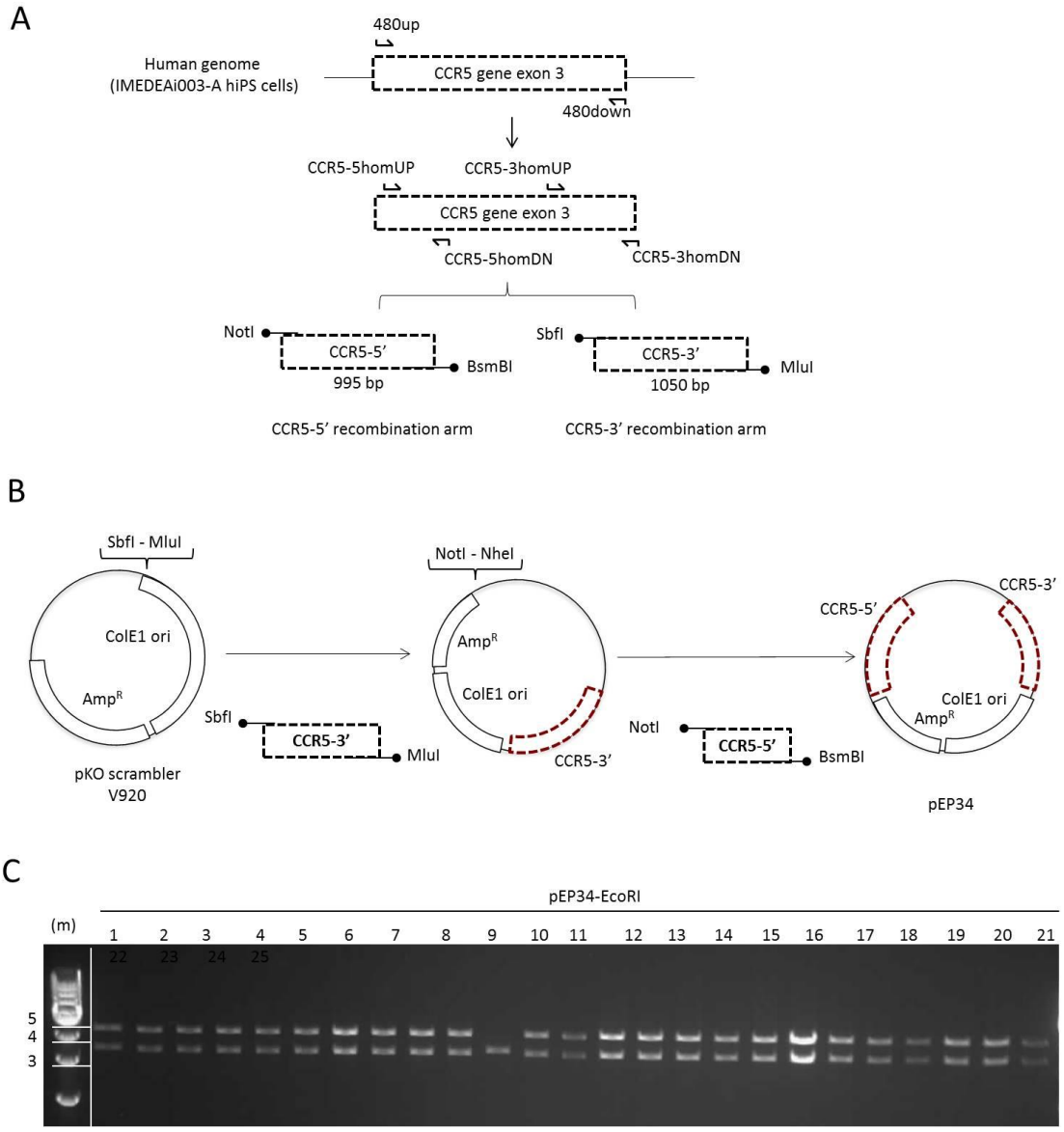

Figure 24| **Construction of vector pEP34. (A)** Amplified exon 3 region of CCR5 gene using the 480up/down primers served as template for the generation of 5' and 3' recombination arm by a second PCR amplification. **(B)** Synthesis of pEP34 by cloning the CCR5-5' and CCR5-3' recombination arms into the plasmid pKOscramblerV920. **(C)** EcoRI digestion of pEP34. Expected bands: 3652 and 4832 bp. Agarose gel (0.8%) electrophoresis with ethidium bromide staining. Molecular size marker, m= 1 kb DNA ladder.

Restriction analysis with EcoRI demonstrated the correct restriction pattern (3652 and 4832 bp) in all clones, except clone 11 (Figure 24 C).

To finish the construction of pEP24, the docking cassette from vector pEP23 (attB phiC31 site, right-ITR element of SB, docking module and puroΔtk selection marker flanked by ITR elements of PB) was cloned into the AatII/PacI restriction sites of pEP34. The final plasmid, pEP24, carried the docking cassette flanked by CCR5 5' and 3' recombination arms (Figure 25 A). Digestion of pEP24 with XhoI revealed two positive clones (24.2 and 24.8) with the correct restriction pattern (3313 and 10897 bp) (Figure 25 C).

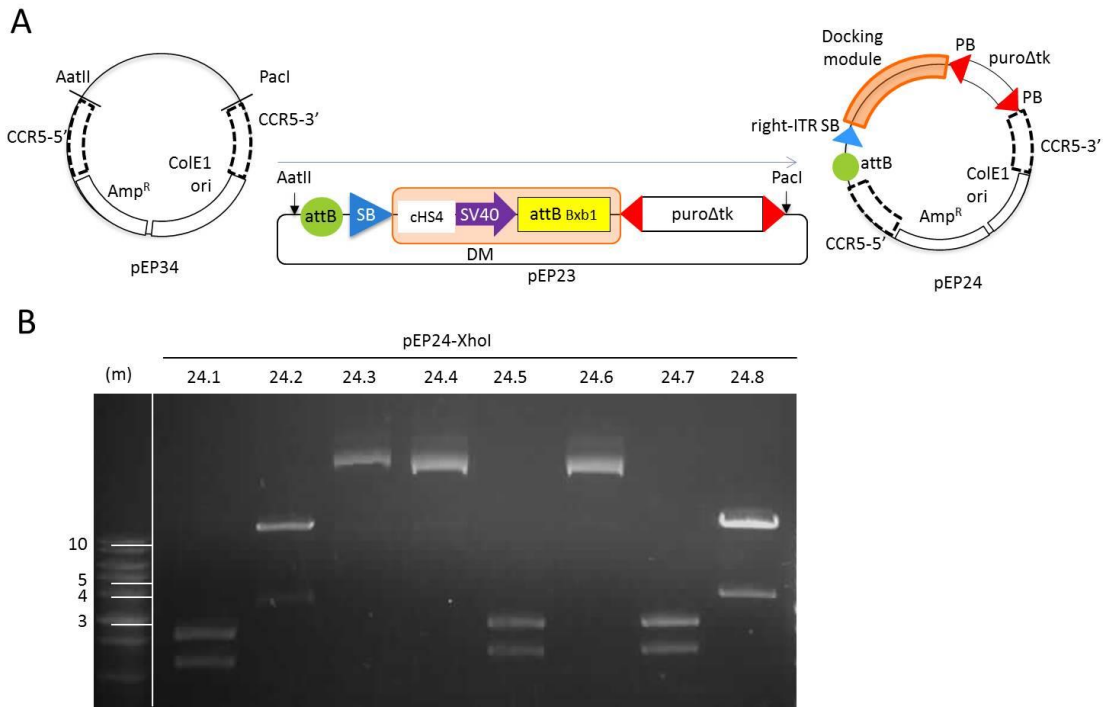

Figure 25| **Final step in the construction of the recombination vector pEP24. (A)** Diagram of the cloning of phiC31-specific attB site, the right-ITR element of SB, a docking module consisting of cHS4 insulator, the SV40 promoter and a Bxb1-specific attB site, as well as a puroΔtk selection marker flanked by ITR elements of PB from pEP23 into pEP34 harboring the two CCR5 recombination arms. **(B)** XhoI digestion of pEP24. Expected bands: 3313 and 10897 bp. Agarose gel (0.8%) electrophoresis with ethidium bromide staining. Molecular size marker, m= 1 kb DNA ladder. Light green rectangle: phiC31-specific attB sites. Blue triangles: SB-specific ITR elements. Red triangles: PB-specific ITR elements. Orange rectangle: docking module. White rectangle: cHS4 insulator. Purple arrow: SV40 promoter. Yellow rectangle: Bxb1-specific attB site.

The last step in the generation of the second recombination vector, pEP28, consisted in the cloning of the SM and the left-ITR element of SB from pEP35 into the AscI/NotI sites of pEP27 (Figure 26 A). The final plasmid, pEP28, consisted of a phiC31-specific attP site, the right-ITR element of PB, the docking module and the puroΔtk cassette preceded by the left-ITR element of PB (Figure 26 A). Restriction analysis by SalI confirmed the correct structure of pEP28 in nine out of twelve clones analyzed (Figure 26 B).

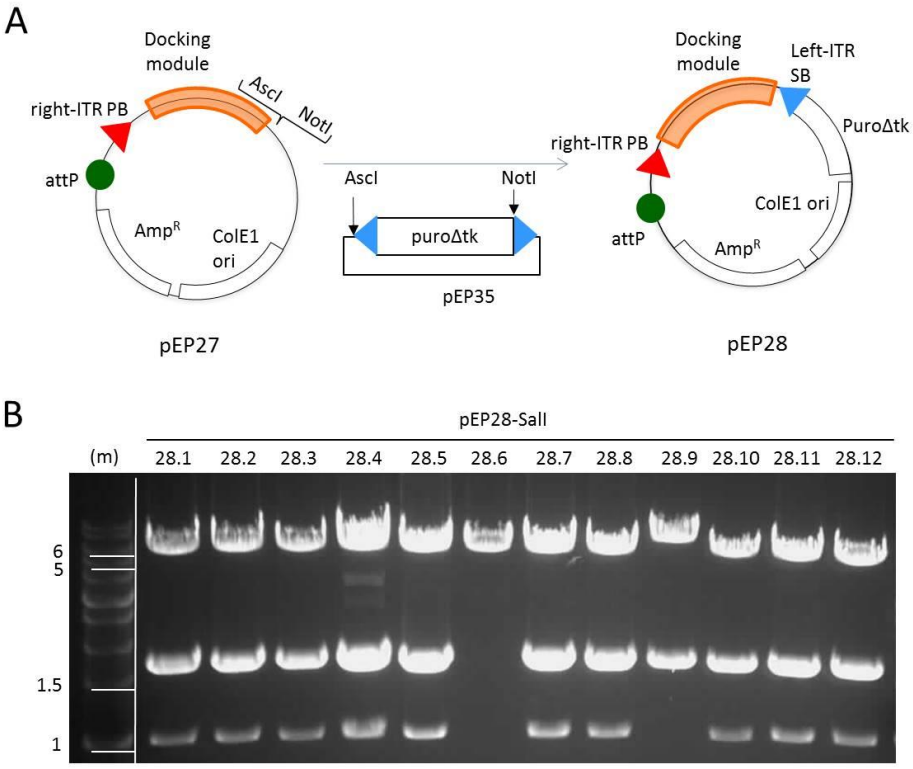

Figure 26| Final step in the construction of the recombination vector pEP28. (A) Diagram of the cloning of puroΔtk and the left-ITR element of SB into pEP27. **(B)** SalI digestion of pEP28 with expected bands: 991, triple band of 1693 and 6777 bp. Agarose gel (0.8%) electrophoresis with ethidium bromide staining. Molecular size marker, m= 1kb DNA ladder. Dark green circles: phiC31-specific attP sites. Blue triangles: SB-specific ITR elements. Red triangles: PB-specific ITR elements. Orange rectangle: docking module.

Finally, the construction of the third recombination vector, pEP50, was performed by cloning the SM (puroΔtk) with the left-ITR element of PB into the pEP31 vector previously linearized by AscI/NotI digestion. pEP50 contained one phiC31-specific attR site, the right-ITR element of PB, the docking module and the puroΔtk with the left-ITR element of SB at its 5; side (Figure 27 A). After digestion with KpnI, four correctly assembled clones showed the expected restriction pattern; of 1515, 1693 and 5276 bp (Figure 27 B).

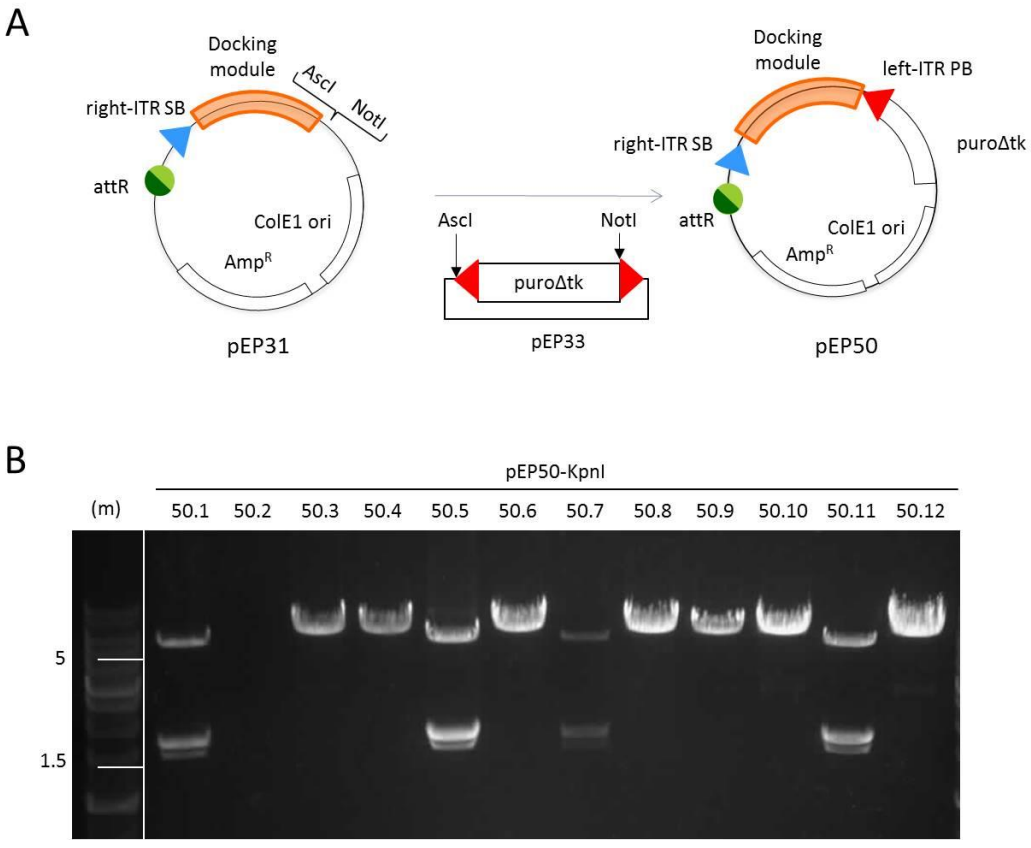

Figure 27| Generation of the third docking vector, pEP50. **(A)** Diagram of the cloning of puroΔtk with the left-ITR element of PB into pEP31. pEP50 consists of the phiC31-specific attR site, the right-ITR element of SB, a docking module and (puroΔtk) with left-ITR element of PB. (B) KpnI digestion of pEP50. Expected bands: 1515, triple band of 1693 and 5276 bp. Agarose gel (0.8%) electrophoresis with ethidium bromide staining. Molecular size marker, m= 1 kb DNA ladder. Dark/light green circles: phiC31-specific attR sites. Blue triangles: SB-specific ITR elements. Red triangles: PB-specific ITR elements. Orange rectangle: docking module.

3.2.2. Generation of the test plasmids for phiC31 & Bxb1 recombinase assay

To validate the functionality of the docking platform assembled at the CCR5 locus, a promoterless loading vector with a fluorescent reporter gene was synthetized. For this purpose, a Bxb1-specific attP attachment site was first generated by annealing of two previously synthesized oligonucleotides (Thermo Fisher Scientific, USA) and inserted via PciI/NheI restriction sites into the pmCherryC1 vector replacing its original CMV promoter (Figure 28 A).

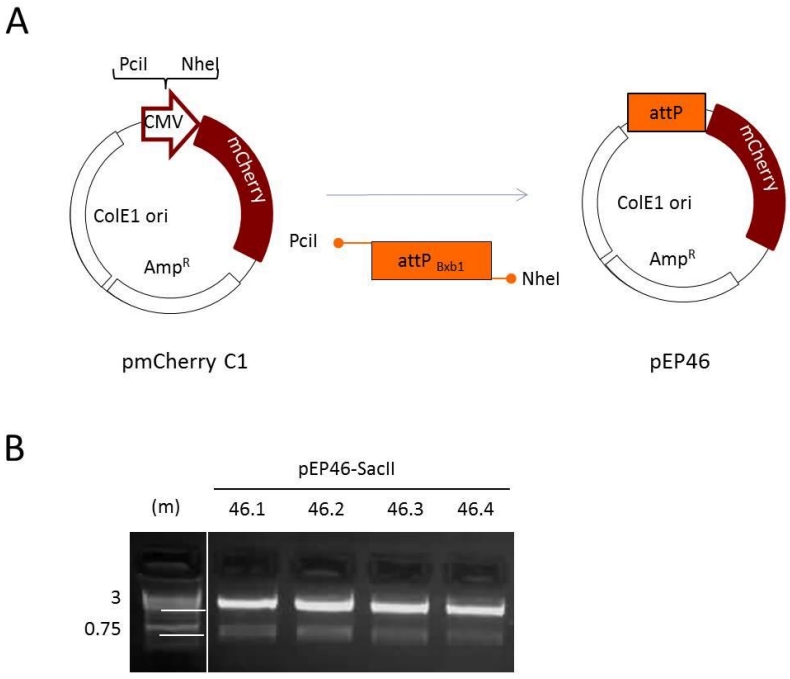

Figure 28| Generation of the testing vector pEP46. (A) Diagram of the removal of the CMV promoter and the insertion of one Bxb1-specific attP site 5' of the mCherry cDNA. **(B)** Agarose gel (0.8%) electrophoresis with ethidium bromide staining. SacII digestion. Expected bands: 816 and 3324 bp. Molecular size marker, m= 1 kb DNA ladder. Bordeaux rectangle: mCherry fluorescence. Orange rectangle: Bxb1-specific attP sites.

To confirm the correct configuration of the plasmid, pEP46 clones were digested with SacII. All clones analysed showed the expected restriction pattern (Figure 28 B).

3.3. Assembly of the docking platform in the human HeLa cell line

As a proof of concept study, the docking platform was first assembled in a basic cellular model, the human HeLa cell line. HeLa cells are of human origin, easy to transfect and have a wide range of applications in biomedical studies. Assembly of the complete docking platform into the human CCR5 locus required sequential loading of the cargos of the three recombination vectors, pEP24, pEP28 and pEP50. The pEP24 targeting vector was integrated by TALENs-assisted HR. pEP28 and pEP50 upload was mediated by phiC31 recombinase. To confirm the reliability of the work in progress, diagnostic components were sequenced after each recombination step.

The first step in the assembly process was initiated with the lipofection of the recombination vector, pEP24, and a pair of CCR5-specific TALENs into HeLa cells. Following puromycin selection, 52 clones were isolated and expanded. The 5' and 3' recombination sites were analysed by PCR (Figure 29 A). Thirteen clones showed correct integration patterns of the 5' end of the docking cassette (Figure 29 B), but only eight of the thirteen showed also correct 3' integration (Figure 29 C). TALENs-assisted HR showed a frequency of 34 (Table 4). Finally, clone 24 was chosen for further experimentation based on its growth and sensibility to FIAU.

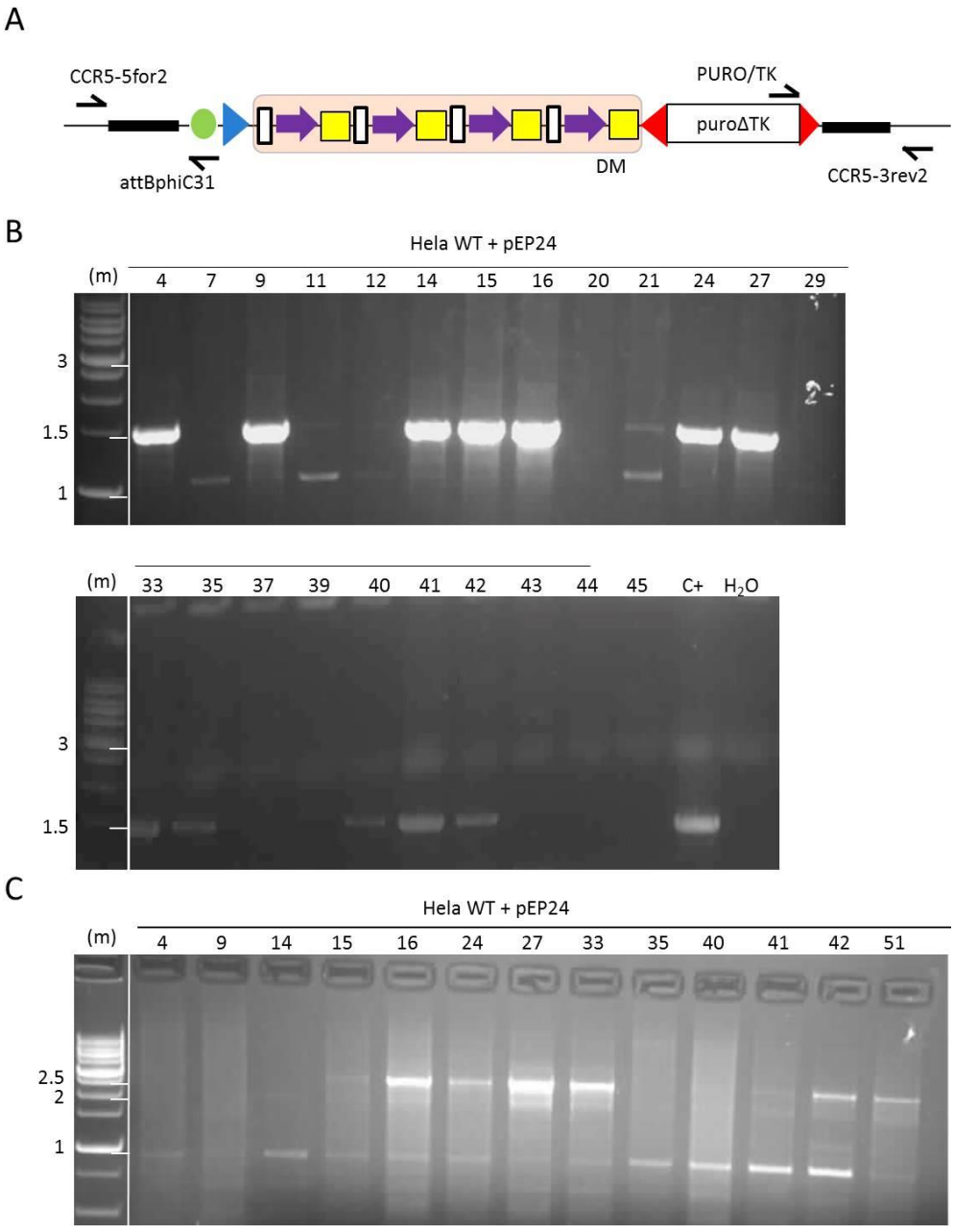

Figure 29| Example of screening by PCR of the clones obtained by recombination at step 1. **(A)** Diagram of the PCR characterization. **(B)** 5' PCR analysis using CCR5-5for2 and attBphiC31 primers (1500 bp). Postive and negative controls (water) are shown at the right end of photo . **(C)** 3' PCR analysis using PURO/TK and CCR5-3rev2 primers (2500 bp). Agarose gel (0.8%) electrophoresis with ethidium bromide staining. Molecular size marker, m= 1 kb DNA ladder.

Following the integration of the first docking module into the human CCR5 locus, the puroΔtk selection csssete was exceised in the HeLa Cl.24 clone by transfection with PB transposase followed by FIAU selection. 48 HeLa clones survived the negative selection process. Successful excision of puroΔtk was confirmed in 6 out of 34 clones analyzed by PCR screening (Table 4, Figure 30). The efficiency of the excision reaction was 17% (Table 4). After sequencing (Figure 31), the HeLa Cl.24.22 clone was finally chosen for continuing the construction up of the docking platform.

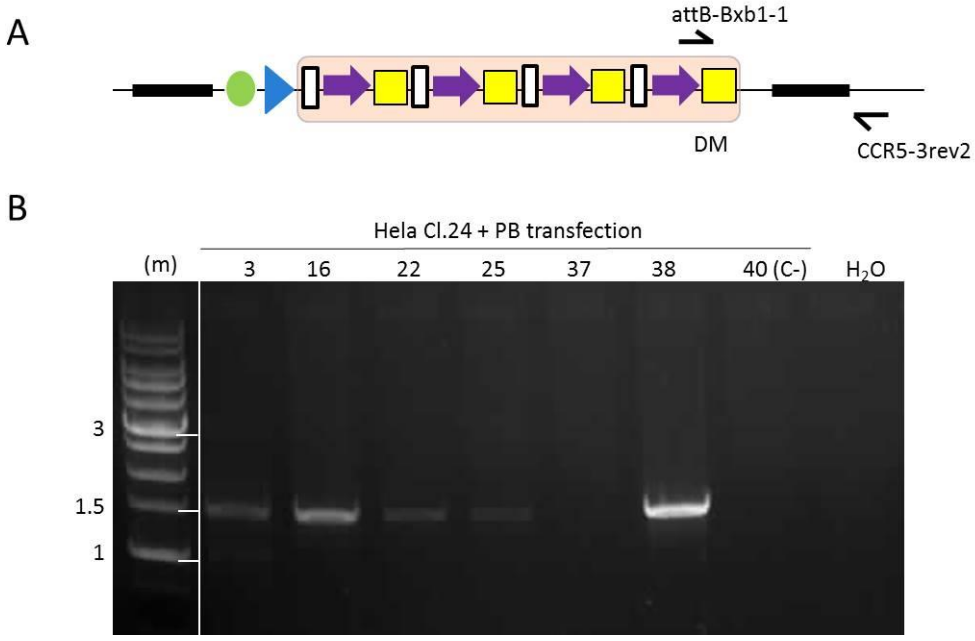

Figure 30| Excision of the selection cassette by PB transposase in HeLa Cl.24. (A) Diagram of the PCR characterization. **(B).** Example of 7 clones analyzed by PCR using the primers attB-Bxb1-1/CCR5-3rev2 (1500 bp), Cl.24.40 and Cl.24.37 were negative. Agarose gel (0.8%) electrophoresis with ethidium bromide staining. Molecular size marker, m= 1 kb DNA ladder.

CLUSTAL O (1.2.4) multiple sequence alignment

```
attBphiC31        --ACCGGTGTCCCAGACGTCTGGGTCGACGATGTAGGTCACGGTCTCGAAGCCGCGGTGC
HeLaCl24.22       GNACCGGTGTCCCAGACGTCTGGGTCGACGATGTAGGTCACGGTCTCGAAGCCGCGGTGC

attBphiC31        GGGTGCCAGGGCGTGCCCTTGGGCTCCCCGGGCGCGTACTCCACCTCACCCATCTGGTCC
HeLaCl24.22       GGGTGCCAGGGCGTGCCCTTGGGCTCCCCGGGCGCGTACTCCACCTCACCCATCTGGTCC

attBphiC31        ATCATGATGGTACCGTCTGG
HeLaCl24.22       ATCATGATGGTACCGTATTG
```

Figure 31| HeLa Cl.24.22 sequence analysis. Theoretical phiC31-specific attB sequence aligned to the Cl.24.22 sequence. In red and bold the 50 bp sequence required for the recombination by phiC31 integrase.

The phiC31-specific attB site is crucial for the next step of platform assembly. In order to integrate a new docking module into the platform, HeLa Cl.24.22 cells, were transfected with a phiC31 integrase expressing plasmid and the second recombination vector, pEP28, that harbors a phiC31-specific attP site. Sucessful attB/attP recombination catalyzed by the phiC31 integrase was confirmed in 4 out of the 24 puromycin resistant clones analyzed. The test was conducted by screening the 3' (Figure 32 B) and 5' (Figure 32 C and D) recombination sites by PCR. Moreover, PCR analysis using primers specific for the 5' of the docking platform in step 1, confirmed the disappearance of the PCR amplicon in the four candidate clones as result of the site-specific recombination event in step 2 (Figure 32 D). Integration efficiency was 16% (Table 4).

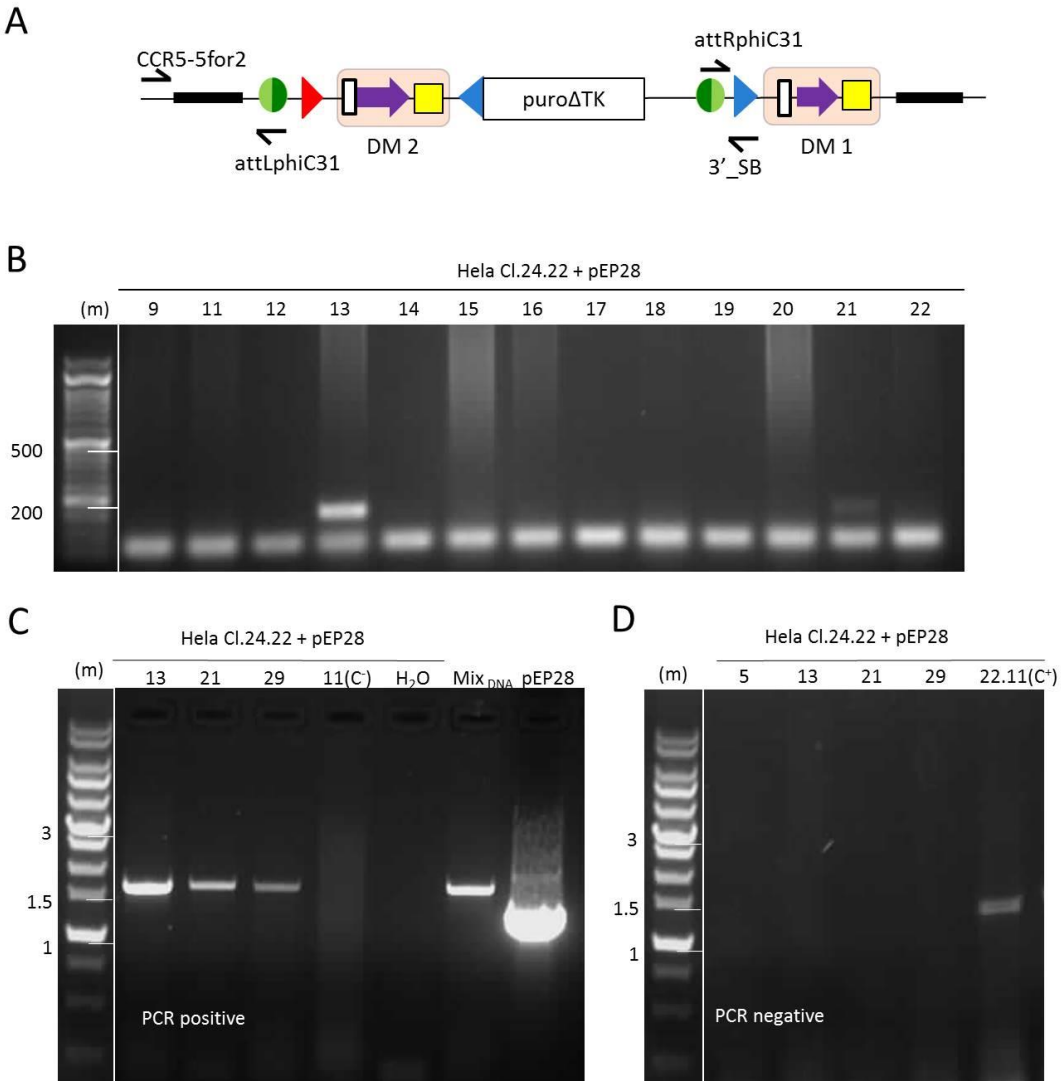

Figure 32| Example of screening by PCR of the clones obtained by recombination at step 2. (A) Diagram of PCR characterization. **(B)** PCR analysis of the 3' recombination arm using attRphiC31/3'_SB primers (amplicon:173 pb). Molecular size marker, m= 50 bp DNA ladder. **(C)** PCR analysis of the 5'recombination arm using CCR5-5for2/attLphiC31 primers (amplicon: 1538 bp). **(D)** PCR analysis using CCR5-5for2/attBphiC31 primers to confirm the disappearance of the 1538 bp amplicon previously detected in step 1 of platform assembly. Agarose gel (2%) electrophoresis with ethidium bromide staining. Molecular size marker, m= 1 kb DNA ladder.

After puromycin selection, the puroΔtk cassette, remaining vector elements and selection cassette in the HeLa Cl.24.22.21 and HeLa Cl.24.22.29 lines were removed by

transfection with SB transposase and exposition to FIAU selection. Successful SE excision was demonstrated in three subclones of HeLa 24.22.21 (24.22.21.3, 24.22.21.7 and 24.22.21.18) and one subclone of HeLa 24.22.29 (24.22.29.7) cells (Figure 33) by lack of PCR amplicon (173 bp) (Figure 32 A) because the primers cannot bind to the template. To confirm the PCR result cells clones were shown to be sensible to puromycin, but resistant to FIAU treatment (Figure 34). Finally, HeLa Cl.24.22.21.3 was selected for completion of the docking platform assembly since it displayed the best growth properties.

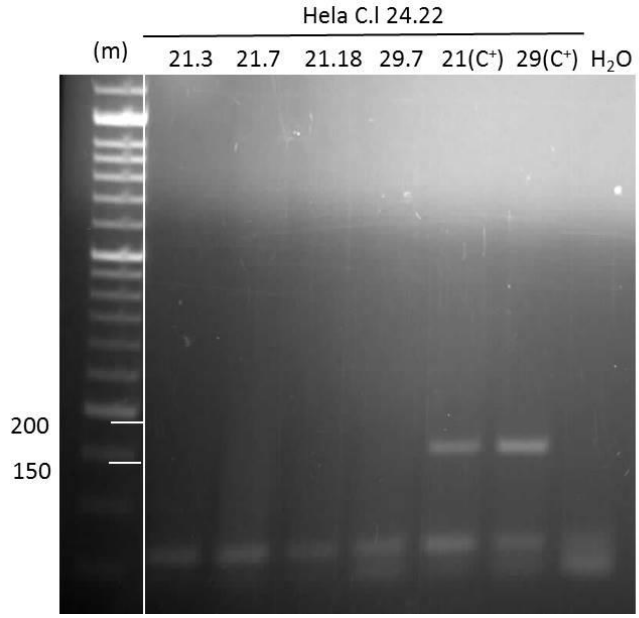

Figure 33| Excision of the selection cassette by SB transposase in HeLa Cl.24.22.21 and Cl.24.22.29 clones Agarose gel (2%) electrophoresis with ethidium bromide staining. PCR analysis using attRphiC31/3'_SB primers (173 bp) demostrated the loss of PCR product after SE excision. As positive control, DNA from HeLa Cl.24.22.21 (21C$^+$) and HeLa Cl.24.22.29 (29 C$^+$) without transfection were used. Molecular size marker, m= 50 bp DNA ladder.

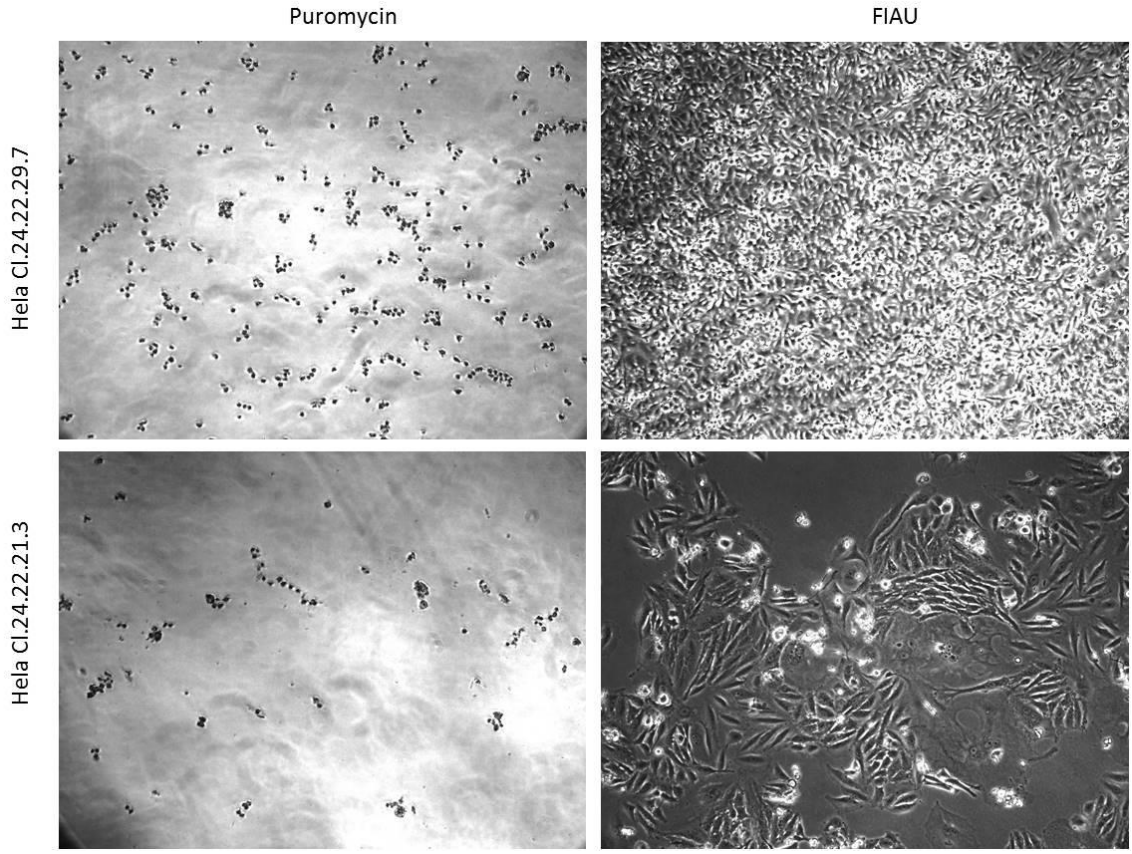

Figure 34| Drug sensibility of HeLa Cl.24.22.21.3 and Cl.24.22.29.7 cells. HeLa Cl24.22.21.3 and Cl24.22.29.7 were exposed to puromycin [2μg/ml] and FIAU [1μM] selection during 6 days to demonstrate the excision of selection elements.

After site-specific recombination between the attB and attP sites occurred in step 2, two novel phiC31-specific recombination sites, attL and attR, were generated in HeLa Cl.24.22.21.3 following transfection of the SB transposase. Subsequently, the integrity of the attL phiC31 site was confirmed by sequencing (Figure 35). attR phiC31 site was excised by SB transposase, therefore there was not necessary to confirm the integrity.

CLUSTAL O (1.2.4) multiple sequence alignment

```
attLphiC31-------GGTGCGGGTGCCAGGGCGTGCCCTTGAGTTCTCTCAGTTGGGGGCGTAGGGTCGCCGAC
HeLacl24.22.21.3-GGTGCGGGTGCCACGGCGTGCCCTTGAGTTCTCTCAGTTGGGGGCGTAGGGCCTCCGAC

attLphiC31-------ATGACACAAGGGGTTGTGACCGGGGTGGACACGTACGCGGGTGCTTACGACCGTCAGTG
HeLaCl24.22.21.3-ATGACACAAGGGGTTGTGACCGGGGTGGACACGTACGCGGGTGCTTACGACCGTCAGTG
```

Figure 35| HeLa Cl.24.22.21.3 sequence analysis. Theoretical attL phiC31 sequence aligned to Cl.24.22.21.3 sequence. In red and bold the 42 bp recognition sequence required for recombination.

Efficient recombination between attL and attR sites, upon which the last recombination step is based, requires the participation of gp3 along with the phiC31 integrase. HeLa Cl.24.22.21.3 cells were transfected with an expression vector containing the fusion protein gp3-phiC31 (pCS_KRI) and the third recombination vector, pEP50 (Figure 17). Following attL/attR recombination, attP and attB sites were regenerated and the third docking module was incorporated into the genome, resulting in a docking platform of 12 docking sites. Following puromycin selection, 15 out of 29 clones analyzed showed successful restauration of the attB phiC31 site (Figure 36 B). The recombination event had an efficiency of 48% (Table 4). In addition to puromycin resistance HeLa clones were sensible to FIAU selection, further demonstrating the integrity of the puroΔtk cassette. The 4 FIAU sensitive clones that showed faster respond to FIAU selection were chosen for confirmation by PCR. Successful attR/attL recombination catalyzed by the phiC31 integrase and gp3 was confirmed by screening the 5' (Figure 36 B and C) and 3' (Figure 36 D).To verify the correct integration in 5', two different PCRs were performed. The first PCR was designed to detect the regenerated attB site (Figure 36 A). Since, pEP50 contains a right-ITR element of SB, which was a unique component of the platform at that stage a second specific PCR test aimed at detecting it was possible (Figure 36 A). 3 out of 4 clones re-confirmed both attB regeneration and the presence of the right-ITR element of SB (Figure 36 B). HeLa Cl.24.22.21.3.65 cells did not show any PCR amplicons in any

PCR analyses, but in contrast a regenerated attB site and the right-ITR element of SB were detected in HeLa Cl.24.22.21.3.98 and Cl.24.22.21.3.145 clones (Figure 36 C). HeLa Cl.24.22.21.3.129, showed attB regeneration but absence of the right-SB-ITR element, and was re-analyzed using a bigger amount of template DNA. The second test confirmed that is was also positive for the ITR (Figure 36 C). Representation of 5' positive clones were also positive to 3' PCR. HeLa Cl.24.22.21.3.65 shows weak band due to less DNA amount (Figure 36 D).

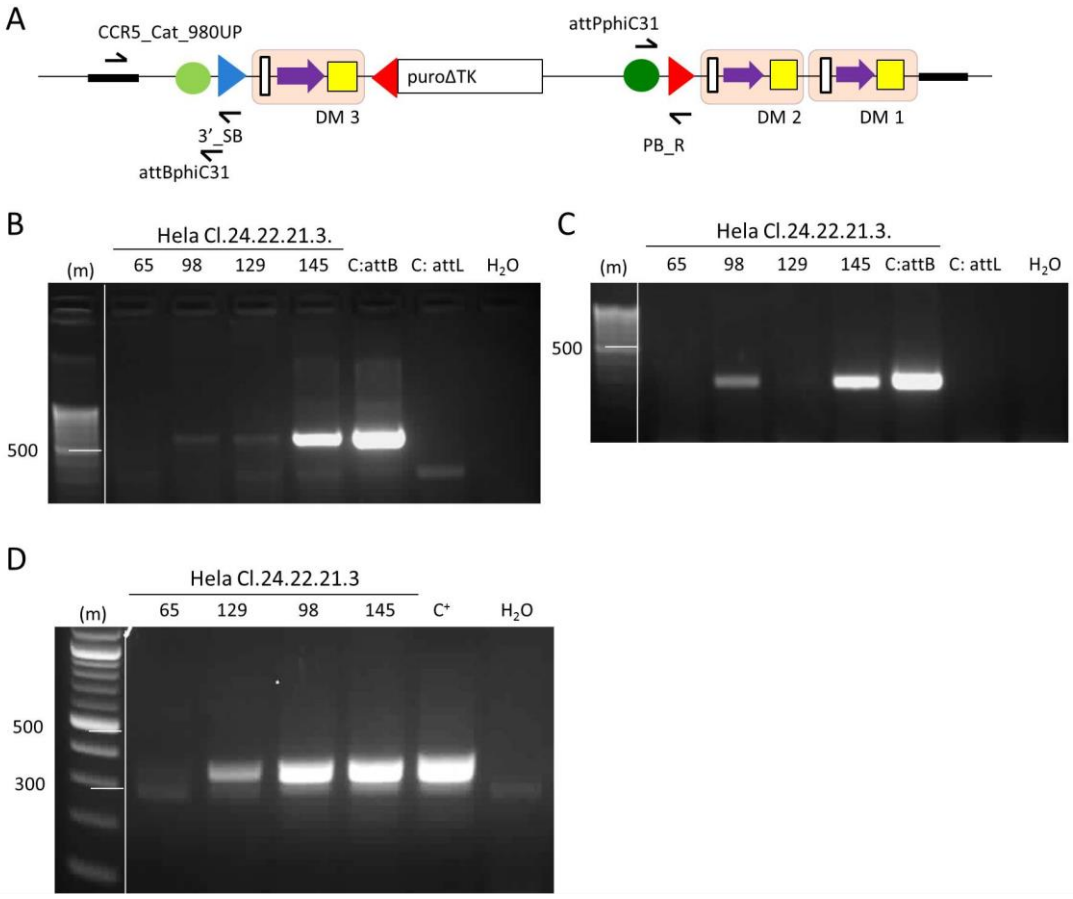

Figure 36| PCR screening of the clones obtained by recombination at step 3. (A) Diagram of the PCR characterization. **(B)** phiC31-specific attB PCR screening with CCR5_Cat_980UP/attBphiC31 primers (amplicon: 581 bp). **(C)** Right-SB-ITR PCR screening with attxphiC31/SB_R (amplicon: 296 bp). **(D)** phiC31-specific attR screening with attPphiC31/PB_R primers (amplicon: 343 bp). Agarose gel (2%) electrophoresis with ethidium bromide staining**.** Molecular size marker, m= 50 bp DNA ladder.

Although deletions were detected at the 3' end of the attB site, the integrity of the core sequence (red bold), crucial for site-specific recombination, was confirmed (Figure 37).

CLUSTAL O (1.2.4) multiple sequence alignment

```
attBphiC31-------------GTCGACGATGTAGGTCACGGTCTCGAAGCCGCGGTGCGGGTGCCAGGGCG
HeLaCl.24.22.21.3.98---GTCGACGATGTAGGTCACGGTCTCGAAGCCGCGGTGCGGGTGCCAGGGCG

attBphiC31-------------TGCCCTTGGGCTCCCCGGGCGCGTACTCCACCTCACCCATCTGGTCCATC
HeLaCl.24.22.21.3.98---TGCCCTTGGGCTCCCCGGGCGCGTACTCCACCTCACCCATCTGGTCCATC

attBphiC31-------------ATGATGGTACCGTCTGGTCCATC
HeLaCl.24.22.21.3.98---*****GGTACCGT****
```

Figure 37| HeLa Cl.24.22.21.3.98 sequence analysis. Theoretical attB phiC31 sequence aligned to Cl.24.22.21.3.98. sequence. In red and bold the 50 bp recognition sequence required for recombination. (*): deleted bases.

To complete the construction of the platform, the SE was excised from the HeLa Cl.24.22.21.3.98 and Cl.24.22.21.3.145 clones by transfection with PB integrase. Successful SE excision was demostrated by lack of PCR amplicon (343 bp) (Figure 36 A) because the primers cannot bind to the template. A total of 52 clones were resistant. PCR analysis demostrated the absence of SE in seven and two FIAU-resistant clones, respectively (Table 4, Figure 38). To confirm PCR result, clones with excised SE were cultured in the presence of puromycin or FIAU, showing sensibility to puromycin and resistance to FIAU.

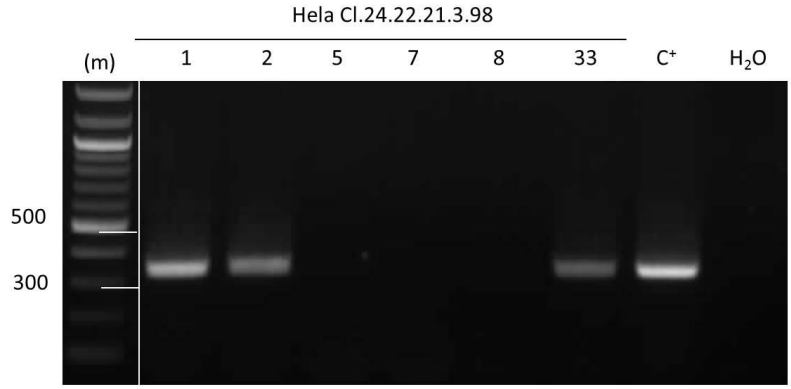

Figure 38| Excision of the selection cassette by PB transposase in HeLa Cl. 24.22.21.3.98. Agarose gel (2%) electrophoresis with ethidium bromide staining. PCR analysis using attPphiC31/PB_R primers (amplicon: 343 bp) demostrated the loss of PCR product after SE excision. As positive control, DNA from HeLa Cl.24.22.21.3.145 (C^+) without transfection. Molecular size marker, m= 50 bp DNA ladder.

Overall, the data presented so far demonstrate the feasibility of the procedure designed to build the platform, as well as the efficiency of the different recombination reactions in uploading and excising genetic material from the human genome. Table 4 provides a summary of the results.

		Number of colonies			Positives		Frequency of positive clones (%)
		Picked colonies	Viable	Analyzed	(5') PCR	(3') PCR	
Step 1	HR	131	52	23	13	8	34
	Excision of SE by PB	48	48	34	NA	6	17
Step 2	phiC31 integrase	33	24	24	4	4	16
	Excision of SE by SB	60	60	21	NA	3	14
Step 3	phiC31 integrase+gp3	29	29	29	15	13*	48**
	Excision SE by PB	52	52	52	NA	9	17

Table 4| Recombination efficiencies at each of the three steps of the synthesis. Step 1. HR: incorporation of the first docking module by TALEN-assisted HR. Excision of the SE by PB transposase. **Step2.** phiC31 integrase: incorporation of the second docking module by site-specific phiC31 recombination. Excision of the SE by SB transposase. **Step3.** phiC31 integrase+gp3: incorporation of the third docking module by site-specific phiC31/gp3 recombination. Excision of the SE by PB transposase. Positives (5') PCR represents the number of clones analyses which showed correct integration at the 5'end. Positive (3') PCR represents the number of clones positive to 5' that were also positive for the 3' end. Frequency of positive clones represents the percentage of clones positive for both 5' and 3' end with respect to total number of clones analyzed. NA: not applicable. (*) only 13 out of the 15 5' positive clones had enough DNA to be tested for the 3' end. (**) This frequency represents the percentage of both positive of the 5' and 3' end (13) out of the 27 analyzed.

3.4. Loading test of the docking platform in HeLa cells.

After integrating the first docking element in the CCR5 locus of HeLa cells, its functionality was tested before continuing with the construction of the platform. For this purpose, HeLa Cl.24.22 cells harboring only one docking module with four docking sites were chosen. In this cell line, it is possible to determine by PCR the exact number and position of the docking units occupied by the transferred cargo (Figure 13). HeLa Cl.24.22 cells were transfected with the reporter vector pEP46 carrying a promoterless mCherry (Figure 28) and the vector expressing Bxb1 integrase. Following G418 selection, stable cells clones were isolated and expanded. Since the mCherry construct lacks a promoter, only successful recombined clones could correctly express the

reporter gene. Indeed, transfection of mCherry without Bxb1 recombinase resulted in the complete absence of mCherry expressing cells (data not shown). Finally, clones with correct growth and morphology were analysed by FACS and fluorescence microscopy based on mCherry expression (Figure 39 and Figure 40).

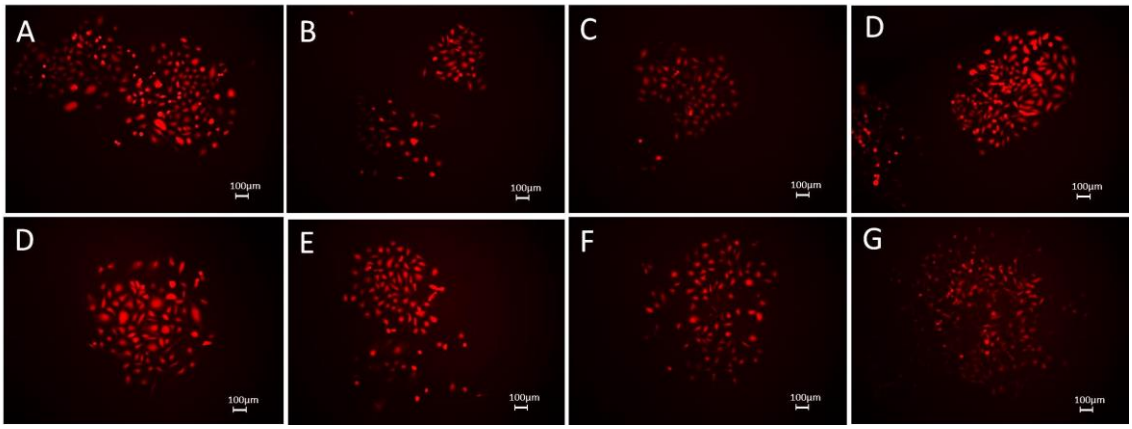

Figure 39| Determination of mCherry fluorescence by epifluorescence microscopy after pEP46 loading in HeLa Cl.24.22 cells. Fluorescence intensity was measured after 10 days of transfection by epifluorescence microscopy using the red channel and constant exposition. Scale bar = 100 µm. **(A)** Clone 1, **(B)** Clone 2, **(C)** Clone 3, **(D)** Clone 4, **(E)** Clone 5, **(F)** Clone 6, **(G)** Clone 7, **(H)** Clone 8.

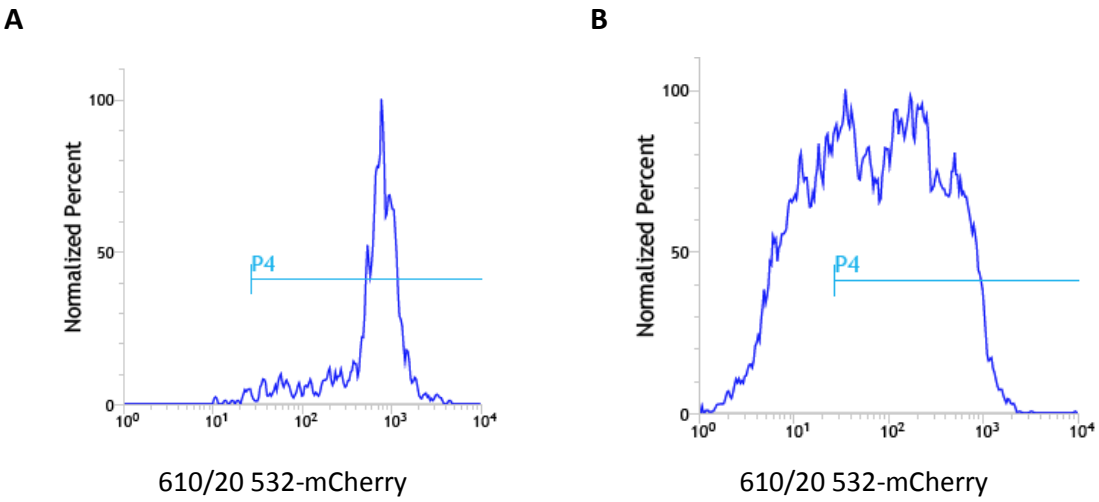

Figure 40| FACS analysis of mCherry expression after pEP46 loading in HeLa Cl.24.22 subclones. Two examples are shown. **(A)** Clone Cl24.22.5. **(B)** Clone Cl24.22.6. Fluorescence intensity was measured by FACS 10 days after transfection.

Of the eight clones analysed, seven showed multiple peaks, which could be interpreted either as evidence of polyclonality, or as variability in the level of expression in a monoclonal cell line. Only clone five showed unimodal emmission spectrum (Table 5).

	HeLa Cl.24.22 subclones							
Subclone number	1	2	3	4	5	6	7	8
Occupied docking sites	2	3	3	4	1	2	2	2

Table 5| Summary of the number of loaded Bxb1 docking sites in HeLa 24.22 clones. HeLa Cl. 24.22 harbored four Bxb1-specific attB sites.

Following DNA extraction, the number and location of the loaded recombination site was determined for all the clones by PCR analyses using site-specific Bxb1-BQ primers (Figure 41).

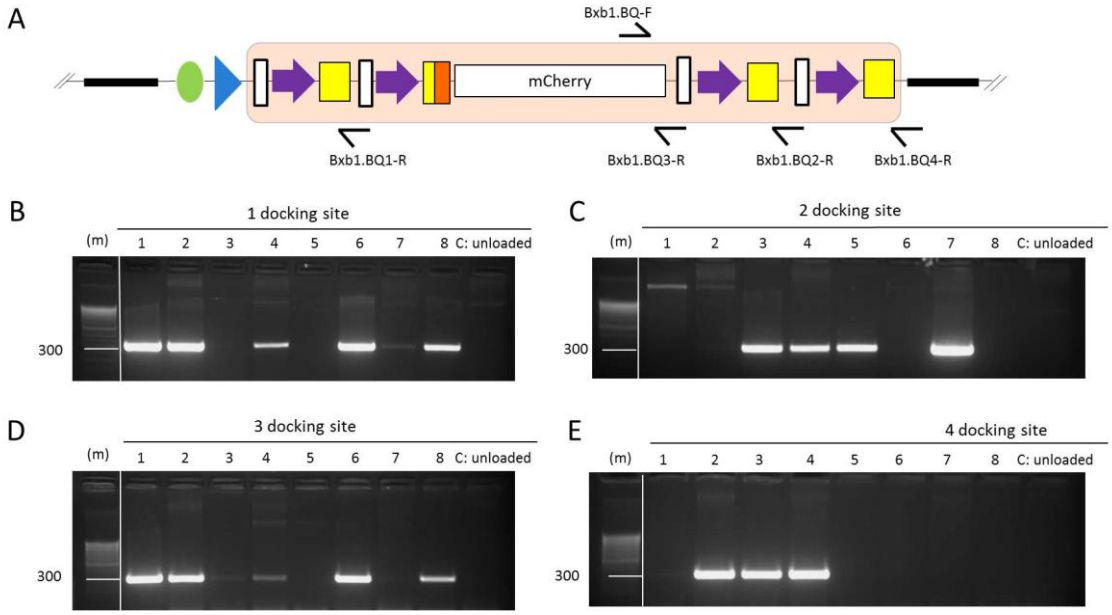

Figure 41| Molecular characterization of docking sites loading in HeLa Cl.24.22 subclones. (A) Diagram of the PCR strategy. **(B, C, D** and **E)** Agarose (2%) gel electrophoresis with ethidium bromide staining of PCR amplification products from sites 1, 2, 3 and 4, respectively. Molecular size marker= 50 bp DNA ladder.

Clone 5, showed a unimodal emission spectrum by FACS and had integrated the mCherry construct only in position 2 of the docking module. The remaining seven clones showed multimodal emissions and a multiple loading status with 1-4 docking sites occupied. It is not clear at this point whether the variability of expression could be ascribed to a possible polyclonality of each one of the seven variable clones, to different expression levels in each one of the docking sites, or simply to variabilities between cells within individual clones. Nevertheless, these results indicate that all four docking sites could be efficiently loaded and were functional, although the functionality of each position should be verified in future experiments.

3.5. Generation of hiPS cell lines

Mammalian cell lines that support reliable and predictable expression of large numbers of transgenes are an enabling technology for a wide range of scientific, industrial and therapeutic applications. To explore additional uses of the docking platform in therapeutic applications, hiPS cell lines were generated (326).

One of the fibroblast-derived MKOS line, IMEDEAAi003-A, was described and characterized regarding their genetic stability, stemness and pluripotency. First, the karyotype was analyzed by standard G-loading and found to be normal (46, XX) (Figure 42).

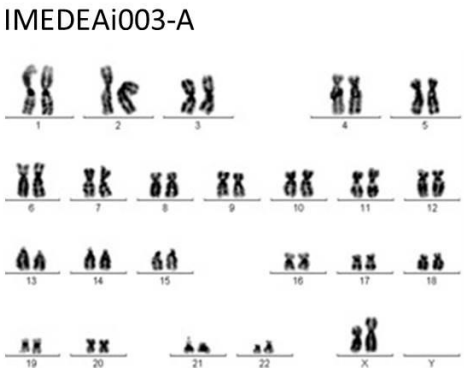

Figure 42| hiPS cell karyotyping. Gband analysis demonstrated that IMEDEAi003-A line display a normal karyotype.

The completion of the reprograming process and the silencing of the reprogramming factors were further evaluated by comparing the expression of endogenous and exogenous OCT4, SOX2, KLF4 and c-MYC by RT-PCR (Figure 43). It has been reported that successful reprogramming is associated with silencing of the retroviral Yamanaka factors (45). By using specific primers to detect the reprogramming transcripts, IMEDEAi003-A cells demonstrated complete silencing of exogenous transgenes. In addition, quantitative RT-PCR was also performed on NANOG, Rex and CRIPTO. In

contrast to primary fibroblast, IMEDEAi003-A showed strong expression of the stemness markers.

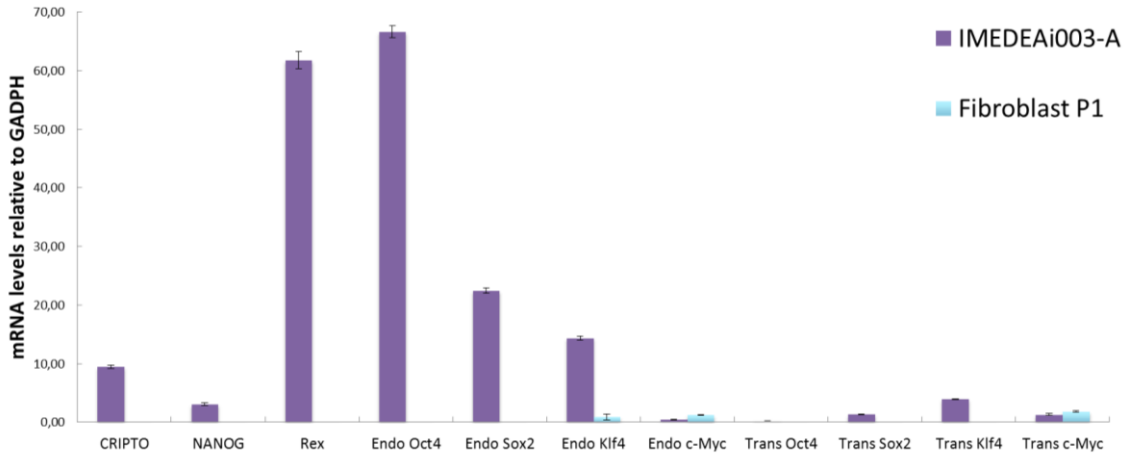

Figure 43| Quantification of pluripotency markers by qRT-PCR. The expression levels of transgenic and endogenous pluripotent factors were quantified by qRT-PCR in WT hiPS cell line, and in primary fibroblasts as negative control. Expression values were plotted relative to GADPH expression.

Stemness was assessed by immunostaining for NANOG, OCT4, SSEA3, SSEA4, TRA-1-60 and TRA-1-81. IMEDEAi003-A expressed high levels of stem cells markers (Figure 44). To test whether IMEDEAi003-A retained the capacity to differentiate into the three embryonic layers (ectoderm, mesoderm and endoderm) a method based on embryoid bodie production was used. After a minimum of ten days of suspension culture, embryod bodies were disrupted mechanically and plated on gelatine. Plated embryoid bodies derivatives were analyzed by immunostaining. Ectodermal lineages were confirmed by the expression of NESTIN and TUJ1, while GATA4 and α-ACTININ illustrated the presence of mesoderm and SOX17 and AFP of endoderm (Figure 44).

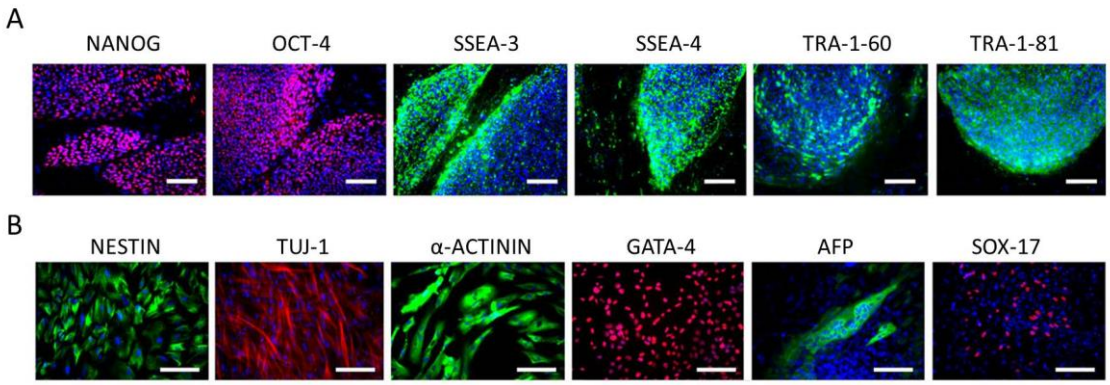

Figure 44| Characterization of the hiPS cell line IMEDEAi003-A by immune-fluorescence studies. (A) Stem cell marker expression analysis in IMEDEAi003-A line confirmed the pluripotent state of the newly generated hiPC cells. **(B)** IMEDEAi003-A cells expressed markers of the three germ layers. Nuclei were counterstained with DAPI shown in blue. (Scale bars: 100 µm)

3.6. Assembly of the first docking module into the CCR5 locus of hiPS cells

The first docking step consisted in the integration of the docking recombination, pEP24, in the CCR5 locus by TALENs-stimulated homologous recombination.

For this purpose, as in the case of the HeLa cell line, the hiPS cell line IMEDEAi003-A was transfected with the pEP24 recombination vector and CCR5-specific TALENs. Following puromycin selection 152 clones were isolated and expanded. 4 clones displayed correct integration of the recombination cassette at the 5' terminal region (Figure 45 A). 4 clones displayed integration of the recombination cassette at 5' (Figure 45 B), whereas 3 out of the 4 clones showed successful integration at the 3' end (Figure 45 C). The hiPS Clone Cl.87, showing correct growth and sensibility to FIAU (data not shown), was chosen for further experimentation.

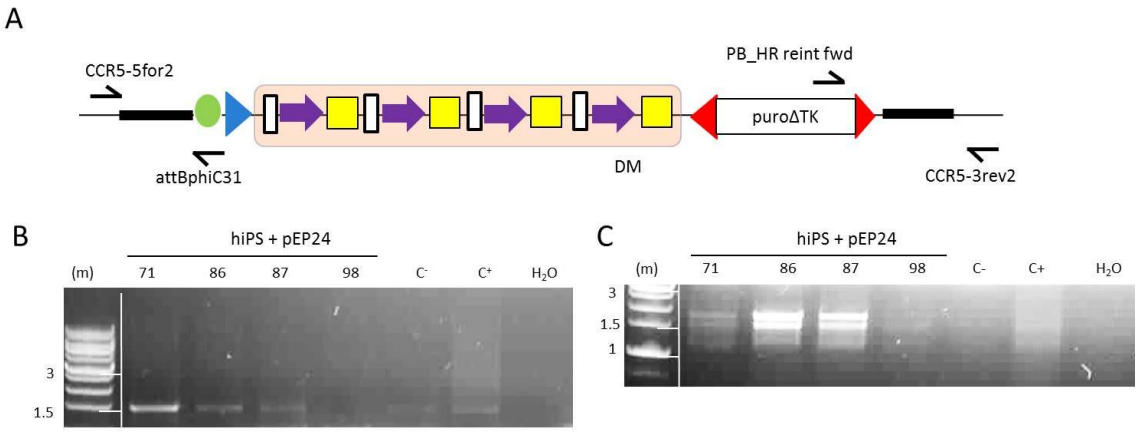

Figure 45| Identification of recombined clones by PCR screening in step 1 of platform construction in hiPS cells. **(A)** Diagram of PCR characterization. **(B)** 5' integration site analysis by PCR using CCR5-5for2/attBphiC31 primers (amplicon: 1500 bp). **(C)** 3' integration site analysis with the primers PB_HR reint fwd/CCR5-3rev2 (amplicon: 1800 bp). Agarose gel (0.8%) electrophoresis with ethidium bromide staining. Positive control: HeLa Cl.24 cells containing the first docking module. Negative control: HeLa WT cells without docking module. Molecular size marker, m= 1 kb DNA ladder.

Next, the SE (3101 bp) was excised by transfection with PB transposase followed by selection with FIAU. Successful excision of the selection elements was confirmed in 24 out of 52 clones analysed by PCR with the attB-Bxb1-1/CCR5-3rev2 pair of primers (Figure 46). All clones showed excision except hiPS Cl.87.35 (Figure 46). hiPS Cl87.38 was used for functional testing of the docking platform since it displayed the best growth properties.

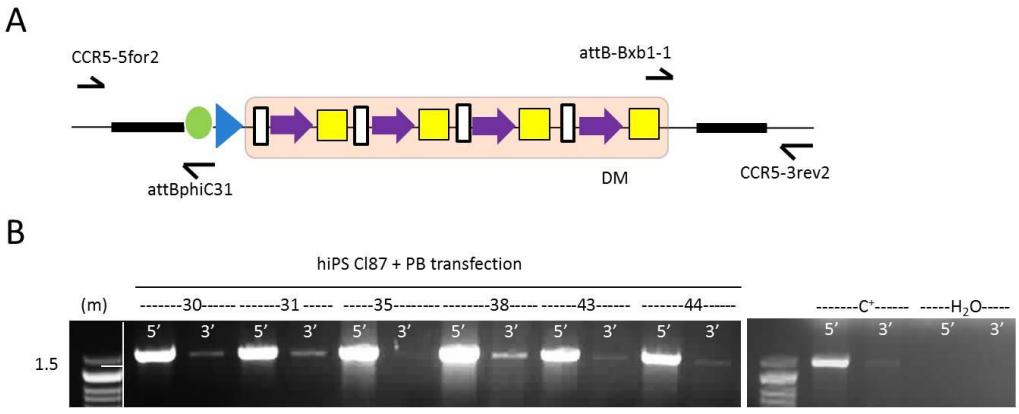

Figure 46| Excision of the selection marker by PBc transposase in hiPS Cl.87 cells. (A) Diagram of the PCR characterization. **(B)** Representation of six clones analyzed by PCR using the CCR5-5for2/attBphiC31 pair of primers (ampicon: 1500 bp) or using attB-Bxb1-1/CCR5-3rev2 (amplicon: 1500 bp) to detect 3' excision site. Positive control: HeLa Cl.24 cells harboring the first docking module. Agarose gel (0.8%) electrophoresis with ethidium bromide staining. Molecular size marker, m= 50 bp DNA ladder.

Table 6 contains a summary of the results obtained in the loading of the first docking module in human iPS cells.

	Number of colonies			Positive		Frequency of positive clones (%)
	Picked colonies	Viable	Analyzed	(5') PCR	(3') PCR	
HR	157	152	47	4	3	5
Excision of SE by PB	72	69	52	NA	24	46

Table 6| Recombination efficiencies during hiPS loading. HR. Recombination vector pEP24 integration by TALEN-assisted HR. **Excision** of SM by phiC31-specifi PB transposase. Picked colonies: individual clones picked after puromycin selection. Viable: number of clones surviving the recovery and expansion process. Analyzed: number of clones analyzed. Positives (5') PCR: number of clones analyses which showed correct integration at the 5'end. Positive (3') PCR: number of clones positive to 5' that were also positive for the 3' end. Frequency of positive clones: percentage of clones positive for both 5' and 3' end with respect to total number of clones analysed. NA: No applicable.

Frequencies obtained were 5 % in homologous recombination and 46 % during the excision of selection marker. After excision, the integrity of the phiC31-specific attB site of hiPS Cl87.38 was confirmed by sequencing (Figure 47).

```
CLUSTAL O (1.2.4) multiple sequence alignment

attBphiC31+RIGHT-ITR SB ----TGCGGGTGCCAGGGCGTGCCCTTGGGCTCCCCGGGCGCGTACT
hiPS Cl.87.38-------------GGTGCGGGTGCCAGGGCGTGCCCTTGGGCTCCCCGGGCGCGTACT

attBphiC31+RIGHT-ITR SB---CCACCTCACCCATCTGGTCCATCATGATGGTACCGTATTGAGTGT
hiPS Cl.87.38-------------CCACCTCACCCATCTGGTCCATCATGATGGTACCGTATTGAGTGT
```

Figure 47| hiPS Cl.87.38 cells sequence analysis. Theoretical attB phiC31 site sequence aligned to hiPS Cl87.38 sequence. In red and bold the 50 bp recognition sequence required for recombination.

3.7. Loading test of the docking platform in hiPS cells.

Once the first docking module was integrated into the CCR5 locus in hiPS cells it was necessary to test its functionality. Similar to the loading experiment already described for HeLa cells, hiPS CL.87.38 cells were transfected with the reporter plasmid pEP46 (attP Bxb1+mCherry+Neo^{R}) and a vector encoding the Bxb1 integrase. Following G418 selection, stable cells clones were isolated and expanded. Since the mCherry construct was promoterless, only successfully recombined clones could express the reporter gene. Indeed, transfection of mCherry without Bxb1 recombinase resulted in the complete absence of mCherry expressing cells. Finally, clones with correct growth and morphology were analysed by FACS and fluorescence microscopy regarding their mCherry expression (Figure 48).

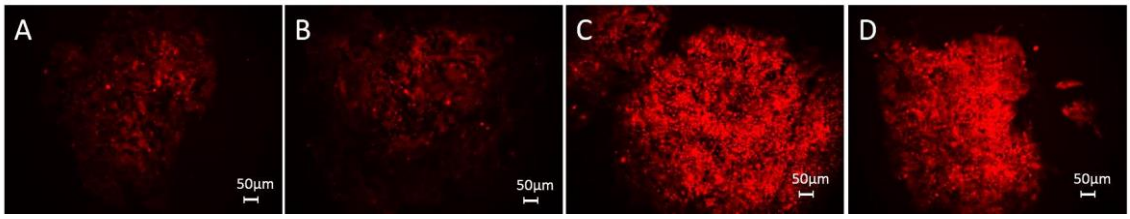

Figure 48| Determination of mCherry fluorescence by epifluorescence microscopy after pEP46 loading in hiPS Cl.87.38 cells. Fluorescence intensity was measured after 10 days of transfection by epifluorescence microscopy using the red channel and constant exposition. Scale bar = 100 µm. **(A)** Clone 13, **(B)** Clone 18, **(C)** Clone 21, **(D)** Clone 22.

Following DNA extraction, the number and location of the loaded Bxb1 sites was determined in 24 clones by PCR analyses, according to the screening strategy already described for HeLa cells (Figure 13, Table 1 and Table 7).

hiPS Cl.87.38 clones' number	Occupied docking site	Total occupied docking site
1	1; 3	2
2	3	1
3	3; 4	2
4	3; 4	2
5	1	1
6	2; 3; 4	3
7	3; 4	2
8	4	1
9	2; 3	2
10	4	1
11	1; 3	2
12	1; 2; 3	3
13	2	1
14	2	1
15	2; 3; 4	3
16	2; 3	2
17	2	1
18	2	1
19	2	1
20	2	1
21	2; 4	2
22	1; 2; 3	3
23	1; 2; 3	3
24	2	1

Table 7| Summary of the number of loaded Bxb1 docking sites in the hiPS 87.38 clone. hiPS Cl.87.38 contained four attB Bxb1 attachment sites. It was transfected with pEP46 that harbored attP Bxb1 attachment site. Blue shadowed lines: unimodal mCherry expression by FACS.

FACS analysis revealed unimodal expression in 11 out of 24 clones, however, PCR analysis showed that some of the clones with multimodal expression (5, 10, 18, 20 and 24) had only one docking site occupied. This combination of factors suggests that the

variance of expression observed could be due to intrinsic variability within the lines and not to differential activity between the four docking units.

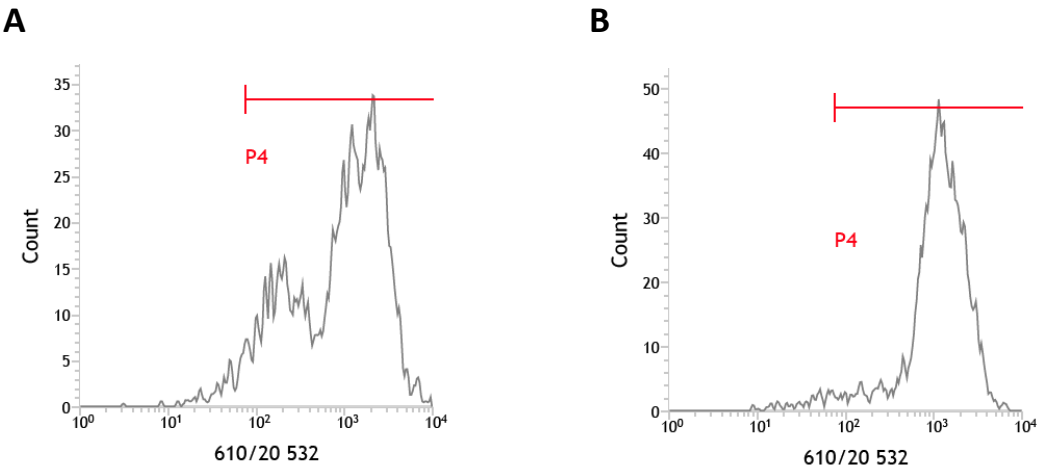

Figure 49| FACS analysis of mCherry expression after pEP46 loading in hiPS Cl.87.38 subclones. Two examples are shown. **(A)** Clone Cl.87.38-22. **(B)** Clone Cl.87.38-21. Fluorescence intensity was measured by FACS 10 days after transfection.

Overall, the results obtained with hiPS cells confirm those previously observed in HeLa cells, and demonstrate that the same strategy for building the docking platform can be successfully used in very different cells lines. In addition, they underline the flexibility and versatility of the docking system and highlight the potential to load various transgenes at the same time.

4. DISCUSSION

The goal of this PhD thesis was the construction of a docking site for the secure and easy loading of genetic information into the genome of mammalian cells. For this purpose, complex arrays of molecular and cellular techniques were integrated into a single protocol. The strategy provides a serial mechanism by which multiple docking modules, each comprised of four docking sites, can be assembled into the human CCR5 locus. First, TALEN-mediated homologous recombination promotes the precise integration of the first docking module. Then, a combination of the site-specific recombination systems phiC31 and transposon systems, piggyBac and Sleeping Beauty, catalyzes the addition of a theoretically indefinite number of docking modules and the excision of unwanted DNA elements that might interfere with the loading and expression of future genes of interest.

The functionality of this multiple site-specific Bxb1 integration platform was confirmed by loading the fluorescence promoterless mCherry reporter gene.

As a proof of concept, the construction of the docking platform was first carried out in the immortalized, easy-to-transfect, HeLa cell line. Once the functionality of the system was confirmed, the platform was assembled in a therapeutically relevant cell-type. With this aim, hiPS cells were generated from healthy donor's fibroblast by reprogramming with the four canonical Yamanaka factors. Bona fide hiPS cell lines were identified and their stemness, pluripotency and correct karyotype were verified.

In the first step of platform assembly, a docking module was integrated into the CCR5 locus by TALEN-assisted HR. In the last decade, ZFNs, TALENs and CRISPR/Cas9 have been intensively used to introduce targeted and specific modifications into the genome of living cells and organisms. However, several reports have shown that ZFNs have lower success rates due to lower cutting efficiency than the other two (68). TALEN and CRISPR/Cas9 endonucleases have recurrently achieved equally good rates of gene editing. TALENs have demonstrated to be useful to correct disease phenotypes

(136; 334). In fact, TALEN have shown up to 90% cutting efficiency with negligible off-target activity in HEK 293 cells (335).

In recent years, our laboratory has designed and developed, for its application in another research project, a novel pair of TALENs to target exon 3 of the CCR5 gene. This pair of TALENs showed high efficiency and specificity in a number of stable cell lines and primary cells and has, therefore, been chosen for this PhD thesis.

Although the docking platform could have been assembled in several loci like AAVS1, Rosa26 or H11, several observations suggested selecting CCR5 as the integration site for the multi-copy docking platform. It has been shown that integration in the AAVS1 locus disrupts the phosphatase 1 regulatory subunit 12C (PPP1R12C) gene (314) with consequences that are still unknown. In addition it can also produce silencing of the transgenes by DNA methylation (318). At the same time, Human Rosa26 and H11 have not been widely used to express transgenes in pluripotent stem cells, and the effects of manipulating these genes are not fully established. On the other hand, gene editing and knock-out studies have demonstrated normal cell function and differentiation upon CCR5 deletion *in vitro* (99; 132), but the most striking functional argument in favour of CCR5 is that a natural null mutation of the gene is present in human individuals with no apparent consequences to their health. (57). Finally, an additional practical reason is the good performance of the CCR5 specific pair of TALENs mentioned in the previous paragraph.

In the experiments conducted during this PhD thesis, correct TALEN-mediated homologous recombination events occurred at a frequency of 34 % in HeLa and 5 % in hiPS cells. These efficacies are obviously much higher than those obtained by classical, unassisted HR (80; 81; 336), especially in hiPS cells (336). These results are comparable to allele targeting efficiency promoted by genome-specific endonucleases ranging from 2-34% (337; 338; 137). The differences between cell lines in targeting efficiency could be attributed to chromatin structure and/or DNA methylation that may negatively affect accessibility of TALENs to DNA (339). The chromatin structure is dynamic and differs between genomic loci (340).

The next step in the protocol was to integrate the second docking module by site-specific recombination catalyzed by phiC31. Serine integrases, such as phiC31, promote recombination between two different DNA recognition sites, attP and attB, resulting in new recombined attL and attR sites (189; 341; 34). PhiC31 integrase mediates unidirectional recombination between attP and attB with 50-80% efficiency (187; 195; 211). In this PhD work, targeting efficiency of only 16% was obtained through site-specific phiC31 recombination in HeLa cells. Transfection method, cell line, plasmid vector length and amount of phiC31 integrase could explain the low recombination frequency obtained at this step. It has been shown that both the integration specificity (214; 215) and efficiency of the phiC31 integrase is cell line dependent (215) due to distinct nuclear envelope and chromatin structures (342).

The second step of the protocol includes the integration into the genome of a phiC31-specific attB attachment site that could later be used as a target for recombination with the phiC31-specific attP site located in the second docking vector, pEP28. This particular arrangement of the attB/attP pair has been included to prevent random integrations of the docking vector into the pseudo-attP sites present in the human genome (211). Pseudo attP sites have been occasionally used when the place of integration is not relevant (343; 344; 345). However, random integration at pseudo-attP sites often involves loss of several base pairs at the integration site and sometimes extensive chromosome rearrangements (214; 346).

PhiC31-specific attachment sites are used to build the module into the chosen genome location and Bxb1-specific attachment sites to upload the genes of interest once the docking platform is completed. PhiC31 can work on a variety of phiC31-specific attP and attB sites but they must have identical central dinucleotides (2 bp overlap) (347). Thus far, six different cores pairs (TT, CT, GT, CA, CC, and TC), have been identified on attP/attB site, and no differences in recombination efficiencies have been reported (34; 189; 191). However, attPs and attBs from different pairs do not work together.

Several studies have illustrated the impact of different kinds of interactions on phiC31 activity. It has been shown that the conserved N-terminal domain of phiC31 is

responsible for DNA cleavage, strand exchange, and protein-protein interactions (207), whereas the large C-terminal domain is required for DNA binding (196). To increase phiC31 activity and specificity distinct mutations were introduced into the N-terminal (209; 210) and C-terminal domains (196; 197). Variants with mutations T2I, V6A, S7Δ, V9E, D10V, D40A, D44A, V49I, and D52A are all located in the N-terminal part of the enzyme. This domain has been shown to have limited ability to recognize pseudo attP sites, but an increased affinity for the wild type-attP sequence. Hence, overall increasing the specificity for the attP site (210). Performing site-specific recombination with this mutant integrase could potentially increase the recombination rates observed in the experiments previously described.

Besides catalysing the recombination between attP and attB sites, phiC31 can also promote site-specific recombination between attL and attR sites although this reversion in polarity, termed excision, requires the recombination directionality factor (RDF) gp3, in addition to the integrase. RDFs promote attL and attR recombination and simultaneously inhibit attP and attB recombination (34; 189; 341). PhiC31 excision activity has been demonstrated in *E.coli* (341) and mammalian cells (219) with recombination efficiencies of 79% and 35-75%, respectively (219; 200).

In step three of the synthesis of the docking platform, the recombination vector, pEP50, was successfully integrated in HeLa cells by phiC31 and gp3 at a frequency of 48%, which is in concordance with the published results (see above).

At the three steps of the construction process, correct integration events were assessed by PCR analysis of the 5' and 3' ends of the integration site. In the case of HR-mediated integration only 8 out of the 13 clones that were positive for the 5' end showed also correct integration of the 3' extreme, whereas in the cases of phiC31 integrase-mediated recombination, all clones with correct 5' integration were also accurately integrated at the 3' end. This data highlights the considerable difference in precession of both site-specific integration systems and supports our original argument in favor of a loading system based on phiC31 and Bxb1 integrases.

Our general design includes the use of two different site-specific recombinase systems: phiC31 for the assembly of the docking platform and Bbx1 to load the desired genetic elements onto the platform after its completion. There are several reasons for these assignments. First, it was necessary that the integrase system employed during the synthesis would also be capable of catalysing the excision of unwanted materials after each round of recombination. Both Bxb1 and phiC31 require an additional factor, RDF, to catalyze the excision reaction. Although the RDF of phiC31, gp3, has been validated in mammalian cells, the RDF of Bxb1 has not (219). In addition, gp3 seems to have some inhibitory effect on phiC31 integration activity thus preventing possible undesired recombination between the regenerated attB and attP (341). This inhibitory effect has also not been observed in Bxb1-gp47 (RDF)-mediated recombination (218; 226; 341). The synthesis of the platform is complex, but once it is finished, the cells equipped with it would be used as a base for multiple pharmaceutical and biomedical applications. To increase its applicability, it was very important to ease the loading process. Since the Bxb1 integrase system is the most efficient recombinase described in literature (59; 187; 227), and attL/attR recombination was only required during the phase of synthesis, phiC31 was chosen for the assembly of the docking platform and Bxb1 for the loading phase.

To test the functionality of the novel docking platform, a proof-of-principle study was performed by loading the fluorescence reporter gene mCherry in cells harboring only one docking module with four attB loading sites. This system offers the great advantage to distinguish by PCR between the single loading sites regarding their loaded/unloaded state. The analysis of a wide variety of HeLa and hiPS cells clones revealed that each cloning position was functional. Moreover, multiple loading events could be detected thereby generating cells with two, three or four copies of mCherry. Most loading recombinase systems described in the literature have a unique site-specific recombination site (59; 223; 348), so they could accommodate only a single recombination event. Few months ago, a new multi-copy loading system was described in CHO cells. It is based on the integration of unique attachment sites, into multiple, not previously characterized, chromosomal locations. The system allows for

the integration of various transgenes at predetermined chromosomal positions (67), but their behaviour as “bona fide” safe harbor sites remains to be fully tested.

The docking functional platform presented in this PhD work can also allow for simultaneous integration of various transgenes. These integrations are performed at the same chromosomal location, which could be useful to study metabolic routes or gene interactions that require close transgene positions. In addition, multiple docking units can be incorporated into the genome after the initial single homologous recombination event, thus, reducing the possibility of off-target effects when compared to several independent integrations at different locations.

The docking module described here consists of four docking sites. Each docking site represents a transcriptional unit containing a cHS4 insulator, the SV40 promoter and a Bxb1 attachment site (Figure 21). Insulators are a class of DNA elements that possess a common ability to protect genes from inappropriate signals emanating from their surrounding environment (349; 350). Insulators can act as a positional enhancer blocker; if the insulator is situated between an enhancer and a promoter, then enhancer-mediated activation of the promoter is impaired by the insulator. If the insulator lies outside the region between enhancer and promoter little or no effect is observed (351). Insulators also have the ability to protect against position effects by acting as barriers, preventing the advance of nearby condensed chromatin that might otherwise silence expression (352; 353). Transgene activity is often affected by chromosomal position effects (349; 354). In conclusion, it is crucial to use insulators between the single loading sites in order to prevent silencing of transgenes integrated in the docking module or run off effects from one promoter over neighbouring ones. In addition, insulator may also protect the genome from the insertional activation of surrounding proto-oncogenes, a very important consideration for cell therapy applications (355). An example of insulator widely used is cHS4, derived from the chicken β-globin locus (330). The cHS4 element is the best characterized, flanking transgene, already tested in many cell lines (353; 356; 357; 358; 359). Besides the insulator type it is important to determine the number and position of the insulator

elements in the docking module. The cHS4 insulator alleviates transcriptional interference occurring between two neighboring transgenes positioned at both sides of the transgenes (358; 360) and it has been described that a single element is enough to achieve higher and long-term transgenes expression levels (358). Finally, only one cHS4 insulator element was placed 5' upstream of SV40 promoter to guarantee the function of each transcriptional unit while minimizing the size of the docking vector.

The choice of promoter has a great impact on both the level and the duration of transgene expression. In our proof of concept study, the docking platform was designed and constructed with the small but strong SV40 promoter, which is frequently used to achieve high levels of expression in a variety of mammalian cell and tissue types (361). The SV40 promoter was cloned 5' upstream of the Bxb1 attachment site, so that only successful recombination events could give rise to cell clones with correct transgene expression. As the loading vector lacks a promoter element, randomly integrated transgenes will not be expressed in the cells. In order to improve transgene expression and prevent potential silencing events due to the viral nature of the SV40 promoters, the docking unit was recently redesigned, substituting the SV40 promoter for the human eukaryotic translation elongation factor 1 alpha 1 (EEF1A1) promoter (362).

In order to obtain recombined SM-free cell clones, the strategy developed in this PhD thesis involves a double selection process after each step of homologous or site-specific recombination. First, recombined cells clones carrying the recombination cassette including the double selection marker, puroΔtk, were selected by puromycin. Second, following PB or SB transposase excision, FIAU selection allowed the isolation of recombined cells clones that had successfully excised the SM. The inclusion of this step in the building strategy had the additional objective of removing from the safe harbor location any DNA vector fragment that might later have undesired effects on transgene expression (363). Besides the elimination of exogenous vector DNA, it was crucial to excise the attachment site attR generated by attP/attB recombination in order to continue with step two of the assembly strategy.

Following TALEN-assisted HR at the CCR5 locus, the cells carried the selection cassette puroΔtk flanked by both ITR elements of PB and the left-ITR element of SB (Figure 14). Once cells were transfected with PB transposase, the selection element was removed leaving only the left-ITR of SB, confirming that transposases' specificities did not overlap and was possible to use them in the same design. After performing the first step of the platform assembly (attP/attB recombination) the ITR configuration completely changed. The selection cassette was now flanked by ITRs (left and right) of SB and the left ITR element of PB (Figure 15). Finally, following step three (attL/attR recombination) the order of ITRs was identical to the situation described after homologous recombination (Figure 16). Even though each transposon has "cut and paste" activity, each species has its own molecular structure and mechanism and ITR sequences are specifically required for the interaction with their cognate transposases (364). For these reasons, PB transposase is unable to interact with ITRs of SB and vice versa. This concept has been validated in this study.

Cell type dependant transposition efficiencies have been observed during the creation of the docking platform. The excision efficiency by PB was 17% in HeLa cells whereas SB-mediated excision was observed in 14% of HeLa clones analyzed. However, transposition efficiencies of 31% (364) and 40-50% (365; 366), have been reported for SB or PB, respectively. One possible explanation for the differences could be that the literature refers to integration experiments, whereas our data corresponds to excisions. Alternatively, it has been described that both transposon systems are sensitive to the overproduction inhibition phenomenon (OPI), a decrease in transposition efficiency above a certain level of cellular transposase concentration (364; 367). Very likely, reducing the amount of transposase could improve the excision efficiency in HeLa cells.

The efficiency obtained in this PhD work for PB-mediated excision in hiPS cells, 46%, was similar to those obtained in published studies, which are significantly higher than in HeLa cells. One possible explanation for such variability between HeLa and hiPS cells could be the interaction of the transpositional machinery with cell-specific host factors

(233). It has been shown that both transposon systems (PB and SB) present different efficiency depending on the cell type in which they are tested (368; 246).

Most gene expression platforms described in the literature are based on a single site-specific recombinase element, previously introduced in the genome by HR (223; 369; 348; 67; 370; 371; 372) . This design has the advantage of its simplicity, but the counter effect of its very limited functionality, since it only permits one single loading. Unlike these platforms, our system can provide as many attachment sites as needed, offering integration of several transgenes at the same locus. In addition, the docking platform presented in this PhD thesis is the first example of a loading system that, thanks to the alternated use of PB and SB, eliminates, together with all remaining vectors elements, the SM (puroΔtk), thus making possible to use the same selection cassette in consecutive loadings.

More recently, three new types of docking platform have been developed. In 2014, X., Duportet et al. designed a novel system integrated by ZFN-aided HR at the AAVS1 locus of HEK293, Hela, CHO and hESC cells. The platform contained a Bxb1-specific attP site and the EEF1A1 promoter upstream of the attachment site. The system was completed by a docking vector carrying a promoterless selection cassette. As a result, expression of the selection element after integration was under control of the docking site constitutive promoter. The authors were able to incorporate four genes of interest by using a multigene expression vector. Nevertheless, the system could only support a single integration event, which severely limited its loading capacity. Our platform incorporates a constitutive promoter at each one of the docking sites and promoterless loading vectors, but unlike Duportet's et al. platform, it offers unlimited loading capacity. In addition, it also admits the use of multigene docking vectors

Zhu et al. (2013) have proposed a system termed Dual Integrase Cassette Exchange (DICE) that consists of adjacent phiC31 and a Bxb1-specific attP attachment sites integrated by TALEN-assisted HR in the H11 locus of ESC and iPS cells. Correct integration required simultaneous recombination at both sites, thus excluding the incorporation of all docking vector elements located outside the area defined by the

attB sites (59). The advantages of DICE would be the increase in efficiency by reducing transgene incorporation in genomic pseudo-attP sites, and the elimination of plasmid backbone integration. Our design also circumvents undesired recombination events with pseudo attP sites by the incorporation of attB instead of attP sites in the docking platform. In addition, the undesired DNA elements are excised by a double transposon system. Although DICE seems to be an attractive model, it only allows single transgene integration and cannot therefore serve for multiple docking events.

The latest docking platform to be described in the literature is based on the integration of single attachment sites into several locations of the genome (67). The authors identified new, putative safe harbor sites in the human genome, among them, a Rosa 26 homolog that could potentially support long-term expression of transgenes. They integrated single Bxb1-specific attP sites simultaneously at three selected loci in CHO cells by CRISPR/Cas9-assisted HR. The system generated a functional platform for site-specific integration into three distinct chromosomal locations. However, the assembly of the platform required the use of CRISPR/Cas9-assisted HR for the establishment of each one of the docking sites, therefore, multiplying by three the possibility of off-targets events. In contrast, the docking platform designed in this PhD thesis utilizes a single TALEN-assisted recombination episode to set up the core of a multi-copy platform, reducing significantly the risk of off-targets events.

This PhD work has designed and developed a novel strategy to achieve accurate multi-site-specific integration of exogenous genetic information into mammalian cells. The assembly of the docking platform combined TALEN-assisted HR with phiC31-mediated site-specific recombination followed by the combined use of PB and SB transposon to excise the selection cassette. The alternating use of PB and SB ITR element has permitted the use of the same selection cassette, puroΔtk, in each recombination step and the excision of any vector elements that could later on interference with transgene expression. Furthermore, the transposon-dependant excision of the selectable element used during the assembly of the docking platform could also be

incorporated into a loading vector, thus allowing the successive integration of genes of interest without selection marker limitations.

Unlike other loading systems, this novel multiple docking platform has been designed to incorporate four docking units at each recombination step. More importantly, the building cycle can be repeated as many times as needed, even after genes of interest have already been loaded. This feature gives the platform the plasticity to adapt to new needs arising from the behaviour of the already loaded elements.

The technology promises improvements in cell line engineering for applications in therapy, diagnostics and biotechnology. A multi-copy system could be very useful in these contexts if multiple integration sites are required in order to maintain independent expression from different parts of a synthetic gene circuit. For example, in a situation where the way in which multiple transcriptional units interact with each other is not known. In this case, it would be possible to sequentially integrate different units modifying also their order. This docking platform could also be used to generate disease models for drug screening and disease pathophysiology both in animal and human pluripotent cells. A docking platform with multi-copy docking sites could be used to generate models of cancer metabolism and anti-cancer treatment by the integration of cancer genes into the docking followed by the incorporation of synthetic lethal genes (373) into other docking site (374).

The system could, as well, be adapted to yeast for the production of biofuels and bio-based chemicals by engineering metabolic pathways in the multi-copy integration platform. Target bio-bases chemicals obtained through the metabolic engineering of *Saccharomyces cerevisiae* are primarily derived from pyruvate, but this production generates by-products that limit its efficiency. In order to improve the glucose consumption rate and reduce the amount of by-products, this docking system could be employed to co-overexpress enzymes involved in the rate limiting step of yeast glycolysis (375). Cell lines with different number of docking sites could as well be useful to study the optimal gene copy number for protein expression. Different cell lines may have alternative tolerances for the number of copies of a given gene. In the same way,

any given cell line would very likely have different tolerance for different proteins. This is already well established in yeast where, for example, it has been demonstrated that 12 copies is the optimal number for secretory expression of porcine insulin precursor (PIP) in *P. pastoris* (376).

Thanks to the results obtained with Hela cells, the validity of the design, and the functionality of the platform were confirmed. Preliminary data obtained from hiPS cells has also shown that the strategy can be applied to different cell lines, suggesting that phiC31 and Bxb1 integrases, together with the PB and SB transposon systems, and the set of recombination and loading vectors already synthesized, might mediate chromosomal integration in a wide range of host environments, including plants, insects and other mammalian cells.

5. CONCLUSIONS

1. TALEN-assisted HR is an efficient method to integrate exogenous genetic material into the human CCR5 locus of HeLa and hiPS cells.
2. The insertion of phiC31-especific attB elements into the human genome provides efficient attachment sites for phiC31 activity and prevents recombination at endogenous pseudo attP sites
3. The combined activity of phiC31 and gp3 is highly effective in catalysing integration reactions in HeLa and hiPS cells.
4. The alternative use of phiC31 specific attB/attP and attR/attL attachment sites permits a continuous integration-excision cycle and an unlimited number of loading events.
5. Alternating the use of Sleeping Beauty or piggyBac specific ITR elements, transposases can be sequentially employed to promote the excision of unwanted DNA elements.
6. Removal of the selection cassette after each loading cycle, allows for the repetitive use of the same selection strategy.
7. Site specific recombination catalyzed by integrases is more accurate and produces less unspecific recombination events that Homologous Recombination.
8. Bxb1 integrase is highly efficient in promoting simultaneous recombination into the same genomic locus of HeLa and human iPS cells.
9. The combination of TALEN-assisted homologous recombination, site-specific recombination mediated by the phiC31 and Bxb1 integrases and excision of unwanted DNA elements by the Sleeping Beauty and piggyBac transposons, provides a practical, rapid and accurate strategy for the creation of multi-copy docking sites at specific mammalian genomic loci.

6. REFERENCES

1. *Synthetic biology and microbioreactor platforms for programmable production of biologics at the point-of-care.* **Perez-Pinera, P., et al.** 2016, Nature Communications, 7: 12211.

2. *Combinatorial metabolic engineering of Pseudomonas putida KT2440 for efficient mineralization of 1, 2, 3-trichloropropane.* **Gong, T., et al.** 2017, Scientific reports, 7(1): 7064.

3. *Genome editing in nonhuman primates, approach to generating human disease models.* **Chen, Y., Niu, Y. and Ji, W.** 2016, Journal of internal medicine, 280(3): 246-251.

4. *Engineering large animal models of human disease.* **Whitelaw, C. B. A., et al.** 2016, The Journal of pathology, 238(2): 247-256.

5. *Generation of a nonhuman primate model of severe combined immunodeficiency using highly efficient genome editing.* **Sato, K., et al.** 2016, Cell Stem Cell, 19(1): 127-138.

6. *Chimeric antigen receptor–modified T cells in chronic lymphoid leukemia.* **Porter, D. L., et al.** 2011, New England Journal of Medicine, 365(8): 725-733.

7. *Gene therapy for ADA-SCID, the first marketing approval of an ex vivo gene therapy in Europe: paving the road for the next generation of advanced therapy medicinal products.* **Aiuti, A., Roncarolo, M. G. and Naldini, L.** 2017, EMBO molecular medicine, e201707573.

8. *Beyond editing to writing large genomes.* **Chari, R. and Church, G. M.** 2017, Nature Reviews Genetics, 18(12): 749.

9. *Hybrid restriction enzymes: zinc finger fusions to Fok I cleavage domain.* **Kim, Y. G., Cha, J. and Chandrasegaran, S.** 1996, Proceedings of the National Academy of Sciences, 93(3): 1156-1160.

10. *Stimulation of homologous recombination through targeted cleavage by chimeric nucleases.* **Bibikova, M., et al.** 2001, Molecular and cellular biology, 21(1): 289-297.

11. *Breaking the code of DNA binding specificity of TAL-type III effectors.* **Boch, J., et al.** 2009, Science, 326(5959): 1509–1512.

12. *Targeting DNA double-strand breaks with TAL effector nucleases.* **Christian, M., et al.** 2010, Genetics, 186(2): 757-761.

13. *A TALE nuclease architecture for efficient genome editing.* **Miller, J. C., et al.** 2011, Nature biotechnology, 29(2): 143-148.

14. *Transcription at different salinities of Haloferax mediterranei sequences adjacent to partially modified PstI sites.* **Mojica, F. J. M., Juez, G. and Rodriguez-Valera, F.** 1993, Molecular microbiology, 9(3): 613-621.

15. *CRISPR/Cas, the immune system of bacteria and archaea.* **Horvath, P. and Barrangou, R.** 2010, Science, 327(5962): 167-170.

16. *The CRISPR/Cas bacterial immune system cleaves bacteriophage and plasmid DNA.* **Garneau, J. E., et al.** 2010, Nature, 468(7320): 67.

17. *CRISPR-mediated modular RNA-guided regulation of transcription in eukaryotes.* **Gilbert, L. A., et al.** 2013, Cell, 154(2): 442-451.

18. *RNA-guided human genome engineering via Cas9.* **Mali, P., et al.** 2013, Science, 339(6121): 823-826.

19. *The new frontier of genome engineering with CRISPR-Cas9.* **Doudna, J. A. and Charpentier, E.** 2014, Science, 346: 1258096.

20. *DNA targeting specificity of RNA-guided Cas9 nucleases.* **Hsu, P. D., et al.** 2013, Nature Biotechnology, 31(9): 827–832.

21. *Comparison of Zinc Finger Nucleases Versus CRISPR-Specific Nucleases for Genome Editing of the Wiskott-Aldrich Syndrome Locus.* **Gutierrez-Guerrero, A., et al.** 2018, Human gene therapy, 29(3): 366-380.

22. *A CRISPRi screen in E. coli reveals sequence-specific toxicity of dCas9.* **Cui, L., et al.** 2018, Nature Communicationsvolume, 9: 1912.

23. *Incorporation of bridged nucleic acids into CRISPR RNAs improves Cas9 endonuclease specificity.* **Cromwell, C. R., et al.** 2018, Nature communications, 9(1): 1448.

24. *Unbiased detection of off-target cleavage by CRISPR-Cas9 and TALENs using integrase-defective lentiviral vectors.* **Wang, X., et al.** 2015, Nature biotechnology, 33(2): 175.

25. *High-fidelity CRISPR–Cas9 nucleases with no detectable genome-wide off-target effects.* **Kleinstiver, B. P., et al.** 2016, Nature, 529(7587): 490-495.

26. *Highly efficient heritable targeted deletions of gene clusters and non-coding regulatory regions in Arabidopsis using CRISPR/Cas9.* **Durr, J., et al.** 2018, Scientific reports, 8(1): 4443.

27. *Dose-Dependent Prevention of Metabolic and Neurologic Disease in Murine MPS II by ZFN-Mediated In Vivo Genome Editing.* **Laoharawee, K., et al.** 2018, Molecular Therapy, 26(4): 1127-1136.

28. *Carbonate-sensitive phytotransferrin controls high-affinity iron uptake in diatoms.* **McQuaid, J. B., et al.** 2018, Nature, 555(7697): 534.

29. *Recombinase-mediated cassette exchange (RMCE)-A rapidly-expanding toolbox for targeted genomic modifications.* **Turan, S., et al.** 2013, Gene, 515(1): 1-27.

30. *Dynamic, structural, and regulatory aspects of lambda site-specific recombination.* **Landy, A.** 1989, Annual review of biochemistry, 58(1): 913-941.

31. *The mechanism of conservative site-specific recombination.* **Craig, N. L.** 1988, Annual review of genetics, 22(1): 77-105.

32. *Expression of a site-specific endonuclease stimulates homologous recombination in mammalian cells.* **Rouet, P., Smih, F. and Jasin, M.** 1994, Proceedings of the National Academy of Sciences, 91(13): 6064-6068.

33. *Site-specific DNA recombination in mammalian cells by the Cre recombinase of bacteriophage P1.* **Sauer, B. and Henderson, N.** 1988, Proceedings of the National Academy of Sciences, 85(14): 5166-5170.

34. *A phage integrase directs efficient site-specific integration in human cells.* **Groth, A. C., et al.** 2000, Proceedings of the National Academy of Sciences, 97(11): 5995-6000.

35. *The origin and behavior of mutable loci in maize.* **McClintock, B.** 1950, Proceedings of the National Academy of Sciences, 36(6): 344-355.

36. *Transposon-mediated mutagenesis of a baculovirus.* **Fraser, M. J., et al.** 1985, Virology, 145(2): 356-361.

37. *Transposable element in fish.* **Koga, A., et al.** 1996, Nature, 383(6595): 30.

38. *Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells.* **Ivics, Z., et al.** 1997, Cell, 91: 501–510.

39. *Alzheimer's disease-related amyloid-β induces synaptotoxicity in human iPS cell-derived neurons.* **Nieweg, K., et al.** 2015, Cell death & disease, 6(4): e1709.

40. *Human iPS cell-derived dopaminergic neurons function in a primate Parkinson's disease model.* **Kikuchi, T., et al.** 2017, Nature, 548(7669): 592-596.

41. *Induced pluripotent stem cells used to reveal drug actions in a long QT syndrome family with complex genetics.* **Terrenoire, C., et al.** 2013, The Journal of general physiology, 141(1): 61-72.

42. *A small molecule screen in stem-cell-derived motor neurons identifies a kinase inhibitor as a candidate therapeutic for ALS.* **Yang, Y. M., et al.** 2013, Cell Stem Cell., 12: 713–726.

43. *Characterization of human induced pluripotent stem cell-derived retinal pigment epithelium cell sheets aiming for clinical application.* **Kamao, H., et al.** 2014, Stem cell reports, 2(2): 205-218.

44. *Expansion and patterning of cardiovascular progenitors derived from human pluripotent stem cells.* **Birket, M. J., et al.** 2015, Nature biotechnology, 33(9): 970.

45. *Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors.* **Takahashi, K. and Yamanaka, S.** 2006, Cell, 126(4): 663-676.

46. *Induction of pluripotent stem cells from adult human fibroblasts by defined factors.* **Takahashi, K., et al.** 2007, Cell, 131(5): 861-872.

47. *Embryo-derived stem cells: of mice and men.* **Smith, A. G.** 2001, Annual review of cell and developmental biology, 17(1): 435-462.

48. *The promise of induced pluripotent stem cells in research and therapy.* **Robinton, D. A. and Daley, G. Q.** 2012, Nature, 481(7381): 295.

49. *Embryonic Stem Cell Lines Derived from Human Blastocysts.* **Thomson, A., et al.** 1998, Science, 282(5391): 1145-1147.

50. *iPS cells: a game changer for future medicine.* **Inoue, H., et al.** 2014, The EMBO journal, 33(5): 409-417.

51. *Treatment of sickle cell anemia mouse model with iPS cells generated from autologous skin.* **Hanna, J., et al.** 2007, Science, 318(5858): 1920-1923.

52. *Characterization of three loci for homologous gene targeting and transgene expression.* **Eyquem, J., et al.** 2013, Biotechonology and Bioengineer, 110(8): 2225-2235.

53. *Disruption of overlapping transcripts in the ROSA βgeo 26 gene trap strain leads to widespread expression of β-galactosidase in mouse embryos and hematopoietic cells.* **Zambrowicz, B. P., et al.** 1997, Proceedings of the National Academy of Sciences, 94(8): 3789-3794.

54. *Functional genomics, proteomics, and regulatory DNA analysis in isogenic settings using zinc finger nuclease-driven transgenesis into a safe harbor locus in the human genome.* **DeKelver, R. C., et al.** 2010, Genome research, 20(8): 1133-1142.

55. *Safe harbours for the integration of new DNA in the human genome.* **Sadelain, M., Papapetrou, E. P. and Bushman, F. D.** 2012, Nature reviews Cancer, 12(1): 51-58.

56. *Characterization of a preferred site on human chromosome 19q for integration of adeno-associated virus DNA by non-homologous recombination.* **Kotin, R. M., Linden, R. M. and Berns, K. I.** 1992, The EMBO Journal, 11: 5071–5078.

57. *The role of mutant CCR5 allele in HIV-1 transmission and disease progression.* **Huang, Y., et al.** 1996, Nature medicine, 2(11): 1240-1243.

58. *Identification and targeting of the ROSA26 locus in human embryonic stem cells.* **Irion, S., et al.** 2007, Nature biotechnology, 25(12): 1477-1482.

59. *DICE, an efficient system for iterative genomic editing in human pluripotent stem cells.* **Zhu, F., et al.** 2013, Nucleic acids research, 42(5): e34-e34.

60. *Precise Correction of Disease Mutations in Induced Pluripotent Stem Cells Derived From Patients With Limb Girdle Muscular Dystrophy.* **Turan, S., et al.** 2016, Molecular therapy, 24(4): 685-696.

61. *High-efficiency transformation of mammalian cells by plasmid DNA.* **Chen, C. and Okayama, H.** 1987, Molecular and cellular biology, 7(8): 2745-2752.

62. *A cDNA cloning vector that permits expression of cDNA inserts in mammalian cells.* **Okayama, H. and Berg, P.** 1983, Molecular and Cellular Biology, 3(2): 280-289.

63. *Duplexes of 21-nucleotide RNAs mediate RNA interference in cultured mammalian cells.* **Elbashir, S. M., et al.** 2001, Nature, 411(6836): 494.

64. *Delivery technologies for genome editing.* **Yin, H., Kauffman, K. J. and Anderson, D. G.** 2017, Nature Reviews Drug Discovery, 16(6): 387.

65. *Generation and comparison of CRISPR-Cas9 and Cre-mediated genetically engineered mouse models of sarcoma.* **Huang, J., et al.** 2017, Nature communications, 8: 15999.

66. *Viral Cre-LoxP tools aid genome engineering in mammalian cells.* **Sengupta, R., et al.** 2017, Journal of Biological Engineering, 11(1): 45.

67. *A multi-landing pad DNA integration platform for mammalian cell engineering.* **Gaidukov, L., et al.** 2018, Nucleic acids research, 46(8): 4072-4086.

68. *A guide to genome engineering with programmable nucleases.* **Kim, H. and Kim, J.S.** 2014, Nature Reviews Genetic, 15: 321–334.

69. *Loss of Tctn3 causes neuronal apoptosis and neural tube defects in mice.* **Wang, B., et al.** 2018, Cell death & disease, 9(5): 520.

70. *Endogenous DNA damage in humans: a review of quantitative data.* **De Bont, R. and Van Larebeke, N.** 2004, Mutagenesis, 19(3): 169-185.

71. *DNA damage and the balance between survival and death in cancer biology.* **Roos, W. P., Thomas, A. D. and Kaina, B.** 2016, Nature Reviews Cancer, 16(1): 20.

72. *The complexity of DNA damage: relevance to biological consequences.* **Ward, J. F.** 1994, International journal of radiation biology, 66(5): 427-432.

73. *The DNA-damage response in human biology and disease.* **Jackson, S. P. and Bartek, J.** 2009, Nature, 461(7267): 1071-1078.

74. *Partners and pathways: repairing a double-strand break.* **Haber, J. E.** 2000, Trends in Genetics, 16(6): 259-264.

75. *The mechanism of vertebrate nonhomologous DNA end joining and its role in V (D) J recombination.* **Lieber, M. R., et al.** 2004, DNA repair, 3(8): 817-826.

76. *V(D)J recombination: double-strand break repair gene products used in the joining mechanism.* **Weaver, D., et al.** 1995, Annals of the New York Academy of Sciences, 764(1): 99-111.

77. *Homologous recombination and non-homologous end-joining pathways of DNA double-strand break repair have overlapping roles in the maintenance of chromosomal integrity in vertebrate cells.* **Takata, M., et al.** 1998, The EMBO journal, 17(18): 5497-5508.

78. *The double-strand-break repair model for recombination.* **Szostak, J. W., et al.** 1983, Cell, 33(1): 25-35.

79. *Homologous integration in mammalian cells without target gene selection.* **Jasin, M. and Berg, P.** 1988, Genes & development, 2(11): 1353-1363.

80. **Kucherlapati, R. and Smith, G.** *Genetic Recombination.* Washington, DC : American Society for Microbiology, 1988.

81. *Prospects for homologous recombination in human gene therapy.* **Vega, M. A.** 1991, Human Genetics, 87(3): 245–253.

82. *Insertion of DNA sequences into the human chromosomal β-globin locus by homologous recombination.* **Smithies, O., et al.** 1985, Nature, 317(6034): 230.

83. *The length of homology required for gene targeting in embryonic stem cells.* **Hasty, P., Rivera-Perez, J. and Bradley, A.** 1991, Molecular and cellular biology, 11(11): 5586-5591.

84. *Reexamination of gene targeting frequency as a function of the extent of homology between the targeting vector and the target locus.* **Deng, C. and Capecchi, M. R.** 1992, Molecular and cellular biology, 12(8): 3365-3371.

85. *Analysis of the role of homology arms in gene-targeting vectors in human cells.* **Shii, A., et al.** 2014, PloS one, 9(9): e108236.

86. *The minimum amount of homology required for homologous recombination in mammalian cells.* **Rubnitz, J. and Subramani, S. U. R. E. S. H.** 1984, Molecular and cellular biology, 4(11): 2253-2258.

87. *Dependence of intrachromosomal recombination in mammalian cells on uninterrupted homology.* **Waldman, A. S. and Liskay, R. M.** 1988, Molecular and cellular biology, 8(12): 5350-5357.

88. *High-throughput engineering of the mouse genome coupled with high-resolution expression analysis.* **Valenzuela, D. M., et al.** 2003, Nature biotechnology, 21(6): 65.

89. *Multi-kilobase homozygous targeted gene replacement in human induced pluripotent stem cells.* **Byrne, S. M., et al.** 2014, Nucleic acids research, 43(3): e21.

90. *Analysis of gene targeting and intrachromosomal homologous recombination stimulated by genomic double- strand breaks in mouse embryonic stem cells.* **Donoho, G. P., Jasin, M. and Berg, P.** 1998, Molecular and cellular biology, 18(7): 4070-4078.

91. *Effects of homology length and donor vector arrangement on the efficiency of double-strand break-mediated recombination in human cells.* **Phillips, J. E. and Calos., and M. P.** 1999, Somatic cell and molecular genetics, 25(2): 91-100.

92. *Introduction of double-strand breaks into the genome of mouse cells by expression of a rare-cutting endonuclease.* **Rouet, P., Smih, F. and Jasin, M.** 1994, Molecular and cell biology, 14(12): 8096–8106.

93. *Zinc-finger nuclease-driven targeted integration into mammalian genomes using donors with limited chromosomal homology.* **Orlando, S. J., et al.** 2010, Nucleic acids research, 38(15): e152-e152.

94. *Donor DNA utilization during gene targeting with zinc-finger nucleases.* **Beumer, K. J., et al.** 2013, G3: Genes,Genomes, Genetics, 3(4): 657-664.

95. *Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining.* **Maruyama, T., et al.** 2015, Nature biotechnology, 33(5): 538.

96. *Homology arms of targeting vectors for gene insertions and CRISPR/Cas9 technology: size does not matter; quality control of targeted clones does.* **Petrezselyova, S., et al.** 2015, Cellular and Molecular Biology Letters, 20(5): 773-787.

97. *CRISPR/Cas9 β-globin gene targeting in human haematopoietic stem cells.* **Dever, D. P., et al.** 2016, Nature, 539(7629): 384.

98. *Non-viral vectors for gene-based therapy.* **Yin, H., et al.** 2014, Nature Reviews Genetics, 15(8): 541-555.

99. *Origins of Programmable Nucleases for Genome Engineering.* **Chandrasegaran, S. and Carroll, D.** 2016, Journal of Molecular Biology, 428(5): 963–989.

100. *FokI dimerization is required for DNA cleavage.* **Bitinaite, J., et al.** 1998, Proceedings of the national academy of sciences, 95(18): 10570-10575.

101. *Zinc finger-DNA recognition: crystal structure of a Zif268-DNA complex at 2.1 A.* **Pavletich, N.P. and Pabo, C.O.** 1991, Science, 252(5007): 809–817.

102. *Expanding the repertoire of target sites for zinc finger nuclease-mediated genome modification.* **Wilson, K. A., et al.** 2013, Molecular Therapy-Nucleic Acids, 2: e88.

103. *Genome editing with engineered zinc finger nucleases.* **Urnov, F. D., et al.** 2010, Nature Reviews Genetics, 11(9): 636-464.

104. *Nuclease target site selection for maximizing on-target activity and minimizing off-target effects in genome editing.* **Lee, C. M., et al.** 2016, Molecular Therapy, 24(3): 475-487.

105. *Selection-free zinc-finger-nuclease engineering by context-dependent assembly (CoDA).* **Sander, J.D., et al.** 2011, Nature methods, 8(1): 67.

106. *Synergy between adjacent zinc fingers in sequence-specific DNA recognition.* **Isalan, M., Choo, Y. and Klug, A.** 1997, Proceedings of the National Academy of Sciences, 94(11): 5617–5621.

107. *ZFN-site searches genomes for zinc finger nuclease target sites and off-target sites.* **Cradick, T. J., et al.** 2011, Bmc Bioinformatics, 12(1): 152.

108. *Chimeric nucleases stimulate gene targeting in human cells.* **Porteus, M.H. and Baltimore, D.** 2003, Science, 300(5620): 763-763.

109. *Highly efficient endogenous human gene correction using designed zinc-finger nucleases.* **Urnov, F.D., et al.** 2005, Nature, 435(7042): 646–651.

110. *An improved zinc-finger nuclease architecture for highly specific genome editing.* **Miller, J. C., et al.** 2007, Nature biotechnology, 25(7): 778-785.

111. *Structure-based redesign of the dimerization interface reduces the toxicity of zinc-finger nucleases.* **Szczepek, M., et al.** 2007, Nature biotechnology, 25(7): 786-793.

112. *Creating designed zinc-finger nucleases with minimal cytotoxicity.* **Ramalingam, S., et al.** 2011, Journal of molecular biology, 405(3): 630-641.

113. *TAL effectors: function, structure, engineering and applications.* **Mak, A. N. S., et al.** 2013, Current opinion in structural biology, 23(1): 93-99.

114. *Efficient design and assembly of custom TALEN and other TAL effector-based constructs for DNA targeting.* **Cermak, T., et al.** 2011, Nucleic acids research, 39(12): e82.

115. *A simple cipher governs DNA recognition by TAL effectors.* **Moscou, M. J. and Bogdanove, A. J.** 2009, Science, 326(5959): 1501-1501.

116. *Xanthomonas AvrBs3 family-type III effectors: discovery and function.* **Boch, J. and Bonas, U.** 2010, Annual review of phytopathology, 48: 419-436.

117. *Targeting G with TAL effectors: a comparison of activities of TALENs constructed with NN and NK repeat variable di-residues.* **Christian, M. L., et al.** 2012, PloS one, 7(9): e45383.

118. *Optimized tuning of TALEN specificity using non-conventional RVDs.* **Juillerat, A., et al.** 2015, Scientific reports, 5: 8150.

119. *Highly efficient generation of heritable zebrafish gene mutations using homo-and heterodimeric TALENs.* **Cade, L., et al.** 2012, Nucleic acids research, 40(16): 8001-8010.

120. *Genetic engineering of human pluripotent cells using TALE nucleases.* **Hockemeyer, D., et al.** 2011, Nature biotechnology, 29(8): 731-734.

121. *Efficient and specific modifications of the Drosophila genome by means of an easy TALEN strategy.* **Liu, J., et al.** 2012, Journal of genetics and genomics, 39(5): 209-215.

122. *Comparing zinc finger nucleases and transcription activator-like effector nucleases for gene targeting in Drosophila.* **Beumer, K. J., et al.** 2013, G3: Genes, Genomes, Genetics, 3(10): 1717-1725.

123. *An efficient strategy for TALEN-mediated genome engineering in Drosophila.* **Katsuyama, T., et al.** 2013, Nucleic acids research, 41(17): e163.

124. *Targeted gene disruption in somatic zebrafish cells using engineered TALENs.* **Sander, J. D., et al.** 2011, Nature biotechnology, 29(8): 697-698.

125. *Heritable gene targeting in zebrafish using customized TALENs.* **Huang, P., y otros, y otros.** 2011, Nature biotechnology, 29(8): 699-700.

126. *TALEN-mediated precise genome modification by homologous recombination in zebrafish.* **Zu, Y., et al.** 2013, Nature methods, 10(4): 329-331.

127. *Knockout rats generated by embryo microinjection of TALENs.* **Tesson, L., et al.** 2011, Nature biotechnology, 29(8): 695-696.

128. *Knockout mice created by TALEN-mediated gene targeting.* **Sung, Y. H., et al.** 2013, Nature biotechnology, 31(1): 23-24.

129. *High-efficiency and heritable gene targeting in mouse by transcription activator-like effector nucleases.* **Qiu, Z., et al.** 2013, Nucleic acids research, 41(11): e120-e120.

130. *One-step generation of mice carrying a conditional allele together with an HA-tag insertion for the delta opioid receptor.* **Su, D., et al.** 2017, Scientific reports, 7: 44476.

131. *TALEN-mediated gene mutagenesis in rhesus and cynomolgus monkeys.* **Liu, H., et al.** 2014, Cell stem cell, 14(3): 323-328.

132. *Gene editing of CCR5 in autologous CD4 T cells of persons infected with HIV.* **Tebas, P., et al.** 2014, New England Journal of Medicine, 370(10): 901-910.

133. *Genome-wide Specificity of Highly Efficient TALENs and CRISPR/Cas9 for T Cell Receptor Modification.* **Knipping, F., et al.** 2017, Molecular Therapy-Methods & Clinical Development, 4: 213-224.

134. *Novel impact of the DNMT3A R882H mutation on GSH metabolism in a K562 cell model established by TALENs.* **Yang, L., et al.** 2017, Oncotarget, 8(18): 30395–30409.

135. *Precise correction of the dystrophin gene in duchenne muscular dystrophy patient induced pluripotent stem cells by TALEN and CRISPR-Cas9.* **Li, H. L., et al.** 2015, Stem cell reports, 4(1): 143-154.

136. *TALEN-mediated functional correction of human iPSC-derived macrophages in context of hereditary pulmonary alveolar proteinosis.* **Kuhn, A., et al.** 2017, Scientific Reports, 7(1): 15195.

137. *A TALEN genome-editing system for generating human stem cell-based disease models.* **Ding, Q., et al.** 2013, Cell stem cell, 12(2): 238-251.

138. *TALENs: a widely applicable technology for targeted genome editing.* **Joung, J. K. and Sander, J. D.** 2013, Nature Review Molecular Cell Biology,14(1): 49–55.

139. *E-TALEN: a web tool to design TALENs for genome engineering.* **Heigwer, F., et al.** 2013, Nucleic acids research, 41(20): e190-e190.

140. *SAPTA: a new design tool for improving TALE nuclease activity.* **Lin, Y., et al.** 2014, Nucleic acids research, 42(6): e47-e47.

141. *Broad specificity profiling of TALENs results in engineered nucleases with improved DNA-cleavage specificity.* **Guilinger, J. P., et al.** 2014, Nature methods, 11(4): 429.

142. *Non-RVD mutations that enhance the dynamics of the TAL repeat array along the superhelical axis improve TALEN genome editing efficacy.* **Tochio, N., et al.** 2016, Scientific reports, 6: 37887.

143. *Repeating pattern of non-RVD variations in DNA-binding modules enhances TALEN activity.* **Sakuma, T., et al.** 2013, Scientific Reports, 3: 3379.

144. *CRISPR provides acquired resistance against viruses in prokaryotes.* **Barrangou, R., et al.** 2007, Science, 315(5819): 1709-1712.

145. *A programmable dual-RNA–guided DNA endonuclease in adaptive bacterial immunity.* **Jinek, M., et al.** 2012, Science, 1225829.

146. *Cas9–crRNA ribonucleoprotein complex mediates specific DNA cleavage for adaptive immunity in bacteria.* **Gasiunas, G., et al.** 2012, Proceedings of the National Academy of Sciences, 109(39): e2579-e2586.

147. *Short motif sequences determine the targets of the prokaryotic CRISPR defence system.* **Mojica, F. J., et al.** 2009, Microbiology, 155: 733-740.

148. *Genome editing with RNA-guided Cas9 nuclease in Zebra fish embryos.* **Chang, N., et al.** 2013, Cell Research. 23(4): 465-472.

149. *High-throughput gene targeting and phenotyping in zebrafish using CRISPR/Cas9.* **Varshney, G. K., et al.** 2015, Genome research, 25(7): 1030-1042.

150. *Genome engineering of Drosophila with the CRISPR RNA-guided Cas9 nuclease.* **Gratz, S. J., et al.** 2013, Genetics, 194(4): 1029-1035.

151. *CRISPR-Cas9 Genome Editing in Drosophila.* **Gratz, S. J., et al.** 2015, Current protocols in molecular biology, 111(1): 31-2.

152. *Heritable gene targeting in the mouse and rat using a CRISPR-Cas system.* **Li, D., et al.** 2013, Nature biotechnology, 31(8): 681-683.

153. *In vivo CRISPR/Cas9 gene editing corrects retinal dystrophy in the S334ter-3 rat model of autosomal dominant retinitis pigmentosa.* **Bakondi, B., et al.** 2016, Molecular Therapy, 24(3): 556-563.

154. *One-step generation of mice carrying mutations in multiple genes by CRISPR/Cas-mediated genome engineering.* **Wang, H., et al.** 2013, Cell, 153(4): 910-918.

155. *Efficient generation of Rosa26 knock-in mice using CRISPR/Cas9 in C57BL/6 zygotes.* **Chu, V. T., et al.** 2016, BMC biotechnology, 16(1): 4.

156. *Generation of gene-modified cynomolgus monkey via Cas9/RNA-mediated gene targeting in one-cell embryos.* **Niu, Y., et al.** 2014, Cell, 156(4): 836-843.

157. *Functional disruption of the dystrophin gene in rhesus monkey using CRISPR/Cas9.* **Chen, Y., et al.** 2015, Human molecular genetics, 24(13): 3764-3774.

158. *Targeted genome engineering in human cells with the Cas9 RNA-guided endonuclease.* **Cho, S. W., et al.** 2013, Nature Biotechnology, 31(3): 230–232.

159. *Multiplex genome engineering using CRISPR/Cas systems.* **Cong, L., et al.** 2013, Science, 339: 819–823.

160. *High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells.* **Fu, Y., et al.** 2013, Nature Biotechnology, 31(9): 822-826.

161. *High-throughput profiling of off-target DNA cleavage reveals RNA-programmed Cas9 nuclease specificity.* **Pattanayak, V., et al.** 2013, Nature biotechnology, 31(9): 839-843.

162. *Improving CRISPR–Cas specificity with chemical modifications in single-guide RNAs.* **Ryan, D.E, et al.** 2018, Nucleic Acids Research, 46(2): 792–803.

163. *Methods for optimizing CRISPR–Cas9 genome editing specificity.* **Tycko, J., Myer, V. E. and Hsu, P. D.** 2016, Molecular cell, 63(3): 355-370.

164. *Partial DNA-guided Cas9 enables genome editing with reduced off-target activity.* **Yin, H., et al.** 2018, Nature chemical biology, 14(3): 311.

165. *Rapid and highly efficient mammalian cell engineering via Cas9 protein transfection.* **Liang, X., et al.** 2015, Journal of biotechnology, 208: 44-53.

166. *Highly efficient RNA-guided genome editing in human cells via delivery of purified Cas9 ribonucleoproteins.* **Kim, S., et al.** 2014, Genome research, 24(6): 1012-1019.

167. *Improving CRISPR-Cas nuclease specificity using truncated guide RNAs.* **Fu, Y., et al.** 2014, Nature biotechnology, 32(3): 279-284.

168. *Digenome-seq: genome-wide profiling of CRISPR–Cas9 off-target effects in human cells.* **Kim, D., et al.** 2015, Nature Methods, 12 : 237 –243.

169. *Enhanced proofreading governs CRISPR–Cas9 targeting accuracy.* **Chen, J.S., et al.** 2017, Nature, 550(7676): 407-410.

170. *Cas-OFFinder: a fast and versatile algorithm that searches for potential off-target sites of Cas9 RNA-guided endonucleases.* **Bae, S., Park, J. and Kim, J. S.** 2014, Bioinformatics, 30(10): 1473-1475.

171. *CasOT: a genome-wide Cas9/gRNA off-target searching tool.* **Xiao, A., et al.** 2014, Bioinformatics, 30(8): 1180-1182.

172. *CRISPRdirect: software for designing CRISPR/Cas guide RNA with reduced off-target sites.* **Naito, Y., et al.** 2014, Bioinformatics, 31(7): 1120-1123.

173. *Identification of Pre-Existing Adaptive Immunity to Cas9 Proteins in Humans.* **Charlesworth, C. T., et al.** 2018, bioRxiv, 243345.

174. *Catalysis by site-specific recombinases.* **Stark, W. M., Boocock, M. R. and Sherratt, D. J.** 1992, Trends in Genetic, 8(12): 432–439.

175. *The integrase family of site-specific recombinases: regional similarities and global diversity.* **Argos, P., et al.** 1986, The EMBO journal, 5(2): 433-440.

176. *Diversity in the serine recombinases.* **Smith, M. and Thorpe, H. M.** 2002, Molecular microbiology, 44(2): 299-307.

177. **Hatfull, G. F. and Grindley, N. D. F.** Resolvases and DNA-invertases: a family of enzymes active in site-specific recombination. [book auth.] R. Kucherlapati and G. R. Smith. *in Genetic recombination.* Washington, D.C : American Society for Microbiology, 1988, pp. 357–396.

178. **Van Duyne, G. D.** A structural view of tyrosine recombinase site-specific recombination. [book auth.] N.L. Craig, et al. *In Mobile DNA II.* Washington, D.C. : American Society of Microbiology., 2002, pp. 93-117.

179. **Grindley, N.D.F.** The movement of Tn3-like elements. Transposition and cointegrate resolution. [book auth.] N.L. Craig, et al. *In Mobile DNA II.* Washington, D.C. : American Society of Microbiology, 2002, pp. 172-302.

180. **Johnson, R. C.** Bacterial site-specific DNA inversion systems. [book auth.] N.R. Craig, et al. *In Mobile DNA II.* Washington, D.C. : American Society of Microbiology, 2002, pp. 230-271.

181. *Site-specific recombination: integration, excision, resolution, and inversion of defined DNA segments.* **Nash, H.A.** 1996, Escherichia coli and Salmonella: cellular and molecular biology, 2: 2363-2376.

182. *Flp recombinase promotes site-specific DNA recombination in embryonic stem cells and transgenic mice.* **Dymecki, S. M.** 1996, Proceedings of the National Academy of Sciences, 93(12): 6191-6196.

183. *Dre recombinase, like Cre, is a highly efficient site-specific recombinase in E. coli, mammalian cells and mice.* **Anastassiadis, K., et al.** 2009, Disease models & mechanisms, 2(9-10): 508-515.

184. **Liu, C.** *Strategies for designing transgenic DNA constructs in Disease. In Lipoproteins and Cardiovascular.* Totowa, NJ. : Humana Press, 2013. pp. 183-201.

185. *Genome organization and characterization of mycobacteriophage Bxb1.* **Mediavilla, J., et al.** 2000, Molecular microbiology, 38(5): 955-970.

186. *The orientation of mycobacteriophage Bxb1 integration is solely dependent on the central dinucleotide of attP and attB.* **Ghosh, P., Kim, A. I. and Hatfull, G. F.** 2003, Molecular cell, 12(5): 1101-1111.

187. *Accuracy and efficiency define Bxb1 integrase as the best of fifteen candidate serine recombinases for the integration of DNA into the human genome.* **Xu, Z., et al.** 2013, BMC biotechnology, 13(1): 87.

188. **Campbell, A.** General aspects of lysogeny. [book auth.] R. Calendar. *In The Bacteriophages.* Oxford, N.Y. : Oxford University Press, 2006, pp. 66-73.

189. *Analysis of the integration function of the Streptomycete bacteriophage phiC31.* **Kuhstoss, S. and Rao, R.N.** 1991, Journal Molecular Biology, 222: 897–908.

190. *Control of directionality in the site-specific recombination system of the Streptomyces phage φC31.* **Thorpe, H. M., Wilson, S. E. and Smith, M.** 2000, Molecular microbiology, 38(2): 232-241.

191. *Rapid metabolic pathway assembly and modification using serine integrase site-specific recombination.* **Colloms, S. D., et al.** 2013, Nucleic acids research, 42(4): e23-e23.

192. *The ins and outs of serine integrase site-specific recombination.* **Rutherford, K. and Van Duyne, G. D.** 2014, Current opinion in structural biology, 24: 125-131.

193. *Architecture of the Hin Synaptic Complex during Recombination: The Recombinase Subunits Translocate with the DNA Strands.* **Dhar, G., Sanders, E. R. and Johnson, R. C.** 2004, Cell, 119(1): 33-45.

194. *Site-specific recombination by ΦC31 integrase and other large serine recombinases.* **Smith, M. C., et al.** 2010, Biochemical Society Transactions, 38(2): 388-394.

195. *A phage protein that binds ϕC31 integrase to switch its directionality.* **Khaleel, T., et al.** 2011, Molecular microbiology, 80(6): 1450-1463.

196. *A motif in the C-terminal domain of ϕC31 integrase controls the directionality of recombination.* **Rowley, P. A., et al.** 2008, Nucleic acids research, 36(12): 2879-3891.

197. *DNA binding and synapsis by the large C-terminal domain of ϕC31 integrase.* **McEwan, A. R, Rowley, P. A. and Smith, M. C.** 2009, Nucleic acids research, 37(14): 4764-4773.

198. *Attachment site recognition and regulation of directionality by the serine integrases.* **Rutherford, K., et al.** 2013, Nucleic acids research, 41(17): 8341–8356.

199. **Azaro, M.A. and Landy, A.** λ integrase and the λ Int family. [book auth.] N.L. Craig, et al. *In Mobile DNA II.* Washington, D.C. : America Society of Microbiology, 2002, pp. 118–148.

200. *Gated rotation mechanism of site-specific recombination by ϕC31 integrase.* **Olorunniji, F. J., et al.** 2012, Proceedings of the National Academy of Sciences, 109(48): 19661-19666.

201. *Mammalian genomes contain active recombinase recognition sites.* **Thyagarajan, B., et al.** 2000, Gene, 244(1): 47-54.

202. *Site-specific genomic integration produces therapeutic Factor IX levels in mice.* **Olivares, E. C., et al.** 2002, Nature biotechnology, 20(11): 1124.

203. *Long-term phenotypic correction in factor IX knockout mice by using phiC31 integrase-mediated gene therapy.* **Keravala, A., et al.** 2011, Gene therapy, 18(8): 842.

204. *Long-term and efficient expression of human β-globin gene in a hematopoietic cell line using a new site-specific integrating non-viral system.* **Dormiani, K., et al.** 2015, Gene therapy, 22(8): 663.

205. *Dystrophin expression in host muscle following transplantation of muscle precursor cells modified with the phiC31 integrase.* **Quenneville, S. P., et al.** 2007, Gene therapy, 14(6): 514.

206. *Genomic integration of the full-length dystrophin coding sequence in Duchenne muscular dystrophy induced pluripotent stem cells.* **Farruggio, A. P., et al.** 2017, Biotechnology journal, 12(4): 1600477.

207. *Tetrameric structure of a serine integrase catalytic domain.* **Yuan, P., Gupta, K. and Van Duyne, G. D.** 2008, Structure, 16(8): 1275-1286.

208. *Enhanced efficiency through nuclear localization signal fusion on phage ϕC31-integrase: activity comparison with Cre and FLPe recombinase in mammalian cells.* **Andreas, S., et al.** 2002, Nucleic acids research, 30(11): 2299-2306.

209. *Role of the N-terminal domain of fC31 integrase in attB-attP synapsis.* **Rowley, P.A. and Smith, M.C.** 2008, Journal Bacteriology, 190: 6918–6921.

210. *Mutational derivatives of PhiC31 integrase with increased efficiency and specificity.* **Keravala, A., et al.** 2009, Molecular therapy, 17(1): 112-120.

211. *Site-Specific Genomic Integration in Mammalian Cells Mediated by Phage phiC31 Integrase.* **Thyagarajan, B., et al.** 2001, Molecular and cellular biology, 21(12): 3926-3934.

212. *Screening of potential pseudo att sites of Streptomyces phage ΦC31 integrase in the human genome.* **Hu, Z. P., et al.** 2013, Acta Pharmacologica Sinica, 34(4): 561.

213. *Adjusting the attB site in donor plasmid improves the efficiency of ΦC31 integrase system.* **Xie, F., et al.** 2012, DNA and cell biology, 31(7): 1335-1340.

214. *Molecular analysis of chromosomal rearrangements in mammalian cells after phiC31-mediated integration.* **Ehrhardt, A., et al.** 2006, Human gene therapy, 17(11): 1077-1094.

215. *Cell type differences in activity of the Streptomyces bacteriophage phiC31 integrase.* **Maucksch, C., et al.** 2008, Nucleic acids research, 36(17): 5462-5471.

216. *Critical Amino Acid Residues Within the ϕC31 Integrase DNA-Binding Domain Affect Recombination Activities in Mammalian Cells.* **Liesner, R., et al.** 2010, Human gene therapy, 21(9): 1104-118.

217. *DAXX interacts with phage ΦC31 integrase and inhibits recombination.* **Chen, J. Z., et al.** 2006, Nucleic acids research, 34: 6298-6304.

218. *Control of phage Bxb1 excision by a novel recombination directionality factor.* **Ghosh, P., Wasil, L.R. and Hatfull, G.F.** 2006, PLoS Biology, 4: e186.

219. *Efficient reversal of phiC31 integrase recombination in mammalian cells.* **Farruggio, A. P., et al.** 2012, Biotechnology journal, 7(11): 1332-1336.

220. *Treatment of mouse limb ischemia with an integrative hypoxia-responsive vector expressing the vascular endothelial growth factor gene.* **Yasumura, E. G., et al.** 2012, PLoS One, 7(3): e33944.

221. *α-l-iduronidase gene-based therapy using the phiC31 system to treat mucopolysaccharidose type I mice.* **Stilhano, R. S., et al.** 2015, The journal of gene medicine, 17(1-2): 1-13.

222. *Recombinase-mediated reprogramming and dystrophin gene addition in mdx mouse induced pluripotent stem cells.* **Zhao, C., et al.** 2014, PLoS One, 9(4): e96279.

223. *A platform for rapid prototyping of synthetic gene networks in mammalian cells.* **Duportet, X., et al.** 2014, Nucleic acids research, 42(21): 13440-13451.

224. *Synapsis in phage Bxb1 integration: selection mechanism for the correct pair of recombination sites.* **Ghosh, P., Pannunzio, N. R. and Hatfull, G. F.** 2005, Journal of molecular biology, 349(2): 331-348.

225. *A diversity of serine phage integrases mediate site-specific recombination in mammalian cells.* **Keravala, A., et al.** 2006, Molecular genetics and genomics, 276(2): 135.

226. *Two-step site selection for serine-integrase-mediated excision: DNA-directed integrase conformation and central dinucleotide proofreading.* **Ghosh, P., Bibb, L. A. and Hatfull, G. F.** 2008, Proceedings of the National Academy of Sciences, 105(9): 3238-3243.

227. *A novel system for simultaneous or sequential integration of multiple gene-loading vectors into a defined site of a human artificial chromosome.* **Suzuki, T., et al.** 2014, PloS one, 9(10): e110404.

228. *Eukaryotic transposable elements and genome evolution.* **Finnegan, D. J.** 1989, Trends Genetic, 5: 103–107.

229. *Retrotransposons revisited: the restraint and rehabilitation of parasites .* **Goodier, J. L. and Kazazian Jr, H. H.** 2008, Cell, 135(1): 23-35.

230. *Molecular evolution and tempo of amplification of human LINE-1 retrotransposons since the origin of primates.* **Khan, H., Smit, A. and Boissinot, S.** 2006, Genome research, 16(1): 78-87.

231. *SINEs: short interspersed repeated elements of the eukaryotic genome.* **Okada, N.** 1991, Trends in ecology & evolution, 6(11): 358-361.

232. *The diversity of LTR retrotransposons.* **Havecker, E. R., Gao, X. and Voytas, D. F.** 2004, Genome biology, 5(6): 225.

233. *Sleeping beauty transposition: biology and applications for molecular therapy.* **Izsvák, Z. and Ivics, Z.** 2004, Molecular Therapy, 9(2): 147-156.

234. *Initial sequencing and analysis of the human genome.* **Lander, E. S., et al.** 2001, Nature, 409(6822): 860-921.

235. *International Human Genome Sequencing. Initial sequencing and analysis of the human genome.* **Consortium.** 2001, Nature, 409: 860–921.

236. *B-globin sleeping beauty transposon reduces red blood cell sickling in a patient-derived CD34+-based in vitro model.* **Sjeklocha, L. M., et al.** 2013, PloS one, 8(11): e80403.

237. *Seamless gene correction of beta-thalassemia mutations in patient-specific iPSCs using CRISPR/Cas9 and piggyBac.* **Xie, F., et al.** 2014, Genome research, 24(9): 1526-1533.

238. *Hyperactive sleeping beauty transposase enables persistent phenotypic correction in mice and a canine model for hemophilia B.* **Hausl, M. A., et al.** 2010, Molecular Therapy, 18(11): 1896-1906.

239. *Ultrasound-targeted hepatic delivery of factor IX in hemophiliac mice.* **Anderson, C. D., et al.** 2016, Gene therapy, 23(6): 510.

240. *Pax3-induced expansion enables the genetic correction of dystrophic satellite cells.* **Filareto, A., et al.** 2015, Skeletal muscle, 5(1): 36.

241. *piggyBac transposons expressing full-length human dystrophin enable genetic correction of dystrophic mesoangioblasts.* **Loperfido, M., et al.** 2015, Nucleic acids research, 44(2): 744-760.

242. *A transposon and transposase system for human application.* **Hackett, P. B., Largaespada, D. A. and Cooper, L. J.** 2010, Molecular therapy, 18(4): 674-683.

243. *Suicidal autointegration of sleeping beauty and piggyBac transposons in eukaryotic cells.* **Wang, Y., et al.** 2014, PLoS Genetics, 10(3): e1004103.

244. *piggyBac internal sequences are necessary for efficient transformation of target genomes.* **Li, X., et al.** 2005, Insect molecular biology, 14(1): 17-30.

245. *Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice.* **Ding, S., et al.** 2005, Cell, 122(3): 473-483.

246. *PiggyBac transposon-mediated gene transfer in human cells.* **Wilson, M. H., Coates, C. J. and George Jr, A. L.** 2007, Molecular therapy, 15(1): 139-145.

247. *A hyperactive piggyBac transposase for mammalian applications.* **Yusa, K., et al.** 2011, Proceedings of the National Academy of Sciences, 108(4): 1531-1536.

248. *High-resolution genome-wide mapping of transposon integration in mammals.* **Yant, S. R., et al.** 2005, Molecular and cellular biology, 25(6): 2085-2094.

249. *Resident aliens: the Tc1/mariner superfamily of transposable elements.* **Plasterk, R. H., Izsvák, Z. and Ivics, Z.** 1999, Trends in genetics, 15(8): 326-332.

250. *Structure–function analysis of the inverted terminal repeats of the Sleeping Beauty transposon.* **Cui, Z., et al.** 2002, Journal of molecular biology, 318(5): 1221-1235.

251. *Excision of Sleeping Beauty transposons: parameters and applications to gene therapy.* **Liu, G., et al.** 2004, The journal of gene medicine, 6(5): 574-583.

252. *Gene transfer efficiency and genome-wide integration profiling of Sleeping Beauty, Tol2, and piggyBac transposons in human primary T cells.* **Huang, X., et al.** 2010, Molecular Therapy, 18(10): 1803-1813.

253. *The hyperactive Sleeping Beauty transposase SB100X improves the genetic modification of T cells to express a chimeric antigen receptor.* **Jin, Z., et al.** 2011, Gene therapy, 18(9): 849.

254. *Germline transgenic pigs by Sleeping Beauty transposition in porcine zygotes and targeted integration in the pig genome.* **Garrels, W., et al.** 2011, PloS one, 6(8): e23573.

255. *Molecular evolution of a novel hyperactive Sleeping Beauty transposase enables robust stable gene transfer in vertebrates.* **Mátés, L., et al.** 2009, Nature genetics, 41(6): 753-761.

256. *Biomedically relevant circuit-design strategies in mammalian synthetic biology.* **Bacchus, W., Aubel, D. and Fussenegger, M.** 2013, Molecular systems biology, 9(1): 691.

257. *Sexually mature individuals of Xenopus laevis from the transplantation of single somatic nuclei.* **Gurdon, J. B., Elsdale, T. R. and Fischberg, M.** 1958, Nature, 182(4627): 64-65.

258. *Viable offspring derived from fetal and adult mammalian cells.* **Wilmut, I., et al.** 1997, Nature, 385(6619): 810.

259. *Establishment in culture of pluripotential cells from mouse embryos.* **Evans, M. J. and Kaufman, M. H.** 1981, Nature, 292(5819): 154-156.

260. *Isolation of a pluripotent cell line from early mouse embryos cultured in medium conditioned by teratocarcinoma stem cells.* **Martin, G. R.** 1981, Proceedings of the National Academy of Sciences, 78: 7634-7638.

261. *Nuclear reprogramming of somatic cells by in vitro hybridization with ES cells.* **Tada, M., et al.** 2001, Current Biology, 11(19): 1553-1558.

262. *Nuclear reprogramming of somatic cells after fusion with human embryonic stem cells.* **Cowan, C. A., et al.** 2005, Science, 309(5739): 1369-1373.

263. *Expression of a single transfected cDNA converts fibroblasts to myoblasts.* **Davis, R. L., Weintraub, H. and Lassar, A. B.** 1987, Cell, 51(6): 987-1000.

264. *GATA-1 reprograms avian myelomonocytic cell lines into eosinophils, thromboblasts, and erythroblasts.* **Kulessa, H., Frampton, J. and Graf, T.** 1995, Genes & development, 9(10): 1250-1262.

265. *Induction of ectopic eyes by targeted expression of the eyeless gene in Drosophila.* **Halder, G., Callaerts, P. and Gehring, W. J.** 1995, Science, 267(5205): 1788-1792.

266. *Stepwise reprogramming of B cells into macrophages.* **Xie, H., et al.** 2004, Cell, 117(5): 663-676.

267. *Reprogramming of human somatic cells to pluripotency with defined factors.* **Park, I. H., et al.** 2008, Nature, 451(7175): 141.

268. *Induced pluripotent stem cell lines derived from human somatic cells.* **Yu, J., et al.** 2007, Science, 318(5858): 1917-1920.

269. *Efficient and rapid generation of induced pluripotent stem cells from human keratinocytes.* **Aasen, T., et al.** 2008, Nature Biotechnology, 26(11): 1276-1284.

270. *A decade of transcription factor-mediated reprogramming to pluripotency.* **Takahashi, K. and Yamanaka, S.** 2016, Nature Reviews Molecular Cell Biology, 17(3): 183.

271. *Reprogramming of pancreatic β cells into induced pluripotent stem cells.* **Stadtfeld, M., Brennand, K., and Hochedlinger, K.** 2008, Current Biology, 18(12): 890-894.

272. *Pluripotent stem cells induced from adult neural stem cells by reprogramming with two factors.* **Kim, J. B., et al.** 2008, Nature, 454(7204): 646.

273. *Reprogramming of peripheral blood cells to induced pluripotent stem cells.* **Staerk, J., et al.** 2010, Cell stem cell, 7(1): 20.

274. *Induction of pluripotent stem cells from primary human fibroblasts with only Oct4 and Sox2.* **Huangfu, D., et al.** 2008, Nature biotechnology, 26(11): 1269.

275. *Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts.* **Nakagawa, M., et al.** 2008, Nature biotechnology, 26(1): 101.

276. *Generation of induced neuronal cells by the single reprogramming factor ASCL1.* **Chanda, S., et al.** 2014, Stem cell reports, 3(2): 282-296.

277. *NKX3-1 is required for induced pluripotent stem cell reprogramming and can replace OCT4 in mouse and human iPSC induction.* **Mai, T., et al.** 2018, Nature cell biology, 20(8): 900.

278. *Reprogramming fibroblasts to pluripotency using arginine-terminated polyamidoamine nanoparticles based non-viral gene delivery system.* **Zhu, K., et al.** 2014, International journal of nanomedicine, 9: 5837.

279. *Reprogramming cell fates by small molecules.* **Ma, X., Kong, L. and Zhu, S.** 2017, Protein & cell, 8(5): 328-348.

280. *Methods for making induced pluripotent stem cells: reprogramming a la carte.* **González, F., Boué, S. and Belmonte, J. C. I.** 2011, Nature Reviews Genetics, 12(4): 231.

281. *Gene therapy: trials and tribulations.* **Somia, N. and Verma, I. M.** 2000, Nature Reviews Genetics, 1(2): 91.

282. *Progress and problems with the use of viral vectors for gene therapy.* **Thomas, C. E., Ehrhardt, A. and Kay, M. A.** 2003, Nature reviews genetics, 4(5): 346-358.

283. *The use of DNA viruses as vectors for gene therapy.* **Ali, M., Lemoine, N.R. and Ring, C. J.** 1994, Gene Therapy, 1(6): 367–384.

284. *Retroviruses.* **Varmus, H.** 1988, Science, 240: 1427–1435.

285. *Integration of murine leukemia virus DNA depends on mitosis.* **Roe, T., et al.** 1993, The EMBO journal, 12(5): 2099-2108.

286. *Retroviral vectors: from cancer viruses to therapeutic tools.* **Miller, A. D.** 2014, Hum Gene Therapy, 25: 989–994.

287. *A third-generation lentivirus vector with a conditional packaging system.* **Dull, T., et al.** 1998, Journal of virology, 72(11): 8463-8471.

288. *Clinical use of lentiviral vectors.* **Milone, M. C. and O'Doherty, U.** 2018, Leukemia, 1.

289. *Isolation of a cytopathogenic agent from human adenoids undergoing spontaneous degeneration in tissue culture.* **Rowe, W. P., et al.** 1953, Proceedings of the Society for Experimental Biology and Medicine, 84 (3): 570-573.

290. *Adenovirusassociated defective virus particles.* **Atchison, R. W., Casto, B. C. and Hammon, W. M.** 1965, Science, 149(3685): 754–755.

291. **Carter, B. J. and Laughlin, C. A.** Adeno-associated virus defectiveness and the nature of the adenovirus helper function. *In The parvoviruses.* Boston, M.A. : Springer, 1984, pp. 67-128.

292. **Muzyczka, N.** Use of adeno-associated virus as a general transduction vector for mammalian cells. *In Viral expression vectors.* Berlin, Heidelberg. : Springer, 1992, pp. 97-129.

293. *Herpes simplex virus types 1 and 2 completely help adenovirus-associated virus replication.* **Buller, R. M., et al.** 1981, Journal of virology, 40(1): 241-247.

294. *Effect of genome size on AAV vector packaging.* **Wu, Z., Yang, H. and Colosi, P.** 2010, Molecular Therapy, 18(1): 80-86.

295. *Nanoparticle orientation to control RNA loading and ligand display on extracellular vesicles for cancer regression.* **Pi, F., et al.** 2018, Nature nanotechnology, 13(1): 82.

296. *Synergistically enhanced selective intracellular uptake of anticancer drug carrier comprising folic acid-conjugated hydrogels containing magnetite nanoparticles.* **Kim, H., et al.** 2017, Scientific reports, 7: 41090.

297. *Toxicity of cationic lipids and cationic polymers in gene delivery.* **Lv, H., et al.** 2006, Journal of Controlled Release, 114(1): 100-109.

298. *Factors affecting the clearance and biodistribution of polymeric nanoparticles.* **Alexis, F., et al.** 2008, Molecular pharmaceutics, 5(4): 505-515.

299. *Gene transfer into muscle by electroporation in vivo.* **Aihara, H. and Miyazaki, J. I.** 1998, Nature biotechnology, 16(9): 867.

300. *Gene transfer into mouse lyoma cells by electroporation in high electric fields.* **Neumann, E., et al.** 1982, The EMBO journal, 1(7): 841-845.

301. *Transfection of mammalian cells with plasmid DNA by scrape loading and sonication loading.* **Fechheimer, M., et al.** 1987, Proceedings of the National Academy of Sciences, 84(23): 8463-8467.

302. *Sonication: a new method for gene transfer to plants.* **Joersbo, M. and Brunstedt, J.** 1992, Physiologia Plantarum, 85(2): 230-234.

303. *Magnetofection: enhancing and targeting gene delivery by magnetic force in vitro and in vivo.* **Scherer, F., et al.** 2002, Gene therapy, 9(2): 102.

304. *Superparamagnetic iron oxide: pharmacokinetics and toxicity.* **Weissleder, R. A., et al.** 1989, American Journal of Roentgenology, 152(1): 167-173.

305. *APOBEC3A intratumoral DNA electroporation in mice.* **Kostrzak, A., et al.** 2017, Gene therapy, 24(2): 74.

306. *Non viral vectors in gene therapy-an overview.* **Ramamoorth, M. and Narvekar, A.** 2015, Journal of clinical and diagnostic research, 9(1): GE01.

307. *Mechanism of DNA release from cationic liposome/DNA complexes used in cell transfection.* **Xu, Y. and Szoka, F. C.** 1996, Biochemistry, 35(18): 5616-5623.

308. *Structure of DNA-cationic liposome complexes: DNA intercalation in multilamellar membranes in distinct interhelical packing regimes.* **Rädler, J. O., et al.** 1997, Science, 275(5301): 810-814.

309. *Receptor-mediated gene delivery and expression in vivo.* **Wu, G. Y. and Wu, C. H.** 1988, Journal of Biological Chemistry, 263(29): 14621-14624.

310. *Design and development of polymers for gene delivery.* **Pack, D. W., et al.** 2005, Nature Reviews. Drug Discov., 4: 581–593.

311. *Chemical vectors for gene delivery: a current review on polymers, peptides and lipids containing histidine or imidazole as nucleic acids carriers.* **Midoux, P., et al.** 2009, British journal of pharmacology, 157(2): 166-178.

312. *Towards safe, non-viral therapeutic gene expression in humans.* **Glover, D.J., Lipps, H.J. and Jans, D.A.** 2005, Nature Reviews Genetic, 6(4): 299–310.

313. *Synthetic DNA delivery systems.* **Luo, D. and Saltzman, W. M.** 2000, Nature biotechnology, 18(1): 33.

314. *Gene insertion into genomic safe harbors for human gene therapy.* **Papapetrou, E. P. and Schambach, A.** 2016, Molecular Therapy, 24(4): 678-684.

315. *Robust, persistent transgene expression in human embryonic stem cells is achieved with AAVS1-targeted integration.* **Smith, J.R., et al.** 2008, Stem Cells, 26(2): 496-504.

316. *Human cardiovascular progenitor cells develop from a KDR+ embryonic-stem-cellderived population.* **Yang, L., et al.** 2008, Nature, 453(7194): 524-528.

317. *Efficient recombinase-mediated cassette exchange at the AAVS1 locus in human embryonic stem cells using baculoviral vectors.* **Ramachandra, C.J., et al.** 2011, Nucleic acids research, 39(16): e107-e107.

318. *Efficient recombinase-mediated cassette exchange in hPSCs to study the hepatocyte lineage reveals AAVS1 locus-mediated transgene inhibition.* **Ordovás, L., et al.** 2015, Stem Cell Reports, 5: 918-931.

319. *Rhesus iPSC safe harbor gene-editing platform for stable expression of transgenes in differentiated cells of all germ layers.* **Hong, S. G., et al.** 2017, Molecular therapy, 25(1): 44-53.

320. *Site-specific integration and tailoring of cassette design for sustainable gene transfer.* **Lombardo, A., et al.** 2011, Nature methods, 8(10): 861-869.

321. *Studies on the propagation in vitro of poliomyelitis viruses: IV. Viral multiplication in a stable strain of human malignant epithelial cells (strain HeLa) derived from an epidermoid carcinoma of the cervix.* **Scherer, W. F., Syverton, J. T. and Gey, G. O.** 1953, Journal of Experimental Medicine, 97(5): 695-710.

322. **Dravid, G., Hammond, H. and Cheng, L.** *Culture of human embryonic stem cells on human and mouse feeder cells. In Human Embryonic Stem Cell Protocols.* s.l. : Humana Press, 2006. pp. 91-104.

323. *Promoting human embryonic stem cell renewal or differentiation by modulating Wnt signal and culture conditions.* **Cai, L., et al.** 2007, Cell research, 17(1): 62-72.

324. *Defining the role of Wnt/beta-catenin signaling in the survival, proliferation, and self-renewal of human embryonic stem cells.* **Dravid, G., et al.** 2005, Stem cells, 23(10): 1489-1501.

325. *Generation of mouse and human induced pluripotent stem cells (iPSC) from primary somatic cells.* **Lorenzo, I. M., Fleischer, A. and Bachiller, D.** s.l. : Stem Cell Reviews and Reports, 9(4): 435-450, 2013.

326. *Generation of two induced pluripotent stem cell (iPSC) lines from p. F508del Cystic Fibrosis patients.* **Fleischer, A., et al.** 2018, Stem cell research, 29: 1-5.

327. **Sambrook J, Fritschi EF and Maniatis T.** Molecular cloning: a laboratorymanual. New York. : Cold Spring Harbor Laboratory Press, 1989.

328. *A method for the isolation and serial propagation of keratinocytes, endothelial cells, and fibroblasts from a single punch biopsy of human skin.* **Normand, J. and Karasek, M.A.** 1995, In Vitro Cell Dev Biol Anim, 31(6): 447-455.

329. *A ROCK inhibitor permits survival of dissociated human embryonic stem cells.* **Watanabe, K., et al.** 2007, Nature biotechnology, 25(6): 681.

330. *Characterization of the chicken β-globin insulator.* **Chung, J. H., Bell, A. C. and Felsenfeld, G.** 1997, Proceedings of the National Academy of Sciences, 94(2): 575-580.

331. *Complete nucleotide sequence of SV40 DNA.* **Fiers, W., et al.** 1978, Nature, 273(5658): 113.

332. *Firefly luciferase gene: structure and expression in mammalian cells.* **De Wet, J. R., et al.** 1987, Molecular and cellular biology, 7(2): 725-737.

333. *A One Pot, One Step, Precision Cloning Method with High Throughput Capability.* **Engler, C., Kandzia, R. and Marillonnet, S.** 2008, PloS one, 3(11): e3647.

334. *Lymphoid regeneration from gene-corrected SCID-X1 subject-derived iPSCs.* **Menon, T., et al.** 2015, Cell stem cell, 16(4): 367-372.

335. *TALENs facilitate targeted genome editing in human cells with high specificity and low cytotoxicity.* **Mussolino, C., et al.** 2014, Nucleic acids research, 42(10): 6767-6773.

336. *Enhanced effciency of human pluripotent stem cell genome editing through replacing TALENs with CRISPRs.* **Ding, Q., et al.** 2013, Cell Stem Cell, 12(4): 339.

337. *Efficient genome engineering by targeted homologous recombination in mouse embryos using transcription activator-like effector nucleases.* **Sommer, D., et al.** 2014, Nature communications, 5: 3045.

338. *Highly efficient biallelic genome editing of human ES/iPS cells using a CRISPR/Cas9 or TALEN system.* **Takayana, K., et al.** 2017, Nucleic acids reasearch, 45 (9): 5198-5207.

339. *A large-scale in vivo analysis reveals that TALENs are significantly more mutagenic than ZFNs generated using context-dependent assembly.* **Chen, S., et al.** 2013 , Nucleic Acids Research, 41(4): 2769–2778.

340. *A chromatin landmark and transcription initiation at most promoters in human cells.* **Guenther, M. G., et al.** 2007, Cell, 130(1): 77-88.

341. *Control of serine integrase recombination directionality by fusion with the directionality factor.* **Olorunniji, F. J., et al.** 2017, Nucleic acids research, 45(14): 8635-8645.

342. *Nuclear envelope and genome interaction in cell fate.* **Talamas, J.A. and Capelson, M.** 2015, Frontier in genetics, 6(95).

343. *Integration specifi city of phage phiC31 integrase in the human genome.* **Chalberg, T. W., et al.** 2006, Journal of molecular biology, 357(1): 28-48.

344. *Creation of engineered human embryonic stem cell lines using phiC31 integrase.* **Thyagarajan, B., et al.** 2008, Stem cells, 26(1): 119-126.

345. *Improved site-specific recombinase-based method to produce selectable marker-and vector-backbone-free transgenic cells.* **Yu, Y., et al.** 2014, Scientific reports, 4: 4240.

346. *A direct comparison of two nonviral gene therapy vectors for somatic integration: in vivo evaluation of the bacteriophage integrase ϕC31 and the Sleeping Beauty transposase.* **Ehrhardt, A., et al.** 2005, Molecular Therapy, 11(5): 695-706.

347. *Switching the polarity of a bacteriophage integration system.* **Smith, M. C., Till, R. and Smith, M. C.** 2004, Molecular Microbiology, 51(6): 1719-1728.

348. *A platform for functional assessment of large variant libraries in mammalian cells.* **Matreyek, K.A., Stephany, J.J. and Fowler, D.M.** 2017, Nucleic Acids Research, 45 (11): e102.

349. *Insulators and boundaries: versatile regulatory elements in the eukaryotic genome.* **Bell, A. C., West, A. G. and Felsenfeld, G.** 2001, Science, 291(5503): 447-450.

350. *Chromatin insulators: linking genome organization to cellular function.* **Phillips-Cremins, J. E. and Corces, V. G.** 2013, Molecular cell, 50(4): 461-474.

351. *A position-effect assay for boundaries of higher order chromosomal domains.* **Kellum, R. and Schedl, P.** 1991, Cell, 64(5): 941-950.

352. *Putting boundaries on silence.* **Sun, F. L. and Elgin, S. C.** 1999, Cell, 99(5): 459-462.

353. *Loss of transcriptional activity of a transgene is accompanied by DNA methylation and histone deacetylation and is prevented by insulators.* **Pikaart, M. J., Recillas-Targa, F. and Felsenfeld, G.** 1998, Genes & development, 12(18): 2852-2862.

354. *Insulators: many functions, many mechanisms.* **West, A. G., Gaszner, M. and Felsenfeld, G.** 2002, Genes & development, 16(3): 271-288.

355. *Sea urchin insulator protects lentiviral vector from silencing by maintaining active chromatin structure.* **Hino, S., et al.** 2004, Gene therapy, 11(10): 819.

356. *Position-independent human β-globin gene expression mediated by a recombinant adeno-associated virus vector carrying the chicken β-globin insulator.* **Inoue, T., et al.** 1999, Journal of human genetics, 44(3): 152.

357. *A chromatin insulator protects retrovirus vectors from chromosomal position effects.* **Emery, D. W., et al.** 2000, Proceedings of the National Academy of Sciences, 97(16): 9150-9155.

358. *cHS4 insulator-mediated alleviation of promoter interference during cell-based expression of tandemly associated transgenes.* **Yahata, K., et al.** 2007, Journal of molecular biology, 374(3): 580-890.

359. *Chromatin insulator elements block transgene silencing in engineered human embryonic stem cell lines at a defined chromosome 13 locus.* **MacArthur, C. C., et al.** 2011, Stem cells and development, 21(2): 191-205.

360. *The impact of cHS4 insulators on DNA transposon vector mobilization and silencing in retinal pigment epithelium cells.* **Sharma, N., et al.** 2012, PloS one, 7(10): e48421.

361. *Progress and prospects: the design and production of plasmid vectors.* **Gill, D. R., Pringle, I. A. and Hyde, S. C.** 2009, Gene Therapy, 16(2): 165-171.

362. *Increased persistence of lung gene expression using plasmids containing the ubiquitin C or elongation factor 1α promoter.* **Gill, D. R., et al.** 2001, Gene therapy, 8(20): 1539-1546.

363. *Site-specific integrase-mediated transgenesis in mice via pronuclear injection.* **Tasic, B., et al.** 2011, Proceedings of the National Academy of Sciences, 108(19): 7902-7907.

364. *Comparative analysis of transposable element vector systems in human cells.* **Grabundzija, I., et al.** 2010, Molecular therapy, 18(6): 1200-1209.

365. *Chromosomal transposition of PiggyBac in mouse embryonic stem cells.* **Wang, W., et al.** 2008, Proceedings of the National Academy of Sciences, 105(27): 9290-9295.

366. *Targeted gene correction of alpha(1)-antitrypsin deficiency in induced pluripotent stem cells.* **Yusa, K., et al.** 2011, Nature, 478: 391–394.

367. *Excision efficiency is not strongly coupled to transgenic rate: cell type-dependent transposition efficiency of sleeping beauty and piggyBac DNA transposons.* **Kolacsek, O., et al.** 2014, Human gene therapy methods, 25(4): 241-252.

368. *Sleeping Beauty, a wide host-range transposon vector for genetic transformation in vertebrates.* **Izsvák, Z., Ivics, Z. and Plasterk, R. H.** 2000, Journal of molecular biology, 302(1): 93-102.

369. *A modular open platform for systematic functional sutdies under physiological conditions.* **Mulholland, C.B., et al.** 2015, Nucleic acids reseach, 43(17): e112-e112.

370. *A novel Bxb1 integrase RMCE system for high fidelity site-specific integration of mAb expression cassette in CHO Cells.* **Inniss, M. C., et al.** 2017, Biotechnology and bioengineering, 114(8): 1837-1846.

371. *Site-specific cassette exchange and germline transmission with mouse ES cells expressing phiC31 integrase.* **Belteki, G., et al.** 2003, Nature Biotechnology, 21(3): 321.

372. *PhiC31 integrase facilitates genetic approaches combining multiple recombinases.* **Monetti, C., et al.** 2011, Methods, 53(4): 380-385.

373. *The concept of synthetic lethality in the context of anticancer therapy.* **Kaelin Jr, W. G.** 2005, Nature reviews cancer, 5(9): 689.

374. *Predicting selective drug targets in cancer through metabolic networks.* **Folger, O., et al.** 2011, Molecular systems biology, 7(1): 501.

375. *Global Metabolic Engineering of Glycolytic Pathway via Multicopy Integration in Saccharomyces cerevisiae.* **Yamada, R., Wakita, K. and Ogino, H.** 2017, ACS synthetic biology, 6(4): 659-666.

376. *Efficient generation of multi-copy strains for optimizing secretory expression of porcine insulin precursor in yeast Pichia pastoris.* **Zhu, T., et al.** 2009, Journal of applied microbiology, 107(3): 954-963.

377. *Long-term phenotypic correction in factor IX knockout mice by using phiC31 integrase-mediated gene therapy.* **Keravala, A., et al.** 2011, Gene therapy, 18(8): 842.