DISSERTATION

Genetic Modification of Stem Cells Utilizing S/MAR DNA Vectors

Submitted to the
Combined Faculties for the Natural Sciences
and for Mathematics
of the Ruperto-Carola University
of Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

Alicia Roig-Merino

June 2018

Dissertation
submitted to the
Combined Faculties for the Natural Sciences and for Mathematics
of the Ruperto-Carola University of Heidelberg, Germany
for the degree of
Doctor of Natural Sciences

presented by

Diplom-Biologin Alicia Roig-Merino
Born in Barcelona, Spain
Oral examination: 28th June 2018

Genetic Modification of Stem Cells Utilizing S/MAR DNA Vectors

Referees: Prof. Dr Christof von Kalle

Dr Richard P. Harbottle

Examining commission

Prof. Dr Christof von Kalle (First examiner)

Prof Dr Ana Martin-Villalba (Chair)

Dr Richard Harbottle (Second examiner)

Dr Michael Milsom (External examiner)

STATEMENT OF AUTHORSHIP

The content of this for this PhD thesis is my original work, and it has been generated at the German Cancer Research Center (Deutsches Krebsforschungszentrum, DKFZ) in Heidelberg, at the Research Group "DNA Vectors, F160", from 1st October 2014 until 28th June 2018 and was funded by the Helmholtz International Graduate School for Cancer Research (DKFZ) and the DNA Vector Lab budget.

Principal Investigator: Dr. Richard Paul Harbottle
DNA Vectors (F160)
Deutsches Krebsforschungszentrum (DKFZ)
Im Neuenheimer Feld 242
69120 Heidelberg
Tel: +49 6221 424978
E-Mail: r.harbottle@dkfz.de

First Examiner: Prof. Dr. Christof von Kalle
Translational Oncology (G100)
Deutsches Krebsforschungszentrum (DKFZ)
Im Neuenheimer Feld 460
69120 Heidelberg
Tel: +49 6221 56 6990
E-Mail: christof.kalle@NCT-Heidelberg.de

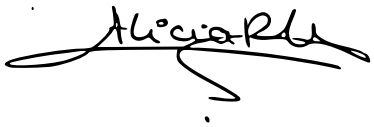
I hereby declare that

1) this thesis has been composed by myself under supervision and describes my own work unless otherwise specified.
2) all references and direct citations have been quoted, and all sources of information have been specifically acknowledged.

Date and place

Heidelberg, 19.04.2018

Signature



ACKNOWLEDGEMENTS

I would like to take the opportunity to express my gratitude to all who have helped and encouraged me during the process of my PhD and the writing of this thesis.

The completion of this work could not have been possible without the participation, help and assistance of so many people whose names may not all be enumerated. Your contributions are sincerely appreciated and gratefully acknowledged.

However, I would like to express my deep appreciation and thank particularly the following:

First and foremost, I would like to thank Dr Richard Harbottle, for his supervision, help, guidance and support, in the professional and personal, during the whole duration of my thesis. I want to express my most sincere gratitude for having believed in me from the very beginning and never stopped doing so, even when I did. I also would like to thank you for the English grammar, vocabulary and phonetics lessons; although I am probably never going to stop saying /təˈmeɪtoʊ/ instead of /təˈmɑːtoʊ/.

Second, I would like to thank my Thesis Advisors and examiners: Prof. Dr Christof von Kalle, Prof. Dr Tristan Mckay, Dr Michael Milsom for their fruitful advice, valuable input, criticism, scientific discussions and direct supervision. Thank you Tris for hosting me in your lab and Mick for your and your lab's collaboration (and infinite source of reagents).

A huge thank you goes to my colleagues, for always facilitating a nice working environment. To my big lab brother Prof. Dr. Bozza, our friendship is as old as the length of this PhD. We travelled the world to conferences, to Italy and we even drove to London with a van; always with the finest gin in our hands. Thanks for your scientific advice, criticism and the most stringent "Bozza Test". To my partner in crime and experiments (and dot detective) Manuela, your motivation and enthusiasm drove us to have the craziest ideas and the experiments I enjoyed the most. You are the "fresh air" in the lab and I hope you will never stop being it. It has been a pleasure sharing my experience and knowledge with you. Tamara, thanks for the emotional and spiritual advice, your words are wise. Sveta, thanks for the suggestions, criticism and sharing your sometimes super-long (but effective) protocols with me.

And of course, to all the former members of the lab and specially the technicians Matthias Ehrbar, Terence Osere and Rita Hörlein and the students who contributed to this work: Bojana, Julia and Alex.

None of this work would have been possible without the acquired knowledge in Stem Cells and their handling, which I gained thanks to Dr. Paul Kaschutnig, Dr. Elena Senis-Herrero, Marleen Büchler, Sina Stäble and Louise Bullen. Thank you Sascha Dehler and Prof. Dr. Ana Martin-Villalba for helping with the stereotactic injections. Thanks Domenic Hartmann and Dr. Jeaninne Lacroix for your help with the mice subcutaneous injections.

It has been also a pleasure to closely work and collaborate with DKFZ core facilities, in particular with Franciscus van der Hoeven and Ulrich Kloz, from the Transgenic Service, who very enthusiastically believed in the project and helped generating the transgenic animals, and Dr. Karin Müller-Decker, Steffi Laier and Andrea Pohl-Arnlod, from the Tumor Modelling unit, who helped with the animal work. Not to forget Dr. Irene Gil-Farina for her collaboration and help with the integration analysis.

I would like to thank the animal technicians and in particular Tatjana Boysen, who did a great job taking care of the animals; and not less important, 'the little furries', for their unconditional collaboration and without which, half of this work would have not been possible.

On the personal side, I would like to thank my parents Montserrat Merino and Robert Roig for their unconditional support. Without it, I would not be here today.

Also a special thanks to my friends, who have been a big support for good and bad times, especially Pere Gil Alba, Barbara Vodicska and Bouchra Tawk. Thanks for the lunches, the coffees and the counselling sessions.

I thank the swing community, my dancing partner Jan Sauer and our always motivated "Swing Pipettes". To my Fofinhas do Swing, Elena, Mireia and Gerhard, thanks for all the fun, the "cotis" and laughs during classes.

I also want to thank those who walked along with me during this past years, and for those who did not, I thank them for the lessons I learned and for making me stronger.

Finally, I would particularly like to thank Valentin Kruspel for being my pillar and for his unconditional support. Thank you for always listening, supporting me and for being there through good and especially bad times. You keep me on track and my feet on the ground.

PUBLICATIONS AND AWARDS

Published presentations

- Roig-Merino, A., Bozza, M., Bullen, L., Büchler, M., Stäble, S., van der Hoeven, F., et al. (2018). ***Genetic modification of Stem Cells utilizing non-integrative and autonomously replicating DNA vectors.*** 21st Annual meeting American Society of Gene and Cell Therapy.

- Roig-Merino, A., Bozza, M., Büchler, M., Stäble, S., van der Hoeven, F., Williams, J., et al. (2018). ***A novel non-integrating non-viral DNA Vector for the persistent genetic modification of embryonic and hematopoietic stem cells.*** 21st Annual meeting American Society of Gene and Cell Therapy.

- Roig-Merino, A., Bozza, M., Dehler, S., Hartmann, D., Lacroix, J., Martin-Villalba, A., & Harbottle, R. (2017). ***Persistent genetic modification of dividing and differentiating cells with non-integrative and autonomously replicating DNA vectors.*** *20th Annual meeting American Society of Gene and Cell Therapy*.

- Roig-Merino, A., Bozza, M., Hartmann, D., Lacroix, J., & Harbottle, R. (2016). ***Genetic modification of dividing and differentiating cells with non- integrative and autonomously replicating DNA vectors.*** 22*nd Annual Meeting of the German Society for Gene Therapy DG-GT*.

- Roig-Merino, A., & Harbottle, R. (2016). **Genetic modification of stems cells using DNA vectors**. 21st Annual DKFZ PhD retreat.

Awards

- PhD stipend. October 2014 – October 2017. DKFZ

- Short-term research fellowship. January-March 2017. DKFZ

SUMMARY

Pluripotent stem cells are considered a prime source of cells for regenerative therapies and gene therapy applications because of their extensive proliferation, the potential for self-renewal and their capability for multi-lineage differentiation. A great advantage of **induced pluripotent stem cells (iPSCs)** is their derivation from a patient's somatic cells, which can be isolated using non-invasive techniques, thus eliminating not only ethical concerns associated with embryonic stem cells but also the risk of immune rejection. Therefore, iPSCs are an attractive tool for personalised medicine, drug screening and to generate disease models. Typically, the modification of pluripotent cells is done by using **integrating viral vectors**. Although vectors based on modified viruses are unquestionably the most effective gene delivery systems in use today, their efficacy at gene transfer is, however, tempered by their potential integration and genotoxicity. **Non-viral DNA vectors** are attractive alternatives to viral gene delivery systems because of their low toxicity, relatively easy production and great versatility. However, their efficiency is still regarded as below the requirements for realistic *in vivo* gene therapy due to deficient delivery exacerbated by the merely transient gene expression of plasmid DNA *in vivo*.

Thus, the development of safer, more efficient and easily and economically prepared persistently expressing genetic vectors remains one of the main strategic tasks of gene therapy research and is the crucial prerequisite for its successful clinical application. An ideal vector for the genetic modification of cells should deliver sustainable therapeutic levels of gene expression without compromising the viability of the host in any way. Permanently maintained, episomal and autonomously replicating DNA vectors, which comprise entirely human elements, might provide the most suitable method for achieving these goals.

This thesis presents the development of a non-viral, non-integrating and autonomously replicating DNA vector system based on the use of a Scaffold Matrix Associated Region (S/MAR), for the persistent **genetic modification** of differentiating and dividing cells, including but not limited to **murine and human Stem Cells (SCs).** Although this DNA Vector is among the best of its class, one of its limitations is that as it is produced in bacteria it comprises a large proportion of bacterial sequences which are unnecessary and undesirable for clinical application. Accordingly, the vector system has been refined, updated and all aspects of its functionality have been improved whilst also reducing its impact on cells following its delivery, resulting in higher levels of more sustained expression than previous versions. Molecular and genetic analysis of S/MAR-labelled cells revealed that the vectors are kept at low copy

numbers, are present in their episomal forms and do not modify or genetically damage the cells or their progeny, as the cells fully retain their pluripotent capabilities and are able to generate chimeric mice. This new vector system is also used to generate iPSCs from murine or patient-derived fibroblasts.

For the first time, this work shows that genetic modification with this DNA vector system provides robust transgene expression which is sustained through the reprogramming and differentiation process *in vitro* and *in vivo*.

ZUSAMMENFASSUNG

Pluripotente Stammzelllen werden wegen ihres beträchtlichen Wachstums, ihrer Fähigkeit der Selbsterneuerung und ihrem Vermögen sich in verschiedene Linien zu differenzieren als wichtigste Zellquelle für regenerative Therapien und Gentherapie-Anwendungen angesehen. Ein großer Vorteil von **induzierten pluripotenten Stammzellen (iPSZ/iPSCs)** ist, dass sie von einer somatischen Zelle eines Patienten stammen, welche durch nicht-invasive Methoden isoliert werden können. Daher werden mit ihrer Hilfe nicht nur das Risiko einer Immunabstoßung, sondern auch ethische Bedenken, welche mit embryonalen Stammzellen einhergehen, umgangen. Aufgrund dessen sind iPSCs attraktive Werkzeuge für die personalisierte Medizin, Medikamentenscreens und für die Herstellung von Krankheitsmodellen. Üblicherweise werden pluripotente Zellen durch **integrierende virale Vektoren** verändert. Während Vektoren, die auf modifizierten Viren basieren zweifellos die effizientesten Übertragungssysteme heutzutage sind, wird der Vorteil ihrer Leistungsfähigkeit beim Gentransfer durch ihre potentielle Integrierung und Genotoxizität gemindert. **Nicht-virale DNA Vektor**en sind aufgrund ihrer geringen Toxizität, relativ leichter Herstellung und großer Vielseitigkeit attraktive Alternativen zu viralen Gentransport-Systemen. Allerdings wird ihre Effizienz aufgrund von unzureichender Lieferung und obendrein lediglich transienter Genexpression von Plasmid DNA *in vivo* immer noch als zu gering erachtet um für realistische *in vivo* Gentherapie in Frage zu kommen.

Daher bleibt die Entwicklung von sichereren, effizienteren und leichter und wirtschaftlicher hergestellten anhaltend exprimierenden Genvektoren eine der strategischen Hauptaufgaben der Gentherapie-Forschung und ist eine wesentliche Voraussetzung für ihre erfolgreiche klinische Anwendung. Ein idealer Vektor für die genetische Veränderung von Zellen sollte nachhaltige therapeutische Mengen an Genexpression liefern, ohne die Viabilität der Wirtszelle in irgendeiner Form zu beeinträchtigen. Dauerhaft erhaltene, episomale und autonom replizierende DNA Vektoren welche ausschließlich humane Elemente enthalten, könnten die geeignetste Methode sein um dieses Ziel zu erreichen.

Diese Arbeit zeigt die Entwicklung eines nicht-viralen, nicht-integrierenden und autonom replizierenden **DNA Vektoren Systems,** welches auf der Verwendung einer **Scaffold Matrix Associated Region (S/MAR)** basiert, für die beständige **genetische Modifizierung** von sich differenzierenden und sich teilenden Zellen, einschließlich, aber nicht beschränkt auf **murine und humane Stammzellen (SZ/SCs)**. Obwohl dieser DNA Vektor einer der Besten seiner Art ist, ist eine seiner Limitierungen,

dass er aufgrund seiner Produktion in Bakterien zu einem Großteil aus bakteriellen Sequenzen besteht welche für die klinische Anwendung unnötig und nicht wünschenswert sind. Dementsprechend wurde das Vektorensystem verfeinert, aktualisiert und alle Aspekte seiner Funktionalität verbessert, während sein Einfluss auf die Zellen nach dem Transport reduziert wurde, was im Vergleich zu früheren Versionen in höheren Levels von nachhaltigerer Expression resultierte. Molekulare und genetische Analysen von S/MAR-gelabelten Zellen haben gezeigt, dass die Vektoren in niedriger Kopienzahl behalten werden, in ihrer episomalen Form vorliegen und die Zellen und ihre Nachfolger nicht verändern oder genetisch beschädigen. Das neue Vektorensystem wurde außerdem verwendet um iPSCs aus murinen oder Patientenfibroblasten zu generieren.

Diese Arbeit zeigt zum ersten Mal, dass die genetische Modifizierung mit diesem DNA Vektorensystem robuste Transgenexpression liefert, welche während des gesamten Reprogrammierungs- und Differenzierungsprozess *in vitro* und *in vivo* aufrechterhalten bleibt.

CONTENTS

ABBREVIATIONS

Chromosomal elements

2A	2A self-cleavage peptide sequence
CHS4	Insulating element Chicken Hypersensitive Site-4
EBNA1	Epstein-Barr Nuclear Antigen 1
EI40	Insulating Element 40
IRES	Internal Ribosome Binding Site
LoxP	Locus of crossover in P1. Recombination site
LTR	Long terminal repeat
MCS	Multicloning Site
ORF	Open reading frame
ori	Origin of Replication
OriP	Plasmid origin of viral replication that contains EBNA binding sites.
pA	Polyadenylation
S/MAR	Scaffold Matrix Attachment Region
UCOE	Ubiquitous Chromatin Opening Element
UTR	Untranslated region

Viral Vectors

AAV	Adeno-associated virus
AdV	Adenovirus
CMV	Cytomegalovirus
EBV	Eppstein-Bar Virus

Genes/Proteins

c-Myc	V-Myc Avian Myelocytomatosis Viral Oncogene Homolog
coGFP	*Cavernularia obesa* variant of the reporter GFP (Ex=488nm, Em=508nm)
dTom	Red reporter gene (Ex=554nm, Em=581nm)
eGFP	An enhanced variant of the reporter GFP (Ex=488nm, Em=508nm)
FancA/C	Fanconi Anaemia complementation group A/C
fLuc	Luciferase variant from *Photinus pyralis*
GAPDH	Glyceraldehyde 3-Phosphate Dehydrogenase

Kat	Katushka turbo Far Red reporter gene (Ex= 588nm, Em= 635nm) from sea anemone *Entacmaea quadricolor*
Klf4	Krueppel-like factor 4
L-Myc	Viral Oncogene Lung Carcinoma Derived Homolog RNA
Lin28	RNA-binding protein Lin28
Nanog	Gene belonging to the Nanog homeobox Family
Oct4	Octamer-binding transcription factor 4 (also known as POU5F1)
OKSM	Reprogramming factors/Yamanaka Factors (Oct4, Klf4, Sox2, cMyc)
rLuc	Luciferase variant from *Renilla reniformis*
SAF-A	Scaffold attachment factor A
Sox2	SRY (Sex-determining region Y)-box 2
SSEA1	Stage-Specific Embryonic Antigen-1 (also known as CD15 and FUT4)
SV40LT	Simian Vacuolating Virus 40 Large T antigen
tGFP	Turbo variant of the reporter GFP (Ex=488nm, Em=508nm)
vLuc	Luciferase variant from *Cypridina hilgendorfii*

Promoters

CAG	CMV early enhancer and chicken β-actin promoter
CMV	Cytomegalovirus promoter
EF1α	Elongation factor 1α promoter
MCMV	Murine Cytomegalovirus promoter
PGK	Phosphoglycerate Kinase promoter
SFFV	Spleen Focus-Forming Virus promoter
UbC	Ubiquitin C promoter

Cell lines

EBs	Embryonic Bodies
HDFs	Human Dermal Fibroblasts
hESCs	Human Embryonic Stem Cells
HSCs	Hematopoyetic Stem Cells
hiPSCs	Human Induced Pluripotent Stem Cells
MEFs	Mouse Embryonic Fibroblasts
mESC	Mouse Embryonic Stem Cells
miPSCs	Mouse induced Pluripotent Stem Cells

NB	Neuroblastoma
NHDF	Non-Human Dermal Fibroblasts
SCs	Stem Cells

Reagents / Media

(D)PBS	Dulbecco Phosphate Buffered Saline
ATRA	All-trans retinoic acid
BSA	Bovine Serum Albumin
dH_2O	Destilled water
DMEM	Dulbecco's modified eagle medium
DMSO	Dimethyl sulphoxide
dNTPs	Deoxynucleotide-5'-triphosphate
FBS/FCS	Fetal bovine/calf serum
LB	Luria-Bertani (broth)
P/S	Penicillin Streptomycin
PEI	Polyethylenimine
Tris	Tris(hydroxymethyl)-amino-methane
Tris-HCl	Tris solution with pH adjusted with HCl acid
VPA	Valproic acid

Molecular terms

°C	Degrees Celsius
Amp^R	Ampicillin resistance
bp	Base pairs
cDNA	Complementary DNA
CpG	5'cytosine-guanine-3'
Da	Daltons (gram per mole)
DNA	Deoxyribonucleic acid
ds	Double-stranded
g	G force or grams
iRNA	Interfering RNA
Kb	Kilo base pairs
kDa	Kilodaltons
Km^R	Kanamycin resistance

M	Molar (moles per litre)
min	Minutes
mM	Millimolar
MBB	Minimal Bacterial Backbone
mRNA	Messenger RNA
ng	nanograms
nm	nanometre
nmol	nanomoles
pDNA	Plasmid DNA
RNA	Ribonucleic acid
rpm	Revolutions per minute
sec	seconds
ss	Single-stranded
UV	Ultraviolet
V	Volts
v/v	Volume per volume
w/v	Weight per volume
µM	Micromolar

Techniques

FACS	Fluorescent activated cell sorting
PCR	Polymerase chain reaction
q(RT)-PCR	Quantitative Reverse-Transcription Polymerase chain reaction
qPCR	Quantitative Polymerase Chain Reaction

Others

ICM	Inner Cell Mass
SCID	Severe combined immunodeficiency
TF	Trophectoderm
ZP	Zona Pellucida

LIST OF FIGURES

LIST OF TABLES

1. INTRODUCTION

1.1 Overview

One of the ultimate goals of gene and cell therapy is the development of safer, more efficient and easily and economically prepared persistently expressing genetic vectors. An ideal vector for the genetic modification of cells should deliver sustainable therapeutic levels of gene expression without compromising the viability of the host cell in any way. Integrating vectors have been the most effective and most commonly used gene delivery tool in the past years. They reached a success peak in the field of gene therapy in 2017, with 671 clinical trials [1] and the approval of gene therapy drugs for the correction of degenerative eye diseases (*Luxturna*) or cancer (*Kymriah and Yescarta*) [2, 3], among others. However, their efficacy is tempered by their potential integration [4] and genotoxicity [5 - 6]. Alternatively, non-viral DNA vectors are attractive alternatives to viral gene delivery systems because of their lower toxicity, relatively easy production and great versatility [7 - 8]. However, their deficient delivery and merely transient gene expression of plasmid DNA *in vivo* [9] hampers their suitability for gene therapy applications.

In this thesis, we aimed to develop an episomal DNA vector system based on entirely human elements which can persistently genetically modify differentiating cells without toxicity and without altering the cells' capabilities. This vector system can be used as a safer alternative to viral vectors for the genetic modification of stem cells or derivation of patients' somatic cells for therapeutic purposes.

1.2 Genetic modification of cells

The **genetic modification** of cells consists of introducing, replacing or removing genetic information to improve or to generate new characteristics. For example, DNA is delivered into cells with the purpose of creating a specific feature **(transgenesis)** or using a gene as a drug to treat a disease **(gene therapy).** Genetic modification has several applications in the fields of medicine, industry, agriculture and research, and can be applied in a wide range of plants, animals and microorganisms. This is done by isolating and cloning a particular gene of interest (GOI) that produces a particular characteristic into a **vector**, which shuttles the genetic information into the cell. Although there are several mechanisms by which cells can be modified, there is no vector able to modify cells, yielding high levels of persistent expression and that has minimal (if not any) effect on the modified cells.

1.2.1 Applications of genetic modification

There are several fields in which genetic modification has important applications. Particularly pertinent for this study are applications in medicine and research, especially for the generation of **transgenic models** and for **gene therapy** purposes. Some relevant applications are summarised in **Table 1.**

Table 1: Applications of genetic modification

Field	Application	Examples	References
Medicine	Drug manufacturing	Recombinant insulin, hormones	Goeddel *et al.,* 1979. [10]
	Hybridomas	Monoclonal antibodies	Roque *et al.,* 2004 [11]
	Chimeric antigen receptors	CAR-T cells	Han and Kwon, 2018 [12]
	Pre-clinical models: genetically modified mice	Study of obesity, cancer, diabetes, ageing, degenerative diseases	Palmiter, 1982 [13]
	Gene Therapy	X-linked SCID Leukaemia, Parkinson, degenerative diseases	Fischer *et al,*. 2010 [14] Black et al,.2014 [15] Cereso et al., 2014 [16] Geiselhart et al., 2012 [17] Milsom and Williams, 2010 [18]
Research	Transgenic models: Genetically modified mice	Gain of function (knock-in)	Alberts, 2008 [19] https://www.genome.gov/12514551/
		Loss of function (knock-out)	
		Tracking/labelling	
Industry	Protein production	Insulin, hormones	Rosano and Ceccarelli, 2014 [20]
	Vaccine production		Ihssen et al., 2010 [21]
	Biofuels		Koppolu and Vasigala, 2016 [22]
Agriculture	Genetically modified crops	Herbicide resistance, stress resistance	Punja, 2001 [23]
	Genetically modified livestock	Drug and protein production	Krimpenfort 1991 [24] Lai,2006 [25]

1.2.2 Types of genetic modification

There are three possible ways in which genetic information can be modified: either by introducing additional information (gene supplementation, gene editing), by deleting or changing endogenous DNA sequences (gene editing); or by interfering with gene expression, either at the transcriptional or post-transcriptional level (gene interference). Regardless of the modification performed, the genetic material, which can be integrated or be kept extrachromosomal, needs to be delivered into the cell via gene delivery mechanisms, which are discussed below.

1.2.2.1 Gene supplementation

Gene supplementation consists of introducing additional genetic information into a cell. Many clinical trials have been focused on treating monogenic diseases by delivering the correct version of a mutated gene. For example, correct expression of FancA was sufficient to restore a functional DNA repair pathway in Fanconi Anaemia patients [26] and the introduction of FancC rescued the diseased phenotype in mice [27]. Likewise, the addition of Rep1 was sufficient to rescue the progressive eye degeneration caused in Choroideremia patients, suffering from an X-linked monogenic disorder affecting this gene [28 - 29]. However, the gene supplementation strategy does not influence correcting diseases caused by dominant negative mutations. In these particular cases, gene editing technologies allow more flexibility and offer a greater versatility to target altered genes resulting in deficient products.

1.2.2.2 Gene editing

Gene editing technology offers the possibility to add, delete and correct pre-existing genetic information. Editing techniques rely on nuclease-induced double-strand breaks (DSBs) that can be repaired via two mechanisms: 1) **Nonhomologous end-joining (NHEJ) repair** results in insertions or deletions *(indels)* of variable length at the site of DSB, which translates into disruption of pre-existing open reading frames (ORFs) and the ablation of the original genetic message. Alternatively, 2) **homology-directed repair (HRD)** can be used to introduce specific sequence alterations in the presence of a homologous DNA template, that following homologous recombination, will insert a particular genetic change or a whole sequence in a specific genomic location.

Genome editing mechanisms, first described in 2005, were based on engineered **Zinc Finger Nucleases (ZFN),** which could induce DSBs to a specific DNA target [30 - 31]. However, their complex engineering and customisation was the primary limiting factor for broader use. Later on, in 2009, the DNA binding domains of bacterial **transcription activator-like effector nucleases (TALENs)** also proved to modify a

desired DNA sequence efficiently [32]. However, a complex design of a specific pair of nucleases per DNA target was required. The latest improvement on gene editing technologies came in 2012 from the hand of Jinek *et al.*, who showed that the bacterial 'immune system' based on clustered regularly interspaced palindromic repeats (CRISPR) and the endonuclease Cas9, could specifically cut DNA sequences when a specific complementary short guide RNA (gRNA) was present [33].

Genome editing techniques overcome some of the leading safety issues, such as the genotoxicity due to random or semi-random integration, the dysregulation of nearby proto-oncogenes, the knock-out of tumour suppressor genes or aberrant splicing.

1.2.2.3 Gene interference

Genetic interference is based on the activity of interfering RNA (iRNA), initially described in *Caenorhabditis elegans* [34]. These effector molecules can interfere with gene expression at a transcriptional or post-transcriptional level. For example, siRNA are dsRNA molecules acting as transcriptional repressors via chromatin modifications [35 - 36]. Also, siRNA can be processed via the endoribonuclease DICER complex, resulting in single-stranded siRNA, which complementary binds to mRNA, recruits the RNA-induced silencing complex (RISC) and results in mRNA cleavage [37]. Another type of iRNA is microRNA, which is involved in regulating gene expression during development and differentiation [38]. Endogenous microRNAs are only partially complementary to mRNA and bind to their 3' untranslated region (3'UTR), resulting in mRNA degradation and translational repression [39].

Because of the ability of iRNA to specifically and potentially knock down specific expression of a gene with a known sequence, they found their place in therapeutical applications. For example, iRNA have been used to treat viral infections [40], neurodegenerative disorders [41], cancer [42] and age-related macular degeneration [43]. Depending on the duration and desired intensity of knock-down activity, iRNA can be delivered using Lentivirus [44], Adenovirus or AAV [45]. Alternatively, iRNA can be delivered using non-viral methods, via complexation with chemical carriers to avoid RNAse degradation [46].

1.2.3 Gene delivery systems

Gene delivery is the process by which foreign genetic material is introduced into cells. The process can be **physical** (*e.g.* microinjection, hydrodynamic injection, electroporation or biolistics), **chemical** (*e.g.* lipid or polymer carriers) or **biological** (*e.g.* viral or non-viral vectors). Although they differ in mechanism, the common underlying principle of all gene delivery systems is that the genetic material needs to overcome a range of extra- and intra-cellular barriers. Examples of extracellular barriers are the cornified cells of the skin [47] and skin nucleases [48], nuclease degradation of the transgene when delivered intravenously [49] or intramuscularly [50], interaction with serum proteins [51] and extravasation from blood circulation [52]. Once the genetic material faces the target cell, it must cross a semipermeable cell membrane [53 - 54], survive endosomal degradation and cytoplasm environment [55] and ultimately reach the nucleus. Entry into the nucleus can be attained either by fusion to nuclear membrane, addition of nuclear localisation signals (NLS) or taking advantage of the temporary disruption of nuclear membrane during cell division [53, 56, 57]. Some of these barriers can be overcome by complexing DNA with polymers or lipids; or directly delivering it into the cytoplasm or nucleus (electroporation, nucleofection) [52].

Because of their efficiency in delivering genetic material and their variety of tropism, viruses are the most commonly used vector for cellular and gene therapy, constituting 67,4% of the current clinical trials [1] (**Figure 1).** However, they have limitations in the size of transgene they can accommodate, which in turn limits the control of genetic material they deliver. More importantly, their use leads to some safety concerns related to insertional mutagenesis [58], associated immune reactions [59, 61], and cytotoxicity [4]. Over the past years, considerable effort has been made in improving chemical and physical delivery methods by making gene delivery more efficient and safer [62]. A summary of different gene delivery systems is shown in **Table 2.**

Table 2: Gene delivery systems

	Mechanism	Advantages	Disadvantages	Examples
Biological	Transfer of genetic material through a natural viral infection pathway using replication-deficient viruses.	• High transduction efficiency • Persistent transgene expression	• Induction of immune responses • Insertional mutagenesis (oncogenesis) • High production costs • Size limitations	• Integrative viruses • Lenti [63], [64] /retrovirus [63] • Non-integrative viruses (AAV, AdV…) [65]
Chemical	Transfer of genetic material complexed with lipid or polymer-based complexes through endocytosis.	• Safer and cheaper to produce • Chemical modification of complexes allows targeted delivery • Effective in *in vitro* experiments	• Short transgene expression • Low transfection efficiency • Low efficiency in non-dividing cells	• Lipofection [66] • PEI [67] • Poly-L-Lysine [68]
Physical	Physical transfer of genetic material through transient pores in the plasma membrane, created by mechanical and physical forces.	• Effective *in vitro* and *in vivo* • Effective in dividing and non-dividing cells • Specific tissue transfection	• Local tissue damage after application of physical/mechanical force • Requires specialised instruments • Optimised protocol and parameters are cell and tissue-specific.	• Needle microinjection [69] • Biolistics/gene gun [70] • Electroporation [71] • Nucleofection [72] • Sonoporation [73] • Magnetofection [74] • Cell squeezeing [75]

Adapted from [62]

1.2.4 Delivery vectors

Gene delivery is the process by which foreign genetic material is transferred into host cells with the help of a shuttle molecule called **vector.** Broadly, delivery vectors are categorised as **viral** or **non-viral** [76]. Viral vectors are extensively used due to their natural ability to infect cells and to deliver genetic material. However, their potential to integrate as well as the associated immune reactions raise safety concerns that limit their use [77]. Non-viral vectors are safer alternatives, although they face a delivery efficiency challenge due to their inherent hydrophilic nature, which results in complexation with delivery vehicles (cationic polymers or lipids) or forced entry via physical methods (electroporation, magnetofection) [78].

1.2.4.1 Viral vectors

Viruses are very efficient vectors because their survival depends on their innate ability to infect cells, which they use as machinery for their expression. Because of their high transduction efficiencies and natural ability to infect and deliver genetic material into a cell nucleus, surpassing all cellular barriers, viral vectors became the most successful and extensively used delivery mechanism – representing 67,4% of the current clinical trials [1], as shown in **Figure 1.**

However, their immunogenicity and cytotoxicity limit their clinical application. For instance, integration close to proto-oncogenes or tumour suppressor genes leads to malignant transformation [58] and even patient death in clinical trials [6].

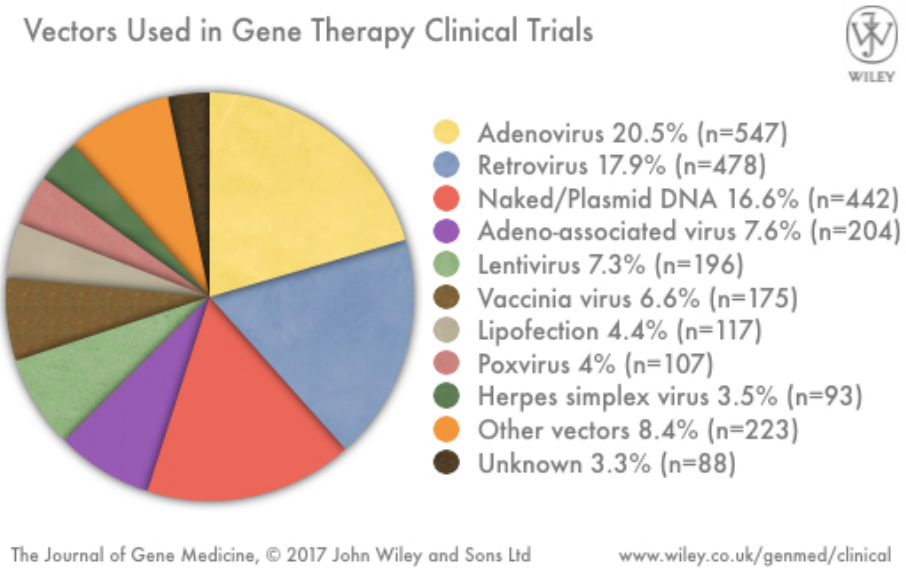


Figure 1: Current vectors used in gene therapy clinical trials

Although viral genomes are modified to accommodate foreign genetic material by deleting regions responsible for viral replication and toxicity, strong systemic inflammatory responses following viral delivery have been observed [59]. These are mostly due to cytotoxic immune responses elicited by viral gene products or the transgene itself [61]; as well as humoral responses due to viral capsid proteins [60].

Currently, there are eight types of viruses that can be used as delivery vectors and are classified into two groups, depending on whether their genomes integrate into the host cell chromosomes; (γ-Retrovirus, Lentivirus) or persist as extrachromosomal entities (AdVs, AAVs, Sendai Virus, Herpes Virus, EBV, and Poxvirus).

1.2.4.1.1 Viral integrative vectors

The genetic material is encoded in a virus, which is delivered and integrated into the host cell genome. Although typically very efficient at transducing cells integrative vectors are potentially dangerous and their use raises safety concerns. They can cause cellular damage due to insertional mutagenesis [4] and its potential for malignant transformation, possibly via activation of neighbouring proto-oncogenes or inactivation of tumour-suppressor genes [58], which in some cases have led to patient deaths [5]. Additionally, they can induce silencing by *de novo* methylation of the inserted gene or neighbouring DNA [79, 80].

- **γ-Retrovirus**

γ-Retrovirus, such as Murine Leukemia Virus (MLV), are enveloped, single-stranded RNA (ssRNA) viruses that belong to the *Retroviridae* family. The genetic information is contained and encapsulated into a modified retroviral ssRNA genome, which cannot replicate [63, 81], and must necessarily be retrotranscribed into cDNA before integration. Retroviruses transduce dividing cells with high efficiency. The main limitations of retroviruses are their small cloning capacity (8kb) and their inability to infect post-mitotic (non-dividing) cells, therefore limiting the spectrum of cells to target (*e.g.*: neurons, myocytes…). Also, they tend to integrate into regulatory regions (5'UTR) of the gene [82, 83], they can integrate in tandem, generating differential expression of the transgene. High titres are difficult to obtain [84].

- **Lentivirus**

Lentiviruses are a subclass of retrovirus, and they belong to the *Retroviridae* family. They are also enveloped ssRNA viruses but their incubation time is slower (*Lente-*, Latin for slow) and they can infect both dividing and non-dividing cells with high efficiency [85] but not quiescent cells [86]. They are less immunogenic than retrovirus but still stimulate the immune system via activation of TLR3 and TLR7 [87]. Their maximum loading capacity is 9kb, and they integrate randomly in coding regions of the genome [88]. However, directed mutations in the integrase gene resulted in impaired integration and existence of episomal forms [89].

1.2.4.1.2 Viral non-integrative vectors

Other viral vectors can infect cells and remain as episomal DNA forms or as RNA molecules in the cytoplasm.

- **Adenovirus (AdV)**

Adenoviruses belong to the *Adenoviridae* family and are non-enveloped viruses with an icosahedral capsid and linear dsDNA genome. They can infect dividing and non-dividing cells with low host specificity, which makes them efficient at transducing almost all tissues. Their cloning capacity is similar to retroviruses (around 8kb), and they can be produced in high titres. AdV enter the cells by recognising widespread Coxsackie and Adenovirus receptors (CAR) and α_v integrins [90], making it difficult to generate cell-specific adenoviral vectors. AdV rarely integrate into the genome but rather stay in their episomal form [91] and yield transient transgene expression. Their capsid mediates a high inflammatory

response, which has lead to severe side effects and patient deaths in clinical trials [59, 92, 93]. Also, there is a pre-existing immunity against AdV5 since commonly infects humans [94].

- **Adeno-associated virus (AAV)**

AAVs are small non-enveloped and non-pathogenic ssDNA or dsDNA (self-complementary) viruses belonging to the *Parvoviridae* family. Their productive infection takes place only in the presence of a helper virus, either Adenovirus (hence their name –*associated*) or Herpes virus [95]. In the absence of helper virus, AAV2 can integrate at specific parts of the genome and establish a latent infection [65]. Although AAV virus can specifically integrate to establish a latent infection, AAV-based vectors are designed to remain episomal, avoiding the risks associated with insertional mutagenesis. They can also infect dividing and non-dividing cells and persist episomally in the nucleus yielding strong transgene expression. AAV can infect a broad range of cells, and each serotype is cell-type specific, making them good candidates for gene therapy [96]. AAV infect cells by interacting with proteoglycans [97] and the internalisation is facilitated by interactions with $\alpha_v ß_5$ and $\alpha_v ß_1$ integrins [98, 99], FGFR1 [100], and laminin receptor [101]. Although rare, integration events are possible (*e.g.*: AAV2). Their cloning capacity is slightly smaller (4 - 5kb), and there is a pre-existing immunity against a range of serotypes [60, 65].

- **Sendai Virus**

Sendai viruses (SeV) belong to the *Paramyxoviridae* family. They have negative ssRNA genomes around 15Kb, without a DNA phase, and their replication takes place in the cytoplasm. They possess strong immunogenic potential and can persist in infected cells for long periods of time, making the removal of the viral particle rather complicated [102].

- **Herpes Virus**

Herpes virus are enveloped viruses with an icosahedral capsid and a linear dsDNA genome. They belong to the *Herpesviridae* family. Upon infection, the viral genome is delivered to the nucleus where it remains episomal, although in rare events integration might occur. They can accommodate large transgenes (150kb) and are neurotropic, which makes them a suitable tool for gene delivery in neurons and nervous-related structures. They can be produced at high titres. There is a pre-existing immunity since herpes virus are the cause of the widely spread cold-sores [103].

- **Epstein-Barr Virus (EBV)**

Epstein-Barr is also an enveloped 170Kb dsDNA virus that belongs to the *Herpesviridae* family. They have a tropism for B-cells [104]. Upon infection, they persist episomally in the nucleus and deliver long-term expression of the transgene [105]. However, adaptative immunity develops after infection, and they have oncogenic potential [106].

- **Poxvirus (Vacciniavirus)**

Poxviruses are large, enveloped, linear dsDNA viruses that belong to the *Poxviridae* family. They replicate in the cytoplasm by using virally encoded polymerases to replicate and transcribe [107], which yield high levels of cytoplasmic transgene expression. Their cloning capacity is larger than Retrovirus (25kb), which makes them a very attractive tool for gene delivery. Although not naturally integrative, recombinant vaccinia vectors are engineered to insert the transgene into the genome via direct site-specific recombination [108]. However, their biology and structure is complex and therefore, they are not so extensively used as gene delivery tools [109].

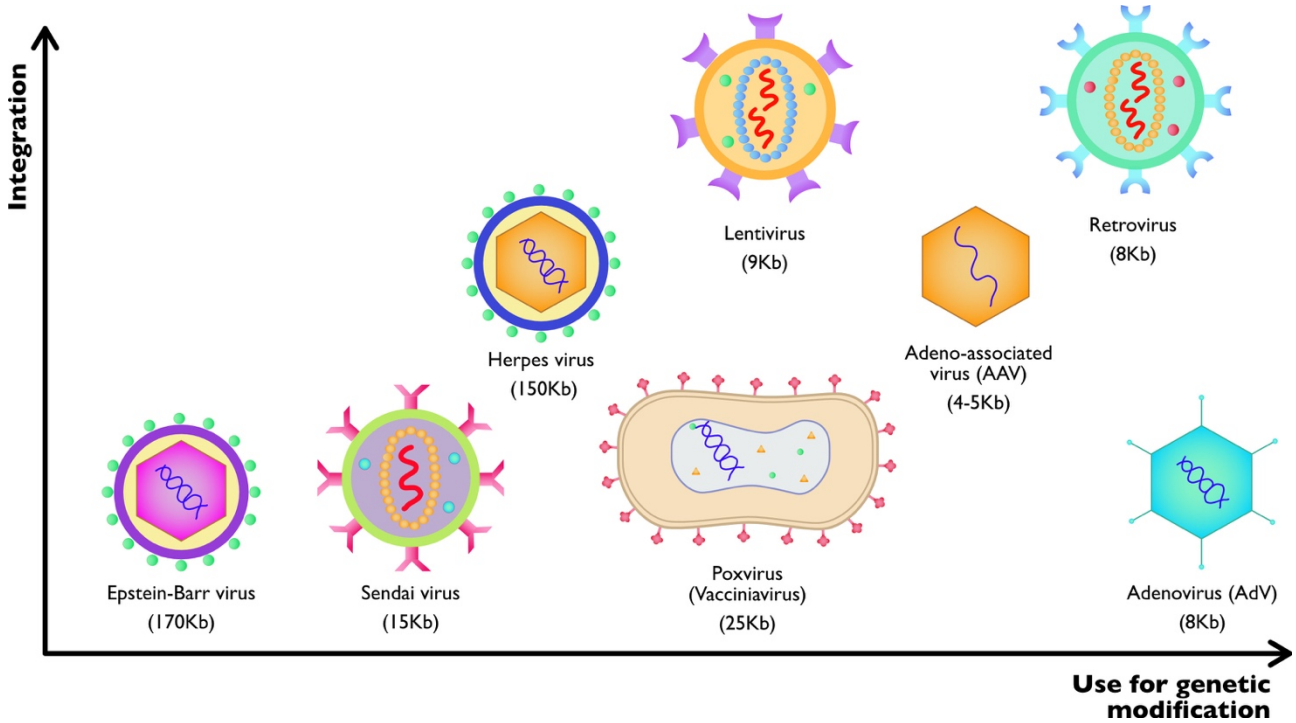


Figure 2: Schematic depiction of most commonly used viral based vectors and their integration potential
Engineered retroviral vectors and adeno-associated viruses (AAV) are the most widely used vectors to provide long-term transgene expression in modified cells. Although efficient, their use is hampered by potential risks such as integration, transgene silencing or immunogenicity. Non-integrative Adenovirus or Sendai Virus are safer, but their transgene expression is transient.

1.2.4.2 Non-Viral vectors

Because of the significant safety concerns associated with viral vectors and the inherent size limitation of accommodating genetic material into a protein capsid, much effort has been invested towards the development of non-viral systems for gene delivery. Non-viral vectors are designed via basic molecular engineering to deliver larger payloads of genetic material, providing flexibility in the size of the transgene, as well as allowing better control of expression due to the incorporation of regulatory sequences. Additional to their decreased pathogenicity and reduced immunotoxicity as compared to viral vectors, non-viral systems are easier and cheaper to produce on a large scale.

Although plasmids can deliver large amounts of genetic content, and they are much less immunogenic, their delivery efficiency is lower than that of viral vectors because of their difficulty to surpass multiple cell barriers [52]. While viruses have evolved to deliver genetic material into host cells proficiently, non-viral vectors have to overcome each physical and physiological obstacle in a cell. When delivered *in vivo*, they have to survive a hostile plasma environment, including shear forces, degrading enzymes and interaction with plasma proteins and antibodies. Then, non-viral vectors still have to retain their ability to cross the cell membrane, resist cytoplasmic degradation and pass through the double nuclear membrane to reach the nucleus (reviewed in [52]). Also, because of their large size and hydrophilic nature, the cells prevent DNA entry through negatively charged membranes. Therefore, much effort is invested in developing techniques to increase the vectors' delivery efficiency via physical delivery or complexation with chemical carriers **(Section 1.2.2).**

1.2.4.2.1 Non-viral integrative vectors

- **Excisable systems**

Excisable systems introduce, invert or release a specific sequence or gene of interest (GOI) into/from a certain genomic locus. The excision takes place at homologous regions that flank the GOI. There are currently four systems described: 1) The λ integrase [110], 2) Cre/LoxP system [111], 3) the Flp/FRT system [112] and 4) the ϕC31 recombinase [113]. Some disadvantages of excisable systems are the DNA foot-print or genomic 'scar' left on the genome upon excision and the need of administrating the recombinase enzyme separately to excise the desired genes at specific time points.

- **Transposons**

Transposable elements or transposons are DNA sequences that can change their position within the genome (Nobel prize Barbara Mcclintock 1983, [114]). This jumping elements often result in duplication of the same genetic material. A commonly used transposon system is the Sleeping Beauty transposon, which is a synthetic transposase reconstructed from extinct transposase sequences from salmon [115]. Although transposons are efficient and deliver long-term expression of the transgene, they can be considered mutagenic as they can be inserted into functional genes or into regulatory regions, disrupting normal gene expression.

1.2.4.2.2 Non-integrative

- **Plasmid DNA**

The simplest form of a non-integrative and non-viral vector is a naked plasmid DNA, which consists of a circular, dsDNA molecule different from the chromosomal DNA that usually remains in an extrachromosomal state. A conventional plasmid is composed of two modules: the expression cassette and the bacterial backbone. The mammalian expression cassette usually comprises an enhancer/promoter region, a 5' untranslated region (5'UTR) including introns, and a transgene, which can be a reporter gene, therapeutic gene or any gene of interest followed by a polyadenylation tail. The bacterial backbone is composed of a bacterial origin of replication (*ori*) and an antibiotic resistance or a selectable marker for plasmid amplification in bacterial strains. However, plasmids have some inherent limitations. For example, the lack of mammalian replicative sequences results in transient expression of the transgene due to vector dilution [116]; and the recognition of unmethylated CpG of bacterial origin are hotspots for immune recognition [117]. Also, plasmids can break upon delivery, which could lead to rearrangements and integrations, although at a rate less than 10^{-5} stable integrants/transfected cell [118]. These limitations have led to modifications to satisfy the requirements for their clinical use [7], which resulted in increased use of DNA vectors in clinical trials **(Figure 1).**

- **Minicircles and miniplasmids**

Several studies suggest that decreasing the plasmid size improves transfection efficiency [119]. In 1997, Daraquet *et al.* used site-specific recombination to eliminate bacterial sequences from plasmids to turn them into minicircles [120]. Minicircles are small circular DNA molecules that are excised from a parental plasmid upon recombination, using the excision mechanisms mentioned above. They are devoid of bacterial sequences (*ori* and bacterial selection), which are left in the residual miniplasmid. Also, the reduction of bacterial unmethylated CpG content reduces the immunogenicity of the

vector [117]. However, the absence of bacterial backbone complicates the vector purification after recombination. Minicircles are not self-replicative, and they get lost in the host cell after few rounds of replication [7]. Overall, vectors devoid of bacterial sequences have better transfection efficiencies *in vitro*, *ex vivo* and *in vivo* [119], yield better levels of transgene expression [120] and are safer for clinical use.

- **Self-replicative episomal vectors**

Conventional non-viral episomal vectors cannot replicate their genomes, leading to dilution of the vector during cell division and transient expression of the transgene. To tackle this issue, replication-driving sequences such as Epstein-Barr Nuclear Antigen 1 (EBNA-1) [105] or the SV40 large T antigen [121, 122], were added to plasmid DNA. These viral element-based vectors proved to be self-replicative and provided long-lasting transgene expression. Although not a virus *per se*, the EBNA-1 viral components have been associated with cellular transformation [123] in lymphocytes.

Table 3: Summary of viral and non-viral gene delivery

Vector type		Example	Advantages	Disadvantages
Viral	**Integrative**	Retrovirus/ Lentivirus	• High transduction efficiency • Long-term expression • Infect non-dividing cells (Lentivirus)	• Small cloning capacity (8kb) • Infect only dividing cells (Retrovirus) • Insertion in tandems • Insertional mutagenesis • Oncogenic potential
	Non-integrative	Adenovirus (AdV)	• Infects dividing and non-dividing cells • Low host specificity (efficient transduction of most tissues) • High titres	• Small cloning capacity (8kb) • Transient expression • Capsid mediates inflammatory response • Pre-existing immunity towards AdV5
		Adeno associated virus (AAV)	• Infects dividing and non-dividing cells • Less inflammatory • Strong transgene expression	• Rare but possible integration events • Limited cloning capacity (4-5kb) • Complicated vector production • Pre-existing immunity against AAV2
		Sendai virus	• RNA genome kept in the cytoplasm	• Strong immunogenic potential • Long-term persistence after infection (labour-intensive particle removal)
		Herpes Virus	• Large transgene capacity (150kb) • Neurotropism • Episomally maintained • High titres	• Pre-existing immunity • Potential to integrate
		Epstein-Barr virus	• Long-term expression • B-cell tropism • Episomally maintained	• Adaptative immunity after infection • Oncogenic potential
		Poxvirus	• High levels of transgene expression in the cytoplasm • Large cloning capacity (25kb)	• Complex structure and biology • Risk of cytopathic effects
Non-viral	**Integrative**	Excisable systems (Cre/LoxP or Flp/FRT)	• Insertion/removal of the flanked region at specific time points	• DNA footprint or genomic scar upon excision • Separate administration of recombinase
		Transposons	• High efficiency • Long-term expression	• Random integration (mutagenic) • Genomic scar
	Non-integrative	Self-replicative episomal vectors	• Self-replicative • Easy to manufacture	• Longer expression might require repeated transfections
		Minicircles	• Reduction in size • Elimination of bacterial sequences	• Purification might be difficult • Short-term expression

1.3 Episomal vectors

Lots of effort has been invested in developing more effective and safer vectors for gene delivery. While **integrative vectors** are widely used and highly efficient, their use is hampered by the risk of insertional mutagenesis [4] and genotoxicity [5, 6]. The use of **non-viral vectors** is safer and allows better flexibility on the size and design of the vector [124], although it compromises the efficiency of gene delivery [9, 52]. In general, **non-viral and non-integrative (episomal) vectors** are preferable, in particular for gene therapy, where the aim is to correct and treat patients whilst avoiding collateral damage caused by the therapeutic vector. However, their use for gene delivery is still regarded to be below requirements, mainly due to a lack of DNA development. While all the attention has been focused on improving both transfection efficiency and the amount of DNA that reaches the nucleus, less work has been done in developing the DNA sequence, configuration and elements that can improve the vector performance and unleash its potential.

This chapter introduces and describes the use of **non-integrative episomal vectors for gene delivery** and highlights their advantages and disadvantages. It introduces the use of chromosomal elements, such as Scaffold Matrix Attachment Regions (S/MARs), to drive episomal maintenance and replication of DNA vectors while preventing epigenetic silencing and enhancing transgene expression.

1.3.1 Advantages of episomal vectors

There are several advantages inherent to non-integrative episomal vectors. Most importantly, they are potentially **less toxic** as they remain extrachromosomal thus **eliminating** the risk of **insertional mutagenesis** [125]. Their **production is relatively easy,** as they can be produced in bacterial strains such as *E.coli* [120, 126]. They are **not subject to size constraints,** allowing larger DNA payloads that would enable controlled modulation of gene expression. For instance, Lufino *et al.* delivered an iBAC vector based on an S/MAR motif coding for the 135kb locus of low-density lipoprotein receptor (LDL-R) using a high-insert capacity Herpes Simplex virus [124]. Additionally, episomal vectors usually persist in multiple copies per cell, as compared to integrative vectors, which integrate at one site/virus [127]. However, episomal vectors also have some significant limitations.

1.3.2 Limitations of episomal vectors

Their efficiency is still regarded to be below the requirements for realistic *in vivo* applications due to inefficient delivery [52, 119]; exacerbated by the typically transient gene expression of plasmid DNA

in vivo (reviewed in [128]). Once non-viral vectors are efficiently delivered into cells, one of the major obstacles is to maintain stable and persistent levels of transgene expression. Upon DNA delivery, transgene expression reaches a maximum peak after 12-48h and then it gradually decreases until its expression is barely detectable after a week, both *in vitro* and *in vivo* [116]. To improve and generate better DNA vectors, it is essential to understand the mechanisms by which vectors can lose expression of their transgenes. The reasons explaining the drop in transgene expression might be diverse.

Vector damage following delivery can lead to DNA degradation, rearrangement or integration into the genome [129]. The DNA can also be lost due to **dilution of vector** copy numbers if no self-replicative mechanisms are incorporated into the plasmid (reviewed in [130]), **promoter CpG methylation** [131, 132], **DNA heterochromatinisation** [133] or **inflammatory responses** to elements of the vector via activation of TLR9 [134, 135], which result in downregulation of expression at a post-transcriptional level [136, 137]. The expression of foreign proteins in circulation can **induce an inflamed state,** resulting in silencing of protein expression and **removal of expressing cells** [135, 138] to prevent the spread of 'infection'. Also, modifying vectors derived from viruses or bacteria can be identified by both the adaptive and the innate immune systems of the target host [117], and these reactions can have a severe consequence on the efficacy of delivery and the host itself [135].

1.3.2.1 Immune reactions towards the vector

Both the delivery system as well as the DNA molecule can be immunogenic. For example, cationic polymers used to complex DNA molecules can elicit immune reactions when delivered *in vivo* [139], and DNA vectors can trigger immune responses upon recognition of high-density unmethylated CpG islands from bacterial origin [117]. Mammalian DNA contains 25 to 30% fewer CpG motifs than bacterial DNA and 80% of this mammalian sequences are found methylated. The presence of highly unmethylated bacterial CpG motifs is recognised by the Toll-like Receptor 9 (TLR-9) [134, 135], which in turn activates a cascade of proinflammatory cytokines and chemokines such as AP-1, TNFα, IFNγ or IL-12; that can downregulate expression post-transcriptionally [136, 137]. The immune response against the vector backbone leads to silencing of the transgene or the elimination of transduced cells [135].

1.3.2.2 Epigenetic events

Epigenetic silencing is a naturally occurring process essential for controlling gene expression. Epigenetic events occur in different contexts, for example: during embryonic development, at specific cellular

states or as a consequence of foreign DNA recognition. During **embryonic development** there are different methylation patterns, beginning with both hypermethylated gametes and embryo, that need to be epigenetically erased after fertilisation due to an extensive genomic demethylation. After resetting the epigenetic landscape like a 'blank canvas', the blastocyst undergoes rapid remethylation after implantation and a gradual remethylation during development to specify different somatic lineages [140]. Remethylation of the genome requires *de novo* methylation at CpG sites due to *cis*-acting methylation sequences. Epigenetic events might also occur at **specific stages**, for example in a differentiated state as opposed to a pluripotent state [141, 143]. Another common epigenetic silencing event occurs following gene delivery into mammalian cells and especially using non-viral vectors. **Transgene silencing** is the consequence of evolutionary mechanisms developed by Eukaryotic cells to detect foreign DNA and abnormal proteins. Silencing takes place due to *de novo* methylation by DNA methyltransferases together with histone modifications (H3K9). These create methylated centres that trigger chromatin condensation, which spreads towards neighbouring chromatin unless blocked by an insulator.

DNA methylation, and therefore silencing of gene expression, is a reversible process. The regulation of gene expression is cell-, time- and context-specific and in turn, determines the properties of the cell [144]. For example, stem cells and progenitor cells are hypomethylated (low levels of DNA methylation), which means that they are less prone to gene silencing. However, these cells become methylated upon differentiation and may silence endogenous as well as exogenous genes (transgenes) [145]. Therefore, using chromosomal elements that counteract or prevent DNA methylation and silencing can be particularly beneficial for the genetic modification of stem cells.

1.3.2.3 Effects of neighbouring chromatin - Positional effects

In genomic DNA and integrated transgenes, position effect variegation (PEV) is the influence of the chromatin environment on gene expression. For example, high levels of transcription can be achieved at nuclear sites specialised in transcription that are protected and separated from heterochromatic regions, where silenced genes are located. In turn, RNA polymerase complexes and TFs influence the chromatin state by recruiting chromatin remodelling and histone modifying enzymes [146], which ensure that the region to be transcribed is accessible. However, compact chromatin has an adverse effect on gene expression and also has a negative influence on regulatory elements flanking the integration site of a transgene [147]. Therefore, it is desirable to identify genetic elements that shield or protect the

transgene from these effects or that control the chromatin structure to develop non-viral vectors to target active chromatin regions and avoid compartments where gene expression is switched off.

However, the influence of chromatin in episomal vectors is not fully understood. The establishment of episomal vectors is a stochastic and infrequent event that strongly depends on which nuclear compartment the vector settles after transfection [148]. Previous studies have shown colocalisation of S/MAR episomal replicons with early replicating foci and regions enriched with active histone marks [149], although no specific chromosome or chromosomal localisation preference could be observed. Subsequently, Hadegorn *et al.* mapped the genomic contact sites of S/MAR-based episomes and showed a preferential localisation in actively transcribed regions, preferentially within promoter sequences, and transcription start sites, which were enriched for open chromatin markers and localised in close proximity to origins of replication [150]. In agreement with the transcription factory model proposed by Cook in 1999, active genes are found co-localized with polymerases and transcription factors in clusters (factories) in which transcription, replication and DNA repair occurred [151]. Hagedorn *et al.* proposed that episomal replicons were established and transcribed in a specific factory and that once the episome found their niche in that factory, their presence would be maintained over mitotic divisions [150].

1.3.3 Evolution of episomal DNA vectors

To overcome some of the forementioned limitations, such as decreasing the immunogenicity of the vector as well as preventing epigenetic silencing or positional effects; several improvements have been made in the vector design. These improvements go in line with abandoning the use of virally-derived vectors and moving towards replicating non-viral DNA molecules.

1.3.3.1 Replication-deficient viruses

The ancestors of episomal vectors are replication-deficient viruses based on AdV or AAV, which are naturally able to be episomally retained by using viral proteins. The replication-deficient versions of these viruses lack the ability of the respective wildtype virus to replicate and generate viral progeny. They are delivered as naked DNA into the cells without encapsidation into viral proteins, and once the viral-replicon reaches the nucleus, it establishes and replicates [152]. However, viral replicons are expensive to produce, and replication-deficient viruses still have safety issues such as leaky immunogenic viral products, which have been reported. Additionally, a low frequency of integration occurs in AAV2 [153].

1.3.3.2 Substitution of viral replicons by viral proteins

The next step in the development of episomal vectors was the substitution of viral replicons by viral proteins responsible for viral replication. For example, the SV40 virus requires the large T antigen (Tag) [154], and the EBV needs the Epstein-Barr Nuclear Antigen 1 (EBNA-1) protein [155] to initiate replication at the viral *ori*. These viral proteins are also needed for the segregation of the viral episomes during cell division **(Figure 3).**

A significant limitation of these vectors is the requirement of a *trans*-acting viral protein EBNA or SV40 Tag, which have transforming potential, which constrains their use *in vivo.* For instance, Snudden *et al.* showed that EBNA-1 could bind to RNA *in vitro* and can, therefore, play a role in post-transcriptional regulation [156]. Sung and Pagano suggested that OriP resembles the enhancer region close to c-Myc, which raises the possibility of EBNA-1 deregulating the expression of the proto-oncogene [157]. Also, EBNA-based episomes are lost at a 4% rate over time in the absence of selective pressure and continuous selection led to the integration of the episomes. EBNA encoding vectors can lead to cellular transformation in lymphocytes [106]. SV40 Tag leads to transformation by interfering with the retinoblastoma and p53 tumour suppressor pathways [158].

1.3.3.3 Episomal vectors based on chromosomal elements (S/MAR)

Because of the safety concerns related to viral-replicons and episomes based on viral proteins, efforts have been directed towards the construction of replicating episomal vectors composed of chromosomal elements.

In the 1980s, autonomously replicating sequences, that act as binding sites for ORC [159], were discovered in yeast. These sequences were subcloned into a plasmid [160] to investigate whether the vector could replicate autonomously; they were either lost or integrated into the genome. In a similar attempt, putative mammalian origins of replications were inserted into plasmids and resulted in few cases of episomal maintenance [161]. A thorough analysis of these putative sequences revealed that the sequences were AT-rich and contained scaffold/matrix attachment regions (S/MARs).

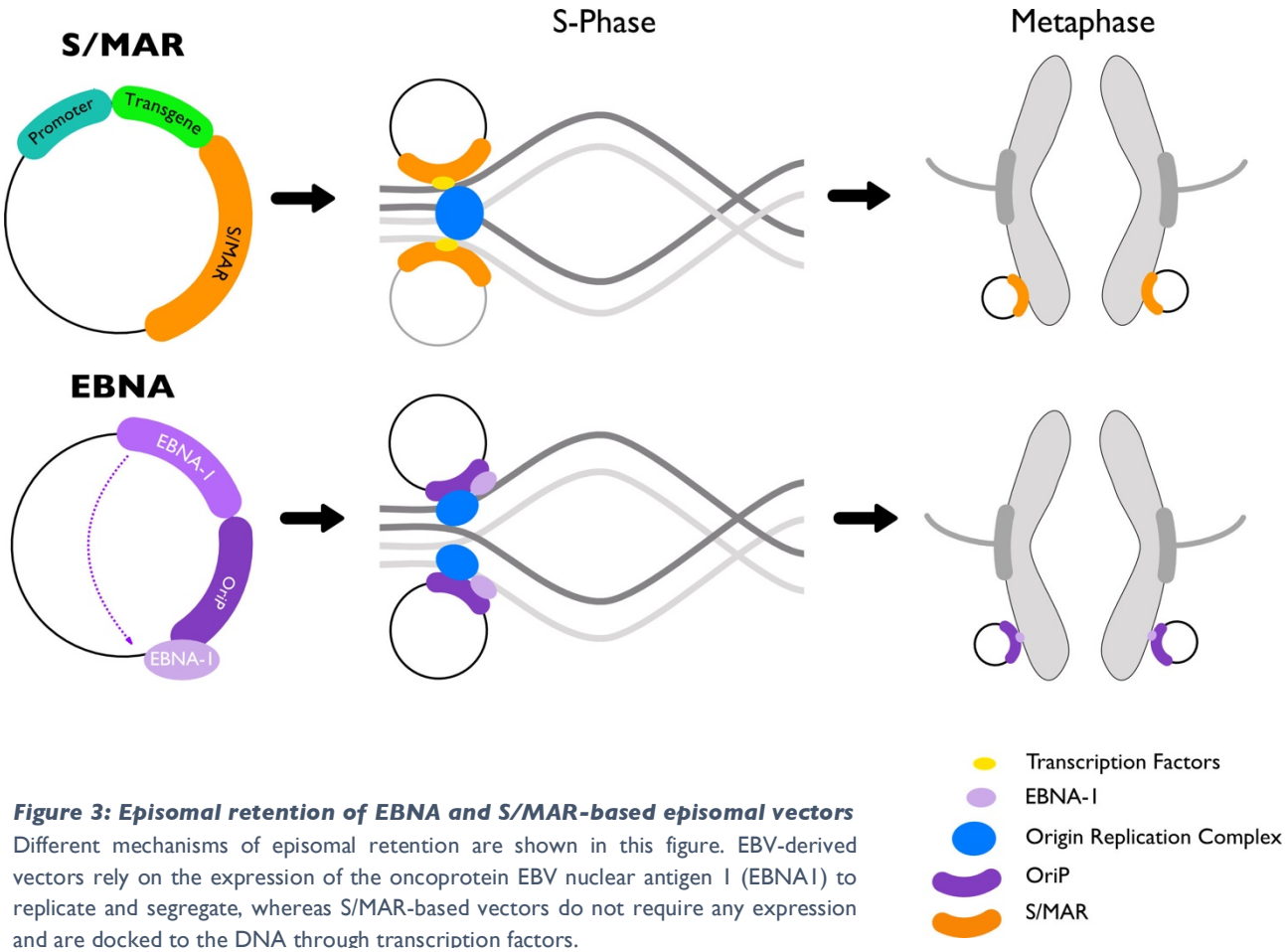


Figure 3: Episomal retention of EBNA and S/MAR-based episomal vectors
Different mechanisms of episomal retention are shown in this figure. EBV-derived vectors rely on the expression of the oncoprotein EBV nuclear antigen 1 (EBNA1) to replicate and segregate, whereas S/MAR-based vectors do not require any expression and are docked to the DNA through transcription factors.

In order to generate an episomal construct in which no viral protein was needed for the vector replication, the group of Hans Lipps replaced the SV40 Tag, required for episomal maintenance, by a S/MAR motif from the 5' region of the human β interferon gene cluster [162], while keeping the SV40 origin of replication. This change gave birth to pEPI **(Figure 4);** the first virus-free episomally maintained DNA vector. This vector provided episomal replication and maintenance [163] in a wide range of cells tested [164], including human hematopoietic stem cells [165], and it was kept at low copy numbers [149]. In addition, it conferred mitotic stability in the absence of selection [162], co-segregated with chromosomes during mitosis **(Figure 3)** and had unlimited cloning capacity [124].

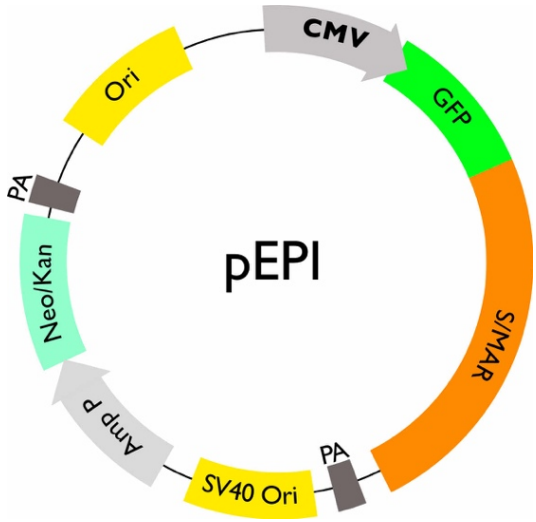


Figure 4: pEPI vector: The first virus-free replicating vector
Cartoon depicting the plasmid pEPI with all necessary components for replication, maintenance and expression. This vector contains the S/MAR region derived from the human interferon β gene, as well as the SV40 origin of replication. The expression of the reporter gene eGFP is driven by the CMV intermediate-early promoter. A large bacterial backbone comprises a double antibiotic selection and a bacterial origin of replication.

Based on the originally described pPEI prototype, several modifications were made to improve its potential applications and are summarised in [166]. For example, the promoter was replaced by a range of more suitable promoters for *in vivo* applications, such as CAG [167], AAT [168], UbC [169]. Additionally, the CpG content from the backbone was depleted [170] and the bacterial elements were removed by generating S/MAR-based minicircles [171].

1.4 Scaffold/Matrix Attachment Regions (S/MARs)

S/MARs are 300-3000bp long evolutionarily conserved AT-rich sequences that play a role in chromosomal and nuclear architecture [172, 173]. These DNA structural motifs bind chromatin to the nuclear matrix, partitioning the genome into independent chromatin loops **(structural function)** [174]**.** These loops are domains involved in several functions, for example, DNA replication, transcription, RNA processing, signalling and transduction **(regulatory functions)** [147, 175]. S/MARs co-localise with transcription factories and they are involved in gene expression by controlling chromatin structure and accessibility as well as recruiting of TFs. Also, due to their AT-rich sequence, they facilitate the unwinding of DNA strands [176] and allow access to DNA replication machinery, favouring DNA replication. S/MARs also function as insulators by protecting a transgene from epigenetic silencing and augmenting transcription [177].

1.4.1 Structural function - S/MAR and loop domain organisation

The nuclear matrix is defined as the insoluble fraction of the nucleus that remains after removing DNA and histones. It's composed of regulatory (hnRNP, TFs) and structural proteins. The nuclear matrix interacts with the chromatin through protein complexes during interphase, and these DNA interaction points are the S/MAR motifs [172], which help to organise the chromatin in independent functional looped domains that contribute to regulating transcription and replication.

1.4.2 Regulatory function - S/MAR mediated transcriptional regulation

Several pieces of evidence indicate that S/MARs play a regulatory role in transcription, rather than just being involved in anchoring chromatin loops to the nuclear matrix. For example, a genomic analysis revealed that the S/MAR density is similar to gene density, suggesting that each gene has its own S/MAR motif **'one gene-one S/MAR hypothesis'** [178]. Also, actively transcribed genes are localised in inaccessible (peripheric) chromatin loops, which provide a permissive and relaxed environment for TFs and transcription machinery to access the DNA **(Figure 5a,b).** On the contrary, inactive genes are topologically compact and inaccessible to TFs **(Figure 5d).** It has been shown that S/MAR motifs adjacent to non-expressed genes are not physically interacting with the nuclear matrix, which suggests that **transcription running through the S/MAR** is a requirement for its activity [179]. It has also been shown that vectors without S/MAR or whose transcription stopped upstream of the S/MAR motif, resulted in integration and subsequent methylation of the promoter, which ultimately resulted in epigenetic silencing and loss of transgene expression [180]. Data from a proof-of-principle experiments

published by Rupprecht *et al.*, in which an inducible TetON promoter was driving the S/MAR transcription unit, showed that when Doxycycline was removed, the vector was lost [181].

Finally, the affinity of S/MARs for the nuclear matrix involves them in other biological processes such as protection from positional effect variegation [177], transcription augmentation [173], long-term maintenance of transgene expression [169, 182], enhancer function and function as origins of replication [183, 184].

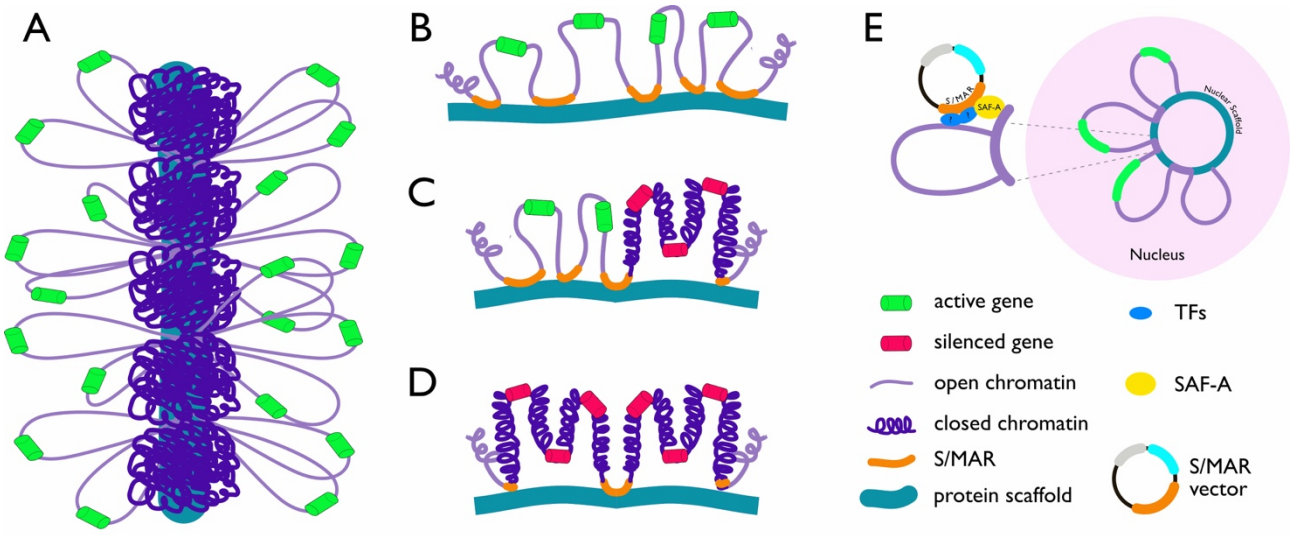


Figure 5: Structural and regulatory functions of S/MAR motifs.
Chromosome model showing inactive (highly compact) and active (looped domains) regions. Active genes (green regions) are localised in the surface of the loop, where RNA is formed within large transcription and splicing complexes **(A).** Spatial patterns of gene expression **(B-D).** The coiled structure indicates closed chromatin and inactive genes are represented as a red cylinder. Open chromatin is depicted as linear structures and active genes as green cylinders. S/MAR elements are represented as orange segments that bind to the matrix that compartmentalise genes by creating transcriptionally active loops. Schematic depiction of episomal replication and maintenance of S/MAR vectors **(E).** The S/MAR motif binds to DNA through scaffold matrix attachment protein A (SAF-A) and other auxiliary replication/transcription proteins.

1.4.2.1 Insulator function

As described above, the interaction of S/MAR vectors with the nuclear matrix is essential for generating independent chromatin domains [183] and insulating those loops from the adverse silencing effects from surrounding chromatin (PEV) [177], as well as to allow local access of TFs to promoters and enhancers. The S/MAR insulating function has been shown on several occasions. For instance, when a reporter gene was integrated into the host cell DNA, the S/MAR could enhance expression of the reporter gene only in stably transfected cells but not in transiently transfected cells [185, 186]. These findings were reinforced with the one gene-one S/MAR hypothesis, that predicts each gene to be regulated by one S/MAR [187]. In other studies with the immunoglobulin κ chain transgene, which

is methylated (inactive) in pre-B cells and hypomethylated (active) in B cells, it was shown that the κ chain transgene was demethylated only when an S/MAR was present after B cells were transfected [188]. Others found that S/MAR could initiate transcription in genes with demethylated CpG [189]. Finally, S/MAR sequences bind to chromatin remodelling proteins such as SAF-A or p300, which influence histone acetylation and nucleosome remodelling [147].

1.4.2.2 Transcription augmentation

Similar to other *cis*-acting transcription regulatory elements, such as promoters and enhancers, S/MARs are usually found in the first intron, and their transcriptional effect depends on the distance from the promoter and direction of transcription [190]. Although it is not clear how exactly S/MARs upregulate gene expression; some suggest that S/MAR can mediate changes in the chromatin by establishing an active locus [191] and by increasing the probability of a permissive chromatin state while minimising the silencing [192]. Others suggest that S/MAR sequences could directly bind to TFs (SAF1A), which interact with the transcription machinery, favouring the accessibility of the plasmid to transcription factories. Finally, fluorescent *in situ* hybridisation (FISH) shows that S/MAR vectors are localised in the periphery of chromosomes, where the chromatin remains easily accessible to transcription complexes [193, 194].

1.4.2.3 S/MAR as origins of replication

The S/MAR sequences are evolutionary conserved, and although they do not have a consensus sequence, they comprise 70% of AT-rich sequences [172, 173], which facilitate dissociation of DNA strands and allow the chromatin to open, unwind and be accessible to the DNA replication machinery. Also, S/MAR motifs contain binding sites (ATTA) for replication initiation proteins. The S/MAR function as an origin of replication allowed the construction of episomally replicating expression vectors for mammalian cells. Its association with the nuclear matrix is essential for the plasmid replication, maintenance and segregation during mitosis.

1.4.2.4 Mitotic stability

The molecular mechanisms by which S/MAR episomes are mitotically stable and equally segregated are not entirely understood. There is evidence suggesting that S/MAR sequences associated with nuclear proteins such as Topoisomerase II, Lamin B1, SATB1 or Histone H1, which would enable the vector co-segregation during mitosis [183]. Like EBNA-1 or SV40 Tag, S/MAR sequences recruit nuclear components that result in destabilisation of the DNA strands and allow the recruitment and

assembly of the replicating machinery (ORC). That means that placing an S/MAR sequence into a DNA vector would allow the episome to be efficiently maintained in active chromatin and to associate with early replicating chromosomal sequences **(Figure 3).** Proof of that is the pEPI vector, which is shown to be kept at low copy numbers and to localise at specific regions of the chromosome, being pseudo-symmetrically distributed in daughter cells after mitosis [149].

1.4.3 Limitations of S/MAR vectors

Although promising, the originally described S/MAR-based vectors still had some limitations. Once the vector was delivered to the target cells and reached the nucleus, it established stochastically as an episome depending on the nuclear compartment where it falls, that means depending on the chromatin structure and ongoing transcription **(position effect).** If the region of the nucleus is actively transcribed, the vector will establish episomally but if the region is transcriptionally inactive, the vector will integrate or get lost during subsequent cell divisions [149]. However, once established the vector will be maintained through limitless cell divisions. [195, 196]. This stochastic localisation in the 'correct' nuclear compartment results in a relatively low efficiency of establishment, around 9.5 ±7,5% for the originally described pEPI vector [197]. To overcome this limitation, several improvements have been made in order to achieve higher levels of expression, improve the long-term expression and avoid the silencing due to position effect variegation. The most recently described S/MAR vectors were published by Hagedorn *et al.*, in which a Ubiquitous Chromatin Opening Element (UCOE) was added to the vector and significantly improved transgene expression by keeping the DNA in the 'active areas' and avoiding heterochromatic silencing. In addition, a chicken insulator (cHS4) was also added which improved the vector establishment via interaction with the nuclear matrix [198].

1.4.4 Potential applications of the vector

Episomal S/MAR vectors are an alternative system to the vectors described before and can be used for a range of purposes generating minimal impact to the host cell or organism. For example, S/MAR episomal vectors offer the possibility to efficiently and persistently (over)express a gene of interest, or to produce shRNA stably, or to express CRISPR/Cas9 for gene editing approaches persistently. Their versatility in genetic modification opens a myriad of suitable applications, amongst them are the genetic modification of **stem cells**, the introduction of new genetic traits **(transgenesis)** into organisms as well as the supplementation of 'corrected' therapeutic genes **(gene therapy)** in monogenic diseases where the mutated gene does not have a dominant adverse effect.

1.5 Stem cells and reprogramming

Pluripotent stem cells, including **embryonic stem cells (ESCs)** and **induced pluripotent stem cells (iPSCs),** are considered to be a prime cell source for regenerative therapies and cell therapy applications because of their extensive proliferation, self-renewal potential and multi-lineage differentiation capabilities [199]. SCs are also a valuable tool for modelling diseases *in vitro* and to better understand embryogenesis and early stages of development. This section provides an overview of stem cell biology and offers a recapitulation of different methods to genetically modify SCs and reprogram iPSCs while also discussing the advantages and limitations.

1.5.1 Definition and properties of SCs

Stem cells are undifferentiated, karyotypically normal cells with self-renewing capacity and the ability to differentiate into representatives of the three embryonic layers. One crucial *stemness* feature is **self-renewal**, which consists of the ability of cells to generate at least one identical daughter cell. Another defining feature is the **unlimited proliferation** while maintaining an undifferentiated state [200]. Stem cells can be classified based on their **potency**, which is their ability to differentiate. Therefore, **totipotent** (or omnipotent) cells are found early in development and can generate both embryonic (endoderm, mesodermal and ectodermal) and extraembryonic (trophectoderm) lineages. **Pluripotent** cells can generate cells from the three germ layers and germline but not extraembryonic tissues. **Multipotent** cells are much less plastic and can give rise to a limited subset of cells within the same germ-layer. **Oligopotent** stem cells can only differentiate into few cell types, and **unipotent cells** can only produce one cell type but still retain the self-renewal capabilities (as opposed to progenitor cells).

Stem cells can also be classified based on their origin or stage in which they appear in the organism's lifetime. **Embryonic Stem Cells (ESCs)** are pluripotent cells isolated from the inner cell mass (ICM) of early-stage blastocysts. **Somatic (adult) Stem Cells (SSCs)** are scarcely found in adult tissues and serve as a cell source for tissue repair. Most adult stem cells are lineage-restricted (multipotent). **Foetal Stem Cells (FSC)** are multipotent cells that can be isolated from foetal blood, bone marrow and other tissues, such as kidney or liver [201].

Finally**, induced (pluripotent) stem cells (iPSC)** are derived from somatic cells that have regained their pluripotent properties after receiving a defined set of reprogramming factors, such as Oct3/4, Sox2, Klf4 and cMyc in murine fibroblasts (MEFs) [142] or Oct3/4, Sox, Klf4, cMyc or Oct3/4, Sox2, Nanog,

Lin28 and L-Myc for human dermal fibroblasts [202, 203]. iPSC offer several advantages when compared to ESC: 1) they eliminate the ethical concerns associated with embryonic stem cells, 2) they are derived from somatic cells that can be obtained via non-invasive techniques, 3) they are individual-specific, eliminating the risk of immune rejection and 4) they can be used for personalised medicine, drug screenings and disease models [204].

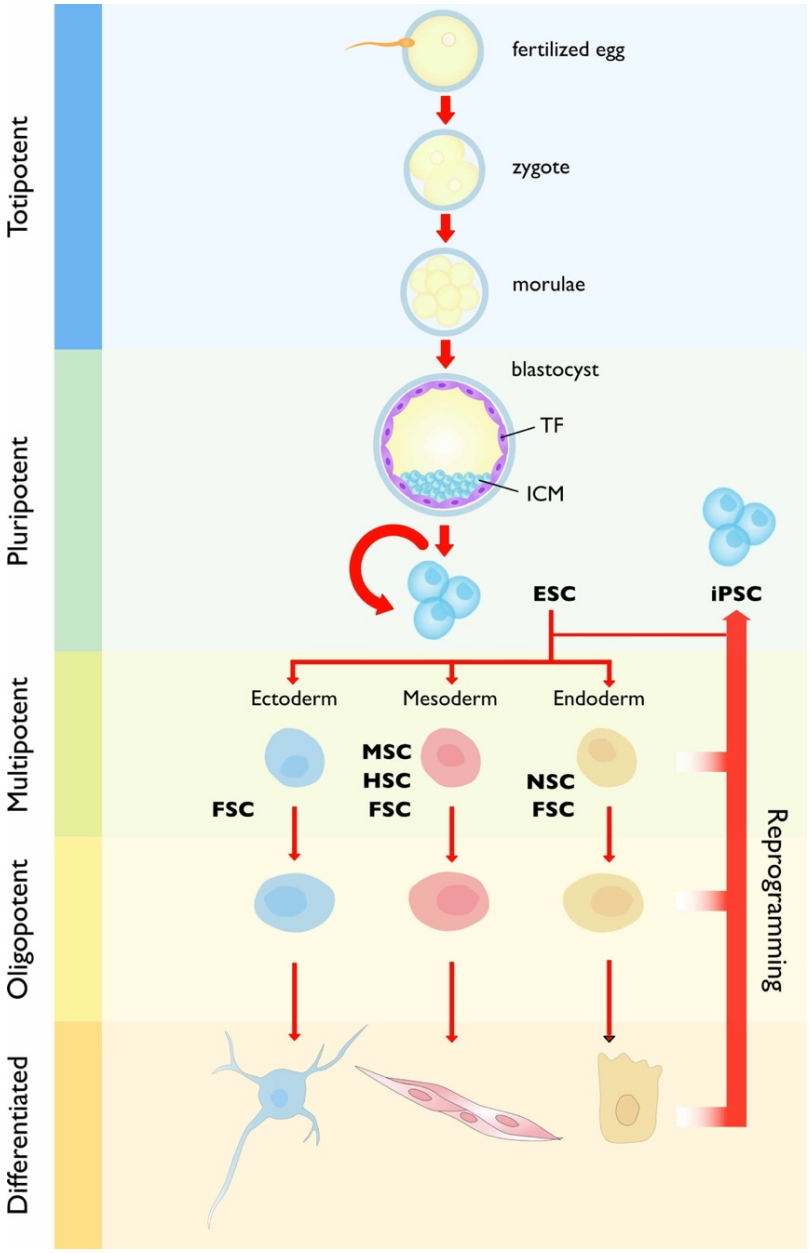


Figure 6: Stem Cell hierarchy

Totipotent cells are found in the first stages of embryonic development (zygote to *morulae*) and retain the ability to differentiate into embryonic (ICM) and extraembryonic (Trophectoderm, TF) tissues. Pluripotent cells, such as embryonic stem cells (ESC) are originally found in the inner cell mass (ICM) of early-stage blastocysts and can only give rise to representatives of the three embryonic layers (endoderm, mesoderm and ectoderm). Multipotent stem cells, such as mesenchymal stem cells (MSC), hematopoietic stem cells (HSC), neuronal stem cells (NSC) or foetal stem cells (FSC) can differentiate into a subset of cells from the same germ layer. Oligopotent cells can only differentiate into few cell types. Terminally differentiated cells, as well as intermediate pluripotent cells, can be reprogrammed into induced pluripotent cells (iPSC) upon administration of a defined set of transcription factors.
Adapted from [205]

1.5.1.1 Cellular properties

Stem cells are **round and compact** cells with **large nuclei**. Murine Stem Cells (mSC) form dense dome-shaped colonies while human stem cells (hSCs) are flat and sharp-edged. They express high levels of **telomerase** to ensure **active proliferation** and **self-renewal**. To maintain SCs in an undifferentiated state, they grow on layers of **feeder cells** (inactivated murine embryonic fibroblasts, MEFs), which on the one hand, support SC growth by providing growth factors and on the other hand, prevent spontaneous differentiation. To reinforce pluripotency and prevent spontaneous differentiation, SC cultures are supplemented with **anti-differentiation cytokines**; for instance, mSCs are supplemented with leukaemia inhibitory factor (LIF), while hSCs are grown in serum replacement supplemented with fibroblast growth factor (FGF2).

1.5.1.2 Pluripotency

SCs are characterised by expression of **pluripotent markers**. Both murine and human SCs express alkaline phosphatase (AP)-related antigens, although they differ in the surface antigens expressed. While murine cells are characterised by the expression of the glycolipid antigen SSEA-1, human cells express SSEA-3, SSEA-4, the keratin sulphate antigens Tra-1-60 and Tra-1-81; and the protein antigens CD9 and CD90 (Thy1), among others.

The gene **expression profile** of pluripotent cells is characterised by intense expression of development-related genes such as Oct3/4, Nanog and Dnmt3b.

Pluripotent cells can form **embryonic bodies (EBs)** when differentiation inhibitors are withdrawn from the media. EBs consist of a core of mitotically active and dividing cells and a periphery of differentiating cells, in which representatives of the three germ layers can be found.

When injected into immunodeficient mice, pluripotent cells spontaneously generate **teratomas,** which are tumours consisting of cells derived from the three germ layers. Teratoma formation is a landmark test for pluripotency. A more stringent pluripotent test consists of the ability of pluripotent cells to generate **chimaeras** when injected into the inner cell mass of embryos. Finally, pluripotent cells can develop into viable non-chimeric mice when injected into tetraploid blastocysts, which can only form extraembryonic tissues, a process known as **tetraploid complementation** [206, 207]**.**

1.5.1.3 Epigenetic status

During fertilisation, hypermethylated gametes, as well as the fertilised zygote, undergo rapid global demethylation to reset the epigenetic landscape of the embryo and turn into a 'blank canvas', in which epigenetic profiles can be 'painted' as cells differentiate.

With this epigenetic reset, early embryos (blastocysts) and therefore stem cells, have the potential to take all cell fate decisions and differentiate into specific lineages [140], gradually losing their potency and acquiring specific epigenetic modifications (remethylation) [208, 209]. However, this differentiation–acquired methylation primes the cells to a somatic fate that must be reverted and reset in primordial germ cells (PGC), which will give rise to the germ-line and must ensure the necessary plasticity to generate games and a totipotent zygote in the next generation.

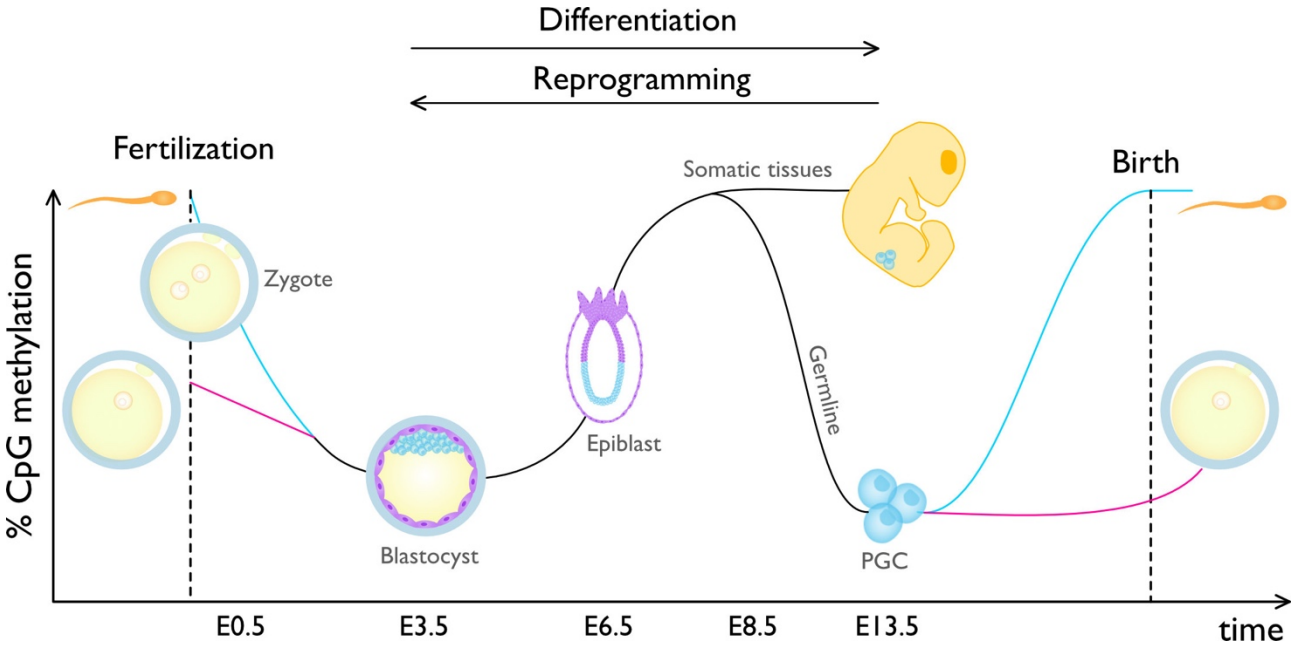


Figure 7: Epigenetic landscape during (murine) development.
The hypermethylated gametes and zygote are quickly demethylated after fertilisation to erase epigenetic marks and turn the genome into 'blank canvas' in which epigenetic profiles can be 'painted' when cells differentiate. Primordial Germ Cells (PGC), which are somatically primed, must be reset back to a 'blank' state so that germ cells can give rise to totipotent zygotes. Somatic cells can be reverted to a pluripotent state by a defined set of reprogramming factors. Derivation of iPSC is a very inefficient process which can be enhanced by using agonist or antagonists of chromatin remodelling molecules.

Maintenance of stemness, lineage commitment (differentiation) and cellular fate are tightly and finely controlled by epigenetic mechanisms such as DNA methylation, histone modifications, remodelling of chromatin structure, non-coding RNA, miRNA and TFs [210]. For instance, **promoters** of pluripotency-associated genes, such as Oct4 or Nanog, are **demethylated** in pluripotent cells, indicating

active expression of pluripotency-associated genes. In addition, **H3 histones** associated with pluripotent genes (Sox2, Oct4 and Nanog) are also demethylated [211, 212]. At the **microRNA level** (miRNA), miRNA302/367 are expressed in embryonic stem cells and are involved in maintaining pluripotency by being direct targets of pluripotency genes and by regulating the cell cycle [213].

Although generally stable *in vivo,* the epigenetic balance defining cell fate can be manipulated and reversed *in vitro,* as demonstrated by Gurdon *et al.,* who showed that somatic cells could be reversed back to a pluri- or totipotent state by transferring somatic nucleus to oocyte environment [214]. This could also be achieved by cell fusion [215] or ultimately, by overexpressing a set of crucial reprogramming factors [142]. However, the frequency of reversing somatic cells back to pluripotency is very low but can be externally influenced by using chemical regulators of chromatin restructuring enzymes [141]. For example, using inhibitors of histone deacetylases (HDACs), such as **Valproic Acid (VPA)** or cofactors of the histone demethylases Jhdm1 a/b, such as **Ascorbic acid/Vitamin C.** Ascorbic acid not only mediates demethylation of H3K26 but is also involved in accelerating the cell cycle, inhibiting senescence (repression of p53/p21) and apoptosis; being an agonist of the Hypoxia Inducible Factor (HIF) pathway as well as upregulating the expression of the aforementioned miRNA 302/367 cluster [216]**.**

1.5.2 Genetic modification of pluripotent cells

Traditional approaches to genetically modify ESC or derive iPSC, require the use of Lentiviral vectors, which integrate into the host cell genome and might potentially cause insertional mutagenesis [5, 58]. Other approaches rely on the use of adeno associated viruses, which still retain the potential to integrate [65]. Alternatively, episomal modification of pluripotent cells is based on the EBNA-1 protein, which also has oncogenic potential [217]. In addition, iPSCs used for clinical applications should be devoid of exogenous reprogramming factors, to exclude residual expression and avoid malignant transformation due to their oncogenic potential [218, 219]. This issue has been partially circumvented by developing gene delivery methods that evolved from integrative viruses to episomal modification of stem cells or transient delivery of reprogramming factors to derive iPSC. This chapter provides an overview of current techniques for stem cell modification and reprogramming as well as the current limitations that these are facing.

1.5.2.1 Genetic modification of Stem Cells

Originally, stem cells were transduced using **retroviruses**. For instance, **Moloney Murine Leukaemia Virus (MLV)**, was used to transduce hematopoietic (HSC) and mesenchymal (MSC) stem cells, but resulted in low transduction efficiency together with the short-lasting expression of the transgene [220]. Also, some groups demonstrated transgene silencing during differentiation [221], probably due to methylation on the viral promoter. Another significant limitation was the random integration in untranscribed regulatory regions (5'UTR) and the consequent possible dysregulation of proto-oncogenes and tumour suppressor genes. An example of retroviruses as double-edged swords was a clinical trial in patients with X-linked SCID, in which bone marrow CD34+ cells were transduced with a retroviral MLV vector carrying the therapeutic γ-chain gene. The corrected autologous cells were successfully transplanted back into the patients, with 8/9 successful correction of the immunodeficiency [222]. Unfortunately, 4/9 treated patients developed acute leukaemia as a consequence of insertional mutagenesis [6, 223]. These limitations, together with the inability of retrovirus to infect non-dividing cells, lead to a switch to **Lentiviral vectors,** which have been the tool of choice to modify stem cells, and in particular HSC.

Lentiviral vectors can infect non-dividing cells, although not quiescent G_0 cells, can accommodate larger DNA loads and preferentially integrate into coding regions of genes. A safer version are **Self-inactivating (SIN) lentiviruses**, which contain a deletion in the transcriptional element 3' LTR to avoid generation of new viral genomes in transduced cells [224]. However, lentiviruses share significant drawbacks with their retroviral relatives regarding their potential integrative mutagenesis and that they typically become silenced either directly at the stem cell stage or during differentiation. One explanation is that virally derived promoters and sequences are highly methylated in stem cells and during differentiation [225]. To circumvent the differentiation-induced silencing, Pfaff *et al.* added the **chromatin insulating element UCOE** into SIN-Lentivirus [226]. They showed that by doing so, transgene expression could be maintained during hematopoietic differentiation (independently of using a viral or a housekeeping promoter) and that the presence of a UCOE element reduced promoter CpG methylation by almost 3-fold. However, the UCOE element holds the potential to disrupt neighbouring gene function by divergent transcription, due to its double divergent promoter nature; as well as to affect splicing of surrounding DNA sequences. Also, clonal differences in transcription and gene silencing were found and attributed to the UCOE element [226].

Adeno-associated vectors (AAVs) are amongst the most used vectors for gene delivery and gene therapy to date. However, their use might not be suitable for all cell types, and especially for stem cells, as confronting data suggests that transduction in stem cells is highly variable as well as AAV might induce toxicity in undifferentiated cells [227]. For example, Walsh *et al.* successfully corrected primary peripheral blood CD34+ progenitor cells from FA patients using AAV [228] and Zhou *et al.* demonstrated that blood CD34+ cells could be transduced with AAV without cytokine stimulation [229]. However, Ellis *et al.* showed that HSC were refractory to AAV [230].

Similar to AAV, **Adenoviral** reports on stem cell modification seem to be inconsistent. Some claim that dividing MSC infected with AdV lose expression after three weeks and that AdV has no detrimental effect on the differentiation potential [231]. Others have reported impaired differentiation [232] and high immunogenicity [233], which led to activation of antigen presenting cells (APC), CD4+, CD8+ and cytotoxic cells and to the degradation of viral particles and subsequent silencing of the transgene.

Other methods to modify stem cells include direct delivery of DNA via microinjection [234], electroporation [235], nucleofection [236], calcium phosphate transfection [237], lipofection [238] or cationic polymers [239]. However, the efficiency of these mechanisms is rather low, and the long-term expression of the transgene must be ensured either by integrative mechanisms (transposition) or replicative plasmids.

1.5.2.2 Derivation of iPSC

Ever since Takahashi and Yamanaka demonstrated that differentiation could be reversed by providing somatic cells with the right combination of reprogramming factors [142], several cocktails of transcription factors as well as gene delivery techniques have been developed to improve the still low efficiency of reprogramming.

Originally, Takahashi *et al.* delivered the reprogramming factors (Oct4, Sox2, Klf4 and c-Myc) by means of **retroviral transduction** [142]. Although simple and effective, this method involves the risk of insertional mutagenesis as well as the oncogenic potential of some reprogramming factors, such as Oct4 and cMyc, which can be a concern for clinical applications. This oncogenicity concerns were circumvented by using **excisable Lentiviral systems**, in which LoxP or FRT sites were introduced at the 3'LTR. Upon administration of the respective recombinase after iPSC were obtained, the integrated reprogramming factors could be floxed out of the genome [240, 241]. Alternatively, the use of

transposons to deliver reprogramming factors proved to be efficient at delivering long-term expression and generating iPSC [242 - 244]. However, both excisable lentivirus and transposons retain the ability to integrate and leave behind a 'genomic scar' after homologous recombination. The remaining concerns about genome integrity led to the development of **non-integrative delivery methods**.

For instance, by using **non-integrative DNA viruses**, such as **adenovirus** [91], or RNA viruses **(Sendai-virus)** [245]. However, AAV still retains the potential to integrate and as Sendai viruses, can trigger immune responses. Therefore, a tedious and labour intensive viral particle removal is required for clinical applications of iPSC generated with these methods. Alternatively, **polycistronic minicircles** [246], **plasmids transfections** [218, 247] or delivery of reprogramming factors in the form of **mRNA** and **proteins** [248] have been also used for safer derivation of iPSC, although compromising the efficiency of delivery and hence, reprogramming. Recently, the use of **episomal DNA vectors** based on **SV40 virus,** Epstein-Barr virus **(EBV)** or **S/MARs**, offered a much more attractive alternative to transient or integrative methods [249, 250]. Only episomal vectors based on OriP/EBNA-1 were able to successfully generate iPSC [251, 252]**,** although their use might influence the cells' behaviour by remodelling the chromatin or altering the cell's transcription profile [253].

1.5.3 Applications

The advances in gene modification combined with the unique properties of stem cells, such as their rapid growth, self-renewal and potential to differentiate into multiple cell types; opened a myriad of possibilities and applications for engineered stem cells. Two possible scenarios where modified stem cells and derived iPSC can be used are described in the following sections. On the one hand, murine stem cells can be genetically manipulated to introduce desired traits **(transgenesis)** or a therapeutic gene for pre-clinical validation of gene therapy strategies. iPSC can be derived from patient somatic cells, corrected *in vitro* and used for autologous transplantation for *ex vivo* **cell therapy** approaches.

1.6 Transgenesis

The breakthrough in Transgenic technology came in the 1970s, when it became common to introduce foreign DNA into cells to generate lines of genetically modified organisms. However, transgenesis was not fully recognised until Palmiter *et al.* introduced the growth hormone gene under the expression of the metallothionein promoter into 1-cell stage embryos by pronuclear injection, and the offspring of such injection showed differential growth [13]. However, some problems quickly emerged. For example, the levels of transgene expression were elevated and uncontrollable, which resulted in giant mice overexpressing the growth hormone. Ever since, transgenic techniques have become extremely important since they provide new approaches for life sciences, which cannot be achieved by just culturing cells *in vitro*. Additionally, transgenic technologies are applied for the large-scale manufacturing of biological compounds for therapeutic and pharmaceutic purposes or the study of genetic regulation in animal development and disease. Not least significant, transgenic animal models are a pre-requisite for testing therapeutic genes or molecules before human clinical trial applications. Although a significant improvement on methods to obtain transgenic animals have been made in the past decades, limitations on the efficiency of gene transfer and the control of gene expression remain [254].

1.6.1 How to generate transgenics

To generate transgenic animals, DNA coding for a desired transgene has to be introduced into the embryos of the desired animal model and must be transmissible to the progeny. To date, there are two techniques to introduce the DNA and generate transgenics: 1) a **direct approach** in which foreign DNA is delivered into the pronucleus of a 1 cell-stage fertilised zygote or 2) an **indirect approach** where DNA is introduced into embryonic stem cells, which later on will be injected into a later stage embryo (Blastocyst).

1.6.1.1 Indirect '*ex vivo*' approach: Modification of Stem Cells

This method relies on the genetic modification of Stem Cells (SCs), which are then reimplanted back into the embryo. First, SCs are removed from the inner cell mass (ICM) of early blastocysts. Then, they are cultured, modified *in vitro*, selected and screened to select for cells containing the desired genetic modification.

Since SCs are pluripotent and responsible for generating all cell types of the embryo, they can be reintroduced back into the embryonic stage from where they are derived, after being genetically modified. This double origin embryo containing unmodified and genetically modified cells is called a **chimaera.** The born offspring will transmit the genotype as long as the modified transgenic cells contributed to form the germ cells. The mechanisms by which the transgene is delivered into the cells are reviewed in **Sections 1.2.3, 1.2.4 and 1.5.2.1.**

1.6.1.2 Direct approach: Targeting the zygote

1.6.1.2.1 DNA microinjection

Pronuclear DNA injection (PNI) relies on the microinjection and integration of linear DNA into fertilised zygotes at 1-cell stage and the subsequent integration of such DNA in the chromosomes [69]. To ensure that the transgene is equally distributed in all cells, it is necessary to introduce the DNA at very early stages, when the embryo is still at a 1-cell stage and before the first cell division, and cleavage occurs. This developmental stage, also called Zygote, occurs following the entrance of the sperm cell into the oocyte and it is easily identifiable because the two pronuclei are still visible individually and have not fused yet. The transgene or foreign DNA can be injected into either pronucleus, although injection into the male pronucleus is more common due to its slightly larger size and proximity to the cell surface. After microinjection, the developing embryos are transferred into foster pseudo-pregnant females. To preserve and propagate the animal line, germ-line transmission must be achieved, meaning that the gonads have integration of the transgene, which can be transmitted to subsequent generations.

1.6.1.2.2 Transposons

Pronuclear injection of a recombinant transposon is one way to ensure the integration and transmissibility of a transgene and has been successfully used to generate transgenic fish, chicken and mammals [255]. Also, Ivics *et al.* generated transgenic pigs by cytoplasmic injection of Sleeping Beauty transposon and achieved germline transmission [256].

1.6.1.2.3 Lentiviral transduction

Lentiviral transduction of 1-cell embryos is efficient in several species, including mice, rats, chicken and large farm animals (reviewed in [257]). High ratios of transgenic offspring can be achieved (20-30%) with almost 90% of the injected animals expressing the transgene [258].

However, the low cargo capacity (7-8kb) of the virus and the epigenetic silencing of integrated constructs limit their potential [225, 259].

1.6.1.2.4 Intracytoplasmic sperm injection (ICSI)

Intracytoplasmic sperm injection of membrane permeabilised spermatozoa incubated with exogenous DNA has been used to generate transgenic mice and pigs [254]. Advantages of ICSI are the possibility to introduce large DNA fragments and the rapid integration of the transgene at one cell stage. A more complex possibility consists of modifying sperm precursor cells *in vitro* and reintroduce them back to the testes for generation of modified sperm.

1.6.1.2.5 Gene targeting

The previous mechanisms result in uncontrolled gene integration; although not entirely random, since the DNA is preferentially integrated into gene-rich regions. A way to precisely target a specific genomic region would be using gene targeting methods [260]. The use of Zinc Finger Nucleases (ZFN), Transcription Activator-Like Effector Nucleases (TALENs) or CRISPR/Cas9 technologies directly into the 1-cell embryo, allow the generation of knock-in and knock-out lines. The problems associated with these targeted techniques are the sophisticated design of specific nucleases, possible off-target effects and tedious genotyping.

1.6.1.2.6 Episomal vectors

Current episomal vectors rely on episomally replicating viral vectors, namely EBV or SV40- derived vectors, or on plasmid vectors based on viral proteins. The development of self-replicative S/MAR based vectors (pEPI) bypassed the use of virally-derived systems. pEPI was delivered via sperm-mediated gene transfer (SMGT) into pig embryos and generated transgenic foetuses [261]. Although expressing the episomal vector, these foetuses were not brought to term and germ-line transmission could not be assessed.

1.6.2 Summary of the section

Generation of transgenic animals is a valuable tool for basic research, pre-clinical models for gene therapy, xenotransplantation, industrial production of drugs or improvement of livestock. Currently, a range of methods is available to deliver DNA directly into the 1-cell embryo or to genetically modify stem cells *in vitro*, which will then be implanted back into embryos. However, these methods rely mostly on integrating techniques, such as transposition or viral vectors (Lentivirus), episomally replicating (AAV) vectors or plasmid-DNA vectors based on viral proteins (EBNA). Some attempts to generate transgenic pigs have been made by using the episomal S/MAR-based vector pEPI. However, aspects like transgene expression in adult animals or germ-line transmission were not addressed.

1.7 Cell and Gene Therapy

Gene therapy can be defined as the therapeutic use of nucleic acids as drugs to replace, repair or regulate dysfunctional genes to prevent or treat diseases. **Cell therapy** is the use of modified cells as vehicles to achieve the same goals.

This novel branch of modern medicine emerged as a consequence of the recombinant DNA revolution together with the improvements in vector delivery and materialised in the 1990s with the first attempts to cure diseases using integrative retroviruses and AAVs [262]. However, serious safety concerns and side effects required a better understanding of biology, virology and immunology to develop better delivery vectors and mechanisms. A decade later, Lentiviral vectors became the predominant choice in the gene therapy arena. Their improved delivery efficiency together with the ability to accommodate larger DNA loads made them the preferred tool to treat immunologic conditions and hemoglobinopathies. More importantly, their capacity to infect non-dividing cells opened up the possibility to modify differentiated cells and to treat neuronal and muscular disorders. On the other hand, improvements on AAV technology favoured their *in vivo* administration to target retinal, hepatic and nervous system diseases. A few years later gene editing techniques came into play. Differently from viral and DNA vectors, which are genome 'additive', editing technologies based on engineered nucleases can add, remove or correct pre-existing genetic information. Like other gene modification technologies, this can be done *in vivo* to modify the genome *in situ*, or to modify cells *ex vivo*.

Originally, gene therapy was envisioned as a treatment for **monogenic disorders**, with the idea that a single administration of the 'corrected gene' would be sufficient to generate a life-long benefit. For example, the first successful cell-based clinical trial for a monogenic disorder, mentioned earlier in this introduction, was aimed to correct an X-chromosome linked SCID by using a single administration of retrovirally corrected CD34+ cells. Nowadays, the palette of genes and diseases appointed as candidates for gene therapy has increased significantly, and polygenic diseases, such as cancer, account for the majority of clinical trials [262, 263].

1.7.1 The two roads of gene therapy

Gene therapy can be performed either by direct *in vivo* administration of the therapeutic gene into the living organism **(direct gene transfer)** or by using *ex vivo* modified (stem) cells as vehicles for gene delivery **(cell transfer)** as depicted in **Figure 8.**

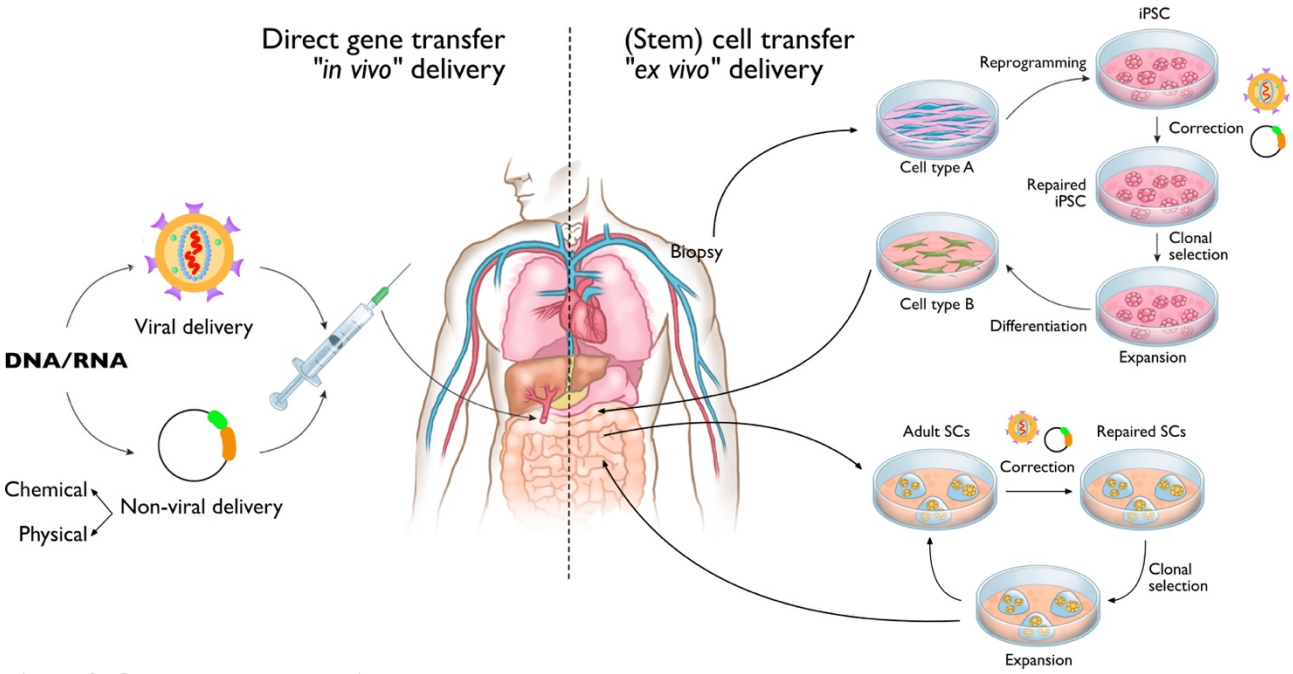


Figure 8: Gene therapy strategies.
Direct gene transfer consists of the *in vivo* delivery of therapeutic DNA/RNA via viral, non-viral or genome editing strategies. Cell transfer, on the other hand, consists of the *ex vivo* correction of either isolated stem cells or somatic cells that can be reprogrammed, corrected, differentiated into other cell types. Upon expansion, these corrected cells are transplanted back into the patient. *Adapted from* [264].

1.7.1.1 Direct gene transfer or '*in vivo*' delivery

In vivo delivery relies on direct tissue targeting via local delivery and avoids the complex problems associated with cell therapy – cell isolation, culturing, modification, expansion and transplantation.

Simplicity is at the same time, its primary weakness; as direct transfer does not allow much control over the therapeutic gene due to integration (Lentivirus) or short-term persistence of episomal forms (AAV). Additional immune responses towards the vector components might occur, which do not happen *in vitro*.

In the past years, clinical trials delivering therapeutic genes to liver, eye and muscle have been accomplished or are on the way to succeed. For instance, different AAV serotypes targeting the liver have been used as tools to deliver factor VIII and IX for the treatment of Haemophilia A and B, respectively. However, in some cases, the expression persisted only for few weeks [60], and the transgene was silenced due to pre-existing neutralising antibodies against the AAV capsid as a consequence of frequent AAV infections in the population [61]. Short term expression was partially solved with short-term immunosuppression [265].

Other examples include the subretinal injection of AAV2 coding for the *RPE65* gene in patients with inherited blindness, who showed clinical improvement after vector administration [266]. In contrast, other clinical trials reported regression of the visual function in some patients [267]. The *in situ* injection of AAV is currently being used in clinical trials for other inherited blindness diseases, such as achromatopsia, X-linked retinitis pigmentosa and choroideremia.

Even though neuromuscular diseases, are multigenic, physiologically more complex and less understood than monogenic disorders, there have been efforts in treating some of these conditions. As an example, AAV2 was used to deliver the dopamine synthesising enzyme Aromatic L-aminoacid decarboxylase (AADC) to enhance the conversion of L-dopamine to dopamine in Parkinson's disease, which is characterised by a loss of dopaminergic neurons [268].

1.7.1.2 Stem cell transfer or '*ex vivo*' delivery

As mentioned earlier, stem cells are a very versatile tool for cell therapy applications, which rely on the infusion or transplantation of living cells into a patient. Although cell-based therapies are inherently more complex than direct *in vivo* approaches due to their handling complexity, the *in vitro* manipulation allows better precision and control over the therapeutic gene than inside the body (*in vivo*). Also, *in vitro* culturing allows getting enough cell numbers required for transplantation. Moreover, some cells (HSC) have 'homing' mechanisms that allow them to return to their original location in the body, conferring regional therapeutical specificity. At the same time, this handling complexity associated with isolation, culturing, correction, expansion and transplantation is a significant limitation and can be complicated further when somatic cells undergo reprogramming and differentiation into a new cell type, before or after genetic correction. Not least important, long culturing can be detrimental and favour the acquisition of deleterious mutations.

Blood transfusions are considered one of the first types of cell-therapy, followed by bone-marrow transplantation and HSC transplantation. In fact, of all types of Stem Cells being used in clinical trials, CD34+ hematopoietic stem cells (HSC) and mesenchymal stem cells (MSC) are mainly exploited due to the ease of collection and high differentiation capabilities [269, 270]. More interestingly, the patient's cells can be used for autologous transplantation to avoid the risk of immune rejection or graft-versus-host disease [271]. However, when a patient carries a genetic disease, autologous cell therapy would not be sufficient as the harvested cells would also carry the genetic mutation after transplantation [272]. To overcome the problem of autologous cell therapy for genetic diseases is necessary to correct the

genetic defect before re-implantation. Although some studies have appointed AAV as successful tools to modify and correct stem cells, they have been replaced by integrating vectors and non-integrating adenoviral vectors [273]. Rapti *et al.* showed that AdVs were more efficient at infecting both undifferentiated and differentiated cells and that Lentivirus was applicable to undifferentiated whereas AAVs were more efficient at delivering to differentiated cells [274]. Also, they showed AAV-induced toxicity in SCs as a consequence of inducing cell cycle arrest and apoptosis.

Alternative to autologous HSC transplantation, patient somatic cells can be isolated, reprogrammed into iPSC, corrected and differentiated to the desired cell type. In some cases, such as Fanconi Anaemia, the genetic mutation impairs the reprogramming capabilities [275], and it is, therefore, necessary to first correct the genetic defect and then derive the iPSC.

1.7.2 Summary of the section

This section reviewed and summarised the advances on development of gene and cell therapy treatments as well as pointed out some of the encountered challenges of therapy vectors, such as safety concerns due to gene integration, short-term expression, silencing during differentiation or immune reactions towards vectors and therapeutic genes. Although successful gene and cell therapy trials have been documented, a significant amount of literature calls into question the suitability of currently used therapy vectors, indicating the need of further vector development and refined vector design. Improved vectors could be used to revisit therapy approaches for the treatment of several types of diseases, such as blood (Fanconi Anemia) and eye disorders (choroideremia), which will be discussed in the following section.

1.8 Potential therapeutic applications of S/MAR vectors

1.8.1 Fanconi Anaemia

Fanconi Anaemia (FA) is a highly heterogeneous autosomal recessive disease characterised by congenital abnormalities, defective haematopoiesis and high risk of developing acute myeloid leukaemia (AML) and solid tumours; both being the primary cause of mortality and morbidity. FA can be caused by a mutation in any of the proteins involved in the FA DNA repair pathway. Although there have been advances in preventive treatment, especially in hematopoietic cell transplantation, a successful curative treatment that solves the problem of genomic instability and risk of developing solid tumours when the patients reach adulthood, still does not exist.

FA is an ideal candidate to be treated with gene therapy because the supplementation of the correct gene is enough to restore the functionality of the pathways as well as to rescue the phenotype. Several gene therapy attempts have been performed using retrovirally-mediated gene transfer of autologous HSC, although success has not yet been achieved. Moreover, virally-mediated gene therapy raises inherent safety concerns due to the random integration of the viral vectors and the potential associated insertional mutagenesis. Therefore, the investigation of gene therapy alternatives it is desirable in this particular disease.

1.8.1.1 Gene therapy strategies

FA is an ideal candidate for gene therapy due to its monogenetic nature and the correction of the disease phenotype by supplementing the right copy of the defective gene. Although there has been successful correction of FancC knockout mice using retrovirally-mediated gene transfer [27], the success in humans has not been as promising [276]. Part of the failure can be attributed to the low success rate of autologous HSC transplantation. HSC in the bone marrow are scarce and hard to mobilise, making it harder to obtain enough cells to transduce and correct *in vitro* and even less successfully corrected cells to transplant back into the patient. Also, retroviral transduction efficiencies are low, although Lentiviruses seem a bit more promising, but still integrate into the genome, deregulating neighbouring genes; or they get silenced due to epigenetic effects [226]. A possible solution to these problems is the derivation of iPSC from the patient's cells [26], which could surpass the difficulty in mobilizing HSC from the bone marrow. However, Müller *et al.* showed that restoration of FA pathway is a pre-requisite for generation of iPSC from FA in mice [275]. FA-affected cells are arrested in G_2 and are very susceptible to DNA damage, which is inherent to the reprogramming process itself. Therefore, FA

affected fibroblasts must be first corrected and then reprogrammed into iPSC. In these studies, the authors correct FancA and FancC knockout cells using retroviral vectors, which randomly integrate into the host cell's genome and can potentially lead to dysregulation of the neighbouring genes. In a context where the cell's DNA is prone to accumulate mutations due to a defective DNA repair pathway, the use of integrative viruses which can themselves mutate the genome is equivalent to the use of a ticking time bomb. Therefore, the use of alternative non-integrative methodologies such as CRISPR/Cas9, designed endonucleases or episomal vectors is desired in this case.

1.8.2 Choroideremia

Choroideremia (CHM) is an X-linked recessive chorioretinal dystrophy caused by mutations in the Rab Escort Protein 1 (REP1) gene. Mutations that lead to loss of REP1 function disrupt normal intracellular trafficking and post-translational lipid modification of Rab small GTPases (Rab proteins) leading to progressive degeneration of the retinal pigment epithelium, photoreceptors, and choroid.

CHM is an ideal target for human gene therapy as there is a detailed understanding of its disease pathophysiology and comprehensive knowledge of its genetics with identified disease-causing variants in the *CHM* gene typically resulting in a loss of REP1. There is also a refined protocol established involving techniques to deliver gene therapy vectors into the subretinal space providing access to the affected cellular layer. Some gene therapy attempts to correct CHM have been performed by using AAV vectors coding for the REP1 gene. Although successful in some patients, some severe adverse effects related to the AAV administration have been reported; indicating the need for further investigation and gene therapy alternatives.

1.8.2.1 Gene therapy strategies

There have been successful attempts to correct deficient cells by lentiviral delivery of a vector coding for REP-1-eGFP *in vitro* and *in vivo* [28]. The transduced cells showed efficient expression of Rep1 as well as rescued functional activity of the protein. Another group used a recombinant adeno-associated virus serotype 8 (rAAV8) vector to correct murine cells *in vivo* and *in vitro* [15]. The first clinical trial for CHM assessed the effects of retinal gene therapy using an adeno associated virus (AAV) vector encoding REP1 (AAV.REP1) in patients with Choroideremia [277]. The choice of vector for this initial trial was greatly influenced by previous ocular gene therapy trials that demonstrated safety and efficacy of subretinal injection of AAV in individuals with other congenital eye disorders.

The initial six-month follow-up report of six patients treated with AAV.REP1 vector described the improvement in both rod and cone function for two of the six treated individuals [278]. In the case of CHM, additional concerns about administrations into the subfoveal space were addressed by the findings that the initial six individuals did not lose a 'clinically significant' degree of visual acuity at six months and the mean thickness of the retina remained unchanged [278]. However, one patient who received a lower dose of the AAV vector did exhibit severe visual acuity loss in the treated eye more than the untreated control. Also, in another cohort of five treated individuals, other measures of visual

function such as microperimetry sensitivity and colour vision decreased and did not recover following subretinal detachment [279].

A more recent trial in Canada, which utilised the same AAV vector, reported a severe adverse event of localised intraretinal inflammatory response with a slow recovery of visual acuity and the structural loss of the Retina Pigmented Epithelia (RPE) complex was reported in one of the treated patients [280].

2. AIMS

2.1 Challenges of current vectors

Current gene therapy vectors are limited not only by the **transient nature of transgene** expression [9] but also by the **impact** that the vector — and its genetic material — has **on the target cell** [59 - 61]. The increasing use of modifying vectors in experimental animal models and in the clinic has raised awareness of **vector-mediated toxicity**. Viruses and bacteria are pathogens, and their target host and its cells have developed a range of mechanisms to detect infection and reduce its damage. Modifying vectors derived from these pathogens can be identified by both the adaptive and the innate immune systems of the target host [117], and these reactions can have a severe consequence on the efficacy of delivery and the host itself. By inducing an inflamed state in a cell, the vector can merely be silenced, or the cell itself can be destroyed to prevent the spread of 'infection' [135, 138]. Additionally, vectors which are specifically designed to integrate into the genome of a host's cells as a means for providing persistent transgene expression also introduce the risk of **random integrative genotoxicity** [4, 6].

Some of these challenges can be overcome by having a closer look at the DNA vector itself and refining its sequence, components and the arrangement of its elements.

2.2 The ideal vector

With clearly established risks involved in the clinical application of integrative or otherwise oncogenic viral vectors and the increasing understanding of vector-mediated toxicity; the design of a novel, safe and efficient gene therapy vector is desired and demands that:

- The delivery system should be **efficient, atoxic, specific** and comprise no bacterial or viral components.
- The vector should **remain extrachromosomal**, to avoid genotoxic effects.
- The genetic component should be **devoid of non-human elements**. Bacterial sequences, which contain unmethylated CpG motifs from bacterial DNA that stimulate innate immune responses, should be removed or minimised.

- The persistence of the vector in the cell must be guaranteed and should be designed to be sustained and replicate in cells episomally. For that, the vector should be **self-replicative and segregate during mitosis.** If the vector is intended to be inheritable through generations, survival through meiosis should also be ensured.

- The vector should remain active throughout the lifespan of a cell, including cellular processes where epigenetic changes and methylation events are common. For that, insulating elements could be included to avoid silencing and negative effects of neighbouring chromatin.

- Specifically designed and selected regulatory sequences, such as promoters, should drive tailored expression of the transgene. If the vector is to be used for clinical applications, expression levels equivalent to those endogenous are desired; whereas if the vector is used to confer new traits, higher levels of expression would be preferred.

- The maintenance and expression of the constructs should have no adverse or toxic consequences for the host at either the cellular or somatic level. That means not altering the cell's properties, such as pluripotency or differentiation potential (in stem cells) or engraftment (tumour cells).

- Size limitation should not be an issue. That would allow inclusion of transgene regulatory sequences to mimic better the cell's expression as well as several expression cassettes that would allow the vector to be multifunctional.

- If the vectors are to be used in clinical applications, they should be designed and implemented specifically for the target disease and cell type.

However, vectors meeting all these criteria are not yet available.

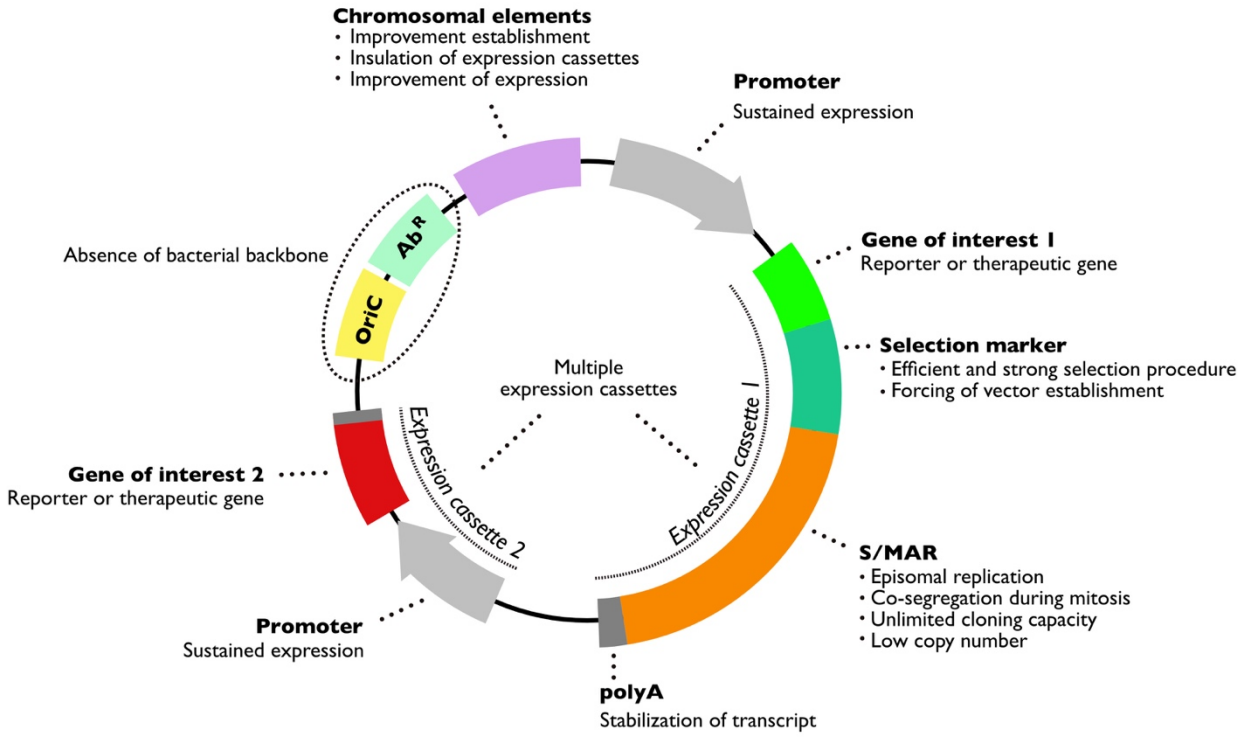


Figure 9: Cartoon depicting an ideal episomal vector.
An ideal episomal vector for a broad range of applications should remain extrachromosomal and avoid integration. Self-replication, episomal retention and co-segregation during mitosis should be ensured by placing an S/MAR element and coupling it to the transcription unit. A polyA tail is added downstream for transcript stabilisation. The choice of promoter and regulatory elements should be adapted and tailored to cell type and purpose, to ensure the right levels of transgene expression. Also, other chromosomal elements, such as insulators, would be desirable to avoid epigenetic silencing. A selection marker should be included when cells were to be modified *in vitro* and should be an active part of the vector establishment. The inclusion of multiple expression cassettes would allow a multifunctional vector. Finally, removal of bacterial sequences, rich in unmethylated CpG, would decrease the plasmid recognition and silencing by the immune system.

2.3 Potential applications of the vector

This work aims to build upon previous studies in which S/MAR vectors have been used to genetically modify dividing cells efficiently and stably [281], without the risk of integration-mediated genotoxicity and providing robust and sustained gene expression [171]. The knowledge in S/MAR DNA vectors combined with the great potential that stem cell technology offers, will be used to develop an alternative, less toxic and safer vector system. This novel vector platform could fulfill all the above-mentioned requirements, can support stable episomal maintenance and persistent gene expression in pluripotent and differentiated cells, for *in vitro* and *in vivo* applications.

3. MATERIALS

3.1 Cells

3.1.1 Human cells

Table 4: Human cells used in this study

Name	Description	Reference/origin
HEK293T	Human embryonic kidney cells expressing adenoviral E1A and E1B and the simian virus 40 (SV40) large T antigen	[282]
HeLa	Henrietta Lacks cervical cancer cell line. Human malignant epithelial cells, derived from an epidermoid carcinoma of the cervix	[283]
Be(2)C	BE(2)-C is a clone of the SK-N-BE(2) Neuroblastoma cell line, established in November of 1972 from a bone marrow biopsy. Does not display MYCN amplification	A kind donation from Dr Jeanninne Lacroix (DKFZ)
LN-18	Grade IV Glioblastoma from human origin	ATCC, CRL-2610. A kind gift from Dr Georgios Giamas
LN-229	Glioblastoma cell line from human origin	ATCC, CRL-2611. A kind gift from Dr Georgios Giamas
U-118	Glioblastoma cell line from human origin	ATCC, HTB-15. A kind gift from Dr Georgios Giamas
U-87	Glioblastoma cell line from human origin	ATCC, HTB-14. A kind gift from Dr Georgios Giamas
U-138	Glioblastoma cell line from human origin	ATCC, HTB-16. A kind gift from Dr Georgios Giamas
T-98	Glioblastoma multiforme cell line from human origin	ATCC, CRL-1690. A kind gift from Dr Georgios Giamas
CLN474 patient HDF	Human Dermal Fibroblasts with p62 amplification	A kind Gift from Prof. Tristan Mckay (MMU)
NHDFs	Neonatal Human Dermal Fibroblasts	Promocell C-12300
hESC	Human Embryonic Stem Cells	A kind Gift from Prof. Tristan Mckay (MMU)
Be(2)C GFP	Be2C cell line labeled with vector 18 (pSMARt_GFP)	generated in this study
Be(2)C Luc	Be2C cell line labelled with vector 50 (pSMARt_Luciferase)	generated in this study
LN-18 GFP	LN-18 Glioblastoma cells labelled with vector 18 (GFP-S/MAR)	generated in this study
LN-229 GFP	LN-229 Glioblastoma cells labelled with vector 18 (GFP-S/MAR)	generated in this study
U-118 GFP	U-118 Glioblastoma cells labelled with vector 18 (GFP-S/MAR)	generated in this study
U-87 GFP	U-87 Glioblastoma cells labelled with vector 18 (GFP-S/MAR)	generated in this study
U-138 GFP	U-138 Glioblastoma cells labelled with vector 18 (GFP-S/MAR)	generated in this study
T-98 GFP	T-98 Glioblastoma cells labelled with vector 18 (GFP-S/MAR)	generated in this study
hiPSC_ EBNA	HDF reprogrammed into hiPSC using EBNA episomal vectors containing Oct4, Klf4, Sox2, Lin28 , L-Myc and shRNA for P53	generated in this study
hiPSC_ S/MAR	HDF reprogrammed into hiPSC using S/MAR episomal vectors containing Oct4, Klf4, Sox2, Lin28, L-Myc and shRNA for P53	generated in this study
hiPSC_ nPOP	HDF reprogrammed into hiPSC using a minimally sized S/MAR vector containing a polycistronic expression cassette with Oct4, Klf4, Sox2 and c-Myc	generated in this study

– Highlighted cell lines are of particular relevance in this study

3.1.2 Murine cells

Table 5: Murine cells used in this study

Name	Description	Reference/origin
PMEF-CFL	Primary Mouse Embryo Fibroblasts, Not Mytomycin C Treated, Strain CF1. Used for Reprogramming experiments.	Merck Millipore, L00183
iMEFs	Immortalised Mouse Embryonic fibroblasts used for feeder layers	A kind gift from Anne Rademacher (PD Dr Karsten Rippe, DKFZ)
BL6 mESC C7	BL6 wildtype Mouse Embryonic Stem Cell line (clone 7)	Kind gift from Paul Kaschtunig, Dr.Milsom Lab (Hi-STEM), DKFZ
E14 mESC	Male embryonic stem cell line derived from mouse strain 129Ola inner-cell masses	A kind gift from Franciscus van der Hoeven (Transgenics Service, DKFZ).
Lung Fibroblasts	BL6 wildtype Lung Fibroblasts	A kind gift from Dr Joschka Willemsen (Binder lab, DKFZ)
iPSC C6	Mouse Embryonic Fibroblasts reprogrammed to iPSC using Lentivirus (OKSMdTom)	generated in this study
E14 v71, v85, v105, v106	mESC labeled with vector 71 (pSMARt_CAG), vector 85 (nSMAR_CAG), vector 105 (nSpliced) or vector 106 (pSMARter)	generated in this study
iPSC GFP	Lung Fibroblasts labelled with vector 72 (pSMARt_SV40LT) and reprogrammed using Lentivirus (OKSMdTom)	generated in this study

– Highlighted cell lines are of particular relevance in this study

3.1.3 Bacteria

Table 6: Bacterial cells used in this study

Name	Description	Reference/origin
ElectroMAX™ DH10B™ Cells	Electrocompetent *E.coli* used for high-efficiency cloning	Life Technologies GmbH 18290-015
MAX Efficiency® DH5a™ Competent Cells	Chemically competent *E. coli* used for routine subcloning	Life Technologies GmbH 18258-012
One Shot® Stbl3™ Chemically Competent E. coli	Chemically Competent *E. coli,* designed especially for cloning direct repeats found in lentiviral expression vectors.	Life Technologies GmbH C7373-03
Stellar Competent Cells	Chemically competent *E. coli* HST08 strain that provides high transformation efficiency used when using InFusion cloning system	Takara Bio Europe Clontech, 636766

3.2 Vectors

3.2.1 DNA vectors

Table 7: DNA vector library

#	Name	Description	Reference/origin
1 2	pMAXcoGFP	Basic pMAX backbone with CMV promoter driving expression of coGFP	Lonza
3	pMAX_coGFP-S/MAR	Basic pMAX backbone with CMV promoter driving expression of coGFP with an S/MAR motif	Cloned in this study
4	pCAG_S/MAR empty	S/MAR based vector containing the CAG promoter	Cloned by Dr Suet P. Wong*
5	pMAX_S/MAR_ intronless	pMAX backbone without 5' intron	Cloned in this study
6	pTurboFP635-N (Katuschka)	Basic backbone containing TurboFP635 reporter gene driven by CMV	A kind gift from Dr Francesca Peri (EMBL). Evrogen.
7	pMAX_CMV:: tGFP-S/MAR	pMAX backbone with turboGFP and S/MAR motif	Cloned in this study
8	pEPI_CMVintron	Basic pEPI backbone with a 5' intron	Cloned in this study
9	pMAX_S/MAR_ G418	pMAX backbone with S/MAR motif and G418 resistance	Cloned in this study
10	pCAG_tGFP_ S/MAR	pEPI backbone with turboGFP under the expression of the CAG promoter	Cloned by Dr Suet P. Wong*
11	pEPI_CMV_S/MAR	pEPI backbone containing MCS and S/MAR motif	Cloned by Dr Suet P. Wong*
12	pMAX_UbC	pMAX backbone with S/MAR motif and the UbC promoter driving expression of coGFP	Cloned in this study
13	pMAX_no Promoter	Promoterless pMAX backbone	Cloned in this study
14	pMAX_UCOE	pMAX backbone with S/MAR motif and the insulating element UCOE	Cloned in this study
15	pSMARt_GFP-2A-G418	pMAC backbone. CMV driving expression of GFP-2A-G418 and coupled to the S/MAR motif	Cloned in this study
16	pSMARt_GFP-IRES-G418	pMAC backbone. CMV driving expression of GFP-IRES-G418 and coupled to the S/MAR motif	Cloned in this study
17	pMAX_GFP-2A-Luc	pMAX backbone with CMV promoter driving GFP-2A-Luciferase (no S/MAR motif).	Cloned in this study
18	pSMARt_GFP-2A-Puro	pSMARt backbone. CMV driving expression of GFP-2A-Puro and coupled to the S/MAR motif. Used for GFP labelling of cells	Cloned by Matthias Bozza* in this study
19	pSMARt_GFP-IRES-Puro	pSMARt backbone. CMV driving expression of GFP-IRES-Puro and coupled to the S/MAR motif	Cloned by Matthias Bozza* in this study
20	pSMARt_GFP-2A-G418-cHS4	pSMARt backbone with CMV::coGFP-2A-G418-S/MAR and the insulating element CSH4	Cloned in this study
21	pSMARt_GFP-IRES-G418-cHS4	pSMARt backbone with CMV::coGFP-IRES-G418-S/MAR and the insulating element CSH4	Cloned in this study
22	pSMARt_PGK:: GFP-IRES-G418	pSMARt backbone with PGK::coGFP-IRES-G418-S/MAR	Cloned in this study
23	pSMARt-Luc-IRES-G418	pSMARt backbone with CMV::fLuc-IRES-G418-S/MAR.	Cloned in this study
24	pPOP	pMAX backbone with SFFV::human Oct4,Klf4,Sox2,cMyc-IRES-dTom-S/MAR	Cloned in this study
25	pMAX_CMV:: hOKSMdTom	pMAX backbone with CMV::human Oct4,Klf4,Sox2,cMyc-IRES-dTom-S/MAR	Cloned in this study
26	pSMARt_UbC:: GFP-IRES-G418	pSMARt backbone with UbC::coGFP- IRES- G418-S/MAR	Cloned in this study

28	pSMARt_GFP-2A-Luc-IRES-Puro	pSMARt backbone with CMV driving coGFP-2A-Luc-IRES-Puro-S/MAR	Cloned in this study
29	pBeast 1	pMAX backbone containing CMV::coGFP-IRES-Puro S/MAR and a second expression cassette with SFFV::hOKSM-IRESdTom	Cloned in this study
30	pSMARt_UbC::GFP-2A-Luc-IRES-G418	pSMARt backbone with CMV driving GFP-2A-Luciferase followed by an IRES-PURO-S/MAR	Cloned in this study
31	pSMARt_UbC::GFP-2A-Puro	pSMARt vector (18) with UbC promoter	Cloned by Terence Osere* in this study
32	pSMARt_UbC::Luc-IRES-G418	pSMARt backbone with UbC driving Luciferase-IRES-G418-S/MAR	Cloned by Terence Osere* in this study
33	pBeast 2	pSMARt backbone with UbC driving GFP-2A-Puro-S/MAR and a second expression cassette with SFFV:: OKSM-IRES-dTom. The reprogramming cassette is flanked by LoxP sites	Cloned by Terence Osere* in this study
34	pMAX_IRES.Puro	pMAX backbone with CMV:: GFP-IRES-Puro-S/MAR	Cloned in this study
35	pPOP	pSMARt backbone with SFFV driving human Oct4, Klf4, Sox2, cMyc-IRES-dTom-IRES-Puro-S/MAR. Reprogramming vector	Cloned in this study
36	pSMARt_CMV::GFP//UbC:Luc	pSMARt backbone with expression cassettes: CMV:GFP-2A-Puro and Ubc::Luciferase	Cloned in this study
37	pGOI	pSMARt backbone with MCS-IRES-Luciferase-S/MAR	Cloned in this study
38	pSMARt_CMV::GFP//UbC:Luc/UCOE	pSMARt backbone with two expression cassettes. CMV::GFP-2A-Puro and Ubc::Luciferase with UCOE	Cloned in this study
39	pSMARt_PGK::rLuciferase	pSMARt backbone with PGK driving renilla Luciferase-IRES-G418-S/MAR	Cloned in this study
40	pMAX_CMV::fLuc-2A-Puro.2LoxP	pSMARt backbone with UbC driving GFP-2A-Puro-S/MAR and a second expression cassette with SFFV::OKSM-IRES-dTom. The reprogramming cassette is flanked by LoxP sites	Cloned by Matthias Bozza* in this study
41	2xLoxP_fLuc.rLuc (no PolyA)	pSMARt backbone containing two expression casettes: CMV driving fLuciferase-2A-Puro-S/MAR and PGK driving rLuc (w/o polyA tail)	Cloned in this study
42	2xLoxP_fLuc.rLuc (with PolyA)	pSMARt backbone containing two expression casettes: CMV driving fLuciferase-2A-Puro-S/MAR and PGK driving rLuc (w/ polyA tail)	Cloned in this study
43	iCre plasmid	Vector containing iCre recombinase, tagged with Myc and containing SV40NLS, driven by the CMV promoter	Cloned by Matthias Bozza* in this study
44	tGFP-2A-iCre	Vector containing GFP-2A-iCre recombinase, tagged with Myc and containing SV40NLS, under expression of the CMV promoter	Cloned by Matthias Bozza* in this study
45	pSMARt_UbC::GFP-2A-Puro.2LoxP	pSMARt backbone with UbC driving GFP-2A-Puro-S/MAR, flanked by LoxP sites	Cloned in this study
46	pMAX_CMV::FANCA-IRES-Luc-S/MAR	pMAX backbone with CMV::Fanconi A-IRES-firefly Luciferase-S/MAR (no selection marker)	Cloned by Terence Osere* in this study
48	PGK_coGFP-S/MAR	pMAX backbone with PGK driving GFP-S/MAR	Cloned in this study
49	PGK_IRES-GFPnls-S/MAR	pSMARt vector with PGK and a MCS-IRES-eGFP-S/MAR	Cloned in this study
50	pSMARt_Luc-2A-Puro-S/MAR.EI40	pSMARt backbone. CMV driving expression of fLuciferase-2A-Puro and coupled to the S/MAR motif and the insulating Element 40. Used for Luciferase labelling of cells	Cloned by Matthias Bozza* in this study
51	pBeast 3.0	pSMARt backbone with UbC driving GFP-2A-Puro-S/MAR and a second expression cassette with SFFV::OKSM-IRES-	Cloned in this study

		dTom. The bacterial backbone and reprogramming cassette is flanked by LoxP sites	
52	pBeast 3.1	pSMARt backbone with UbC driving GFP-2A-Puro-S/MAR and a second expression cassette with SFFV::OKSM-IRES-dTom. The reprogramming cassette is flanked by LoxP sites and the insulating element 40 separating both expression cassettes	Cloned by Bojana Pavlovic* in this study
53	pSMARt_EF1a::GFP-2A-Puro-S/MAR	pSMARt backbone (18) with EF1a driving expression of GFP-2A-Puro-S/MAR	Cloned by Bojana Pavlovic* in this study
54	pBeast 3.2	pSMARt backbone with UbC driving GFP-2A-Puro-S/MAR and a second expression cassette with SFFV::OKSM-IRES-dTom. The reprogramming cassette is flanked by LoxP sites and the insulator UCOE separating both expression cassettes	Cloned by Bojana Pavlovic* in this study
55	pBeast 3.3	pSMARt backbone with EF1a driving GFP-2A-Puro-S/MAR and a second expression cassette with SFFV::OKSM-IRES-dTom. The reprogramming cassette is flanked by LoxP sites and the insulator Element 40 separating both expression cassettes	Cloned by Bojana Pavlovic* in this study
56	pBeast 3.4	pSMARt backbone with EF1a driving GFP-2A-Puro-S/MAR and a second expression cassette with SFFV::OKSM-IRES-dTom. The reprogramming cassette is flanked by LoxP sites w/o insulating element separating both expression cassettes	Cloned in this study
57	pBeast 3.5	pSMARt backbone with EF1a driving GFP-2A-Puro-S/MAR and a second expression cassette with SFFV::OKSM-IRES-dTom. The reprogramming cassette is flanked by LoxP sites and the insulator UCOE separating both expression cassettes	Cloned in this study
58	pSMARt_CMV/EI40	pSMARt backbone with CMV driving expression of GFP-2A-Puro-S/MAR. Contains the insulating Element 40	Cloned in this study
59	pSMARt_UbC/EI40	pSMARt backbone with UbC driving expression of GFP-2A-Puro-S/MAR. Contains the insulating Element 40	Cloned in this study
60	pSMARt_EF1a/EI40	pSMARt backbone with EF1a driving expression of GFP-2A-Puro-S/MAR. Contains the insulating Element 40	Cloned in this study
61	pSMARt_CMV/UCOE	pSMARt backbone with CMV driving expression of GFP-2A-Puro-S/MAR. Contains the insulator UCOE	Cloned in this study
62	pSMARt_EF1a/UCOE	pSMARt backbone with EF1a driving expression of GFP-2A-Puro-S/MAR. Contains the insulator UCOE	Cloned in this study
63	pSMARt_CMV::dTom-IRES-Puro	pSMARt backbone with CMV driving expression of dTomato-IRES-Puro-S/MAR	Cloned in this study
64	pSMARt_CMV::Katy-IRES-Puro	pSMARt backbone with CMV driving expression of Katuschka-IRES-Puro-S/MAR	Cloned in this study
65	pPOP 2	pSMARt backbone with SFFV driving human Oct4, Klf4, Sox2, cMyc-IRES-dTom-S/MAR (w/o Puromycin selection). Reprogramming vector	Cloned in this study
66	pSMARt_EF1a::dTom-IRES_Puro	pSMARt backbone with EF1a driving expression of dTomato-IRES-Puro-S/MAR	Cloned in this study
67	pSMARt_UbC::dTom-IRES-Puro	pSMARt backbone with UbC driving expression of dTomato-IRES-Puro-S/MAR	Cloned in this study
68	pSMARt_EF1a::Kat-IRES-Puro	pSMARt backbone with EF1a driving expression of Katuschka-IRES-Puro-S/MAR	Cloned in this study
69	pSMARt_UbC::Kat-IRES_Puro	pSMARt backbone with UbC driving expression of Katuschka-IRES-Puro-S/MAR	Cloned in this study
70	PGK_FANCA-IRES-GFPnls-S/MAR	pMAX backbone with PGK::FANCA-IRES-eGFPnls-S/MAR	Cloned in this study
71	pSMARt::CAG-coFGP-2A-S/MAR	pSMARt backbone with CAG driving expression of coGFP-2A-Puro-S/MAR. Used for mESC labelling.	Cloned in this study

72	pSMARt_SV40LT-GFP-S/MAR	pSMARt backbone with CMV driving expression of SV40LT-2A-GFP-SMAR and the insulating element 40. used for immortalisation and labelling of MEFs	Cloned by Matthias Bozza* in this study
73	pSMARt-hSK	pMAX backbone with the CAG promoter driving expression of human Sox2-2A-Klf4	Cloned in this study
74	pSMARt-hML	pMAX backbone with the CAG promoter driving expression of human nMyc-2A-Lin28	Cloned in this study
75	pSMARt-hOct4-shP53	pMAX backbone with the CAG promoter driving expression of human Oct4 and U6 driving shRNA for p53	Cloned in this study
76	pSMARt-shP53	pSMARt backbone with CMV driving GFP-2A-Puro-S/MAR and U6 driving shRNA for p53	Cloned in this study
77	pCXLE-gw	(Empty Backbone) Non-integrating (episomal) expression vector with Gateway cassette	A kind gift from Prof. Tristan Mckay (MMU) [284]
78	pCXLE-hOct4-shP53	Integration-free (episomal) expression of human OCT3/4 and shRNA against p53	A kind gift from Prof. Tristan Mckay (MMU) [203]
79	pCXLE-hSK	Integration-free (episomal) expression of human SOX2 and KLF4	A kind gift from Prof. Tristan Mckay (MMU) [203]
80	pCXLE-hML	Integration-free (episomal) expression of human L-MYC and LIN28	A kind gift from Prof. Tristan Mckay (MMU) [203]
81	nanoPOP1 (n35)	Minimal pSMARt backbone with SFFV:OKSM-IRES-dTom-IRES-Puro-S/MAR. No bacterial backbone	Nature Technology (NTC), Lincoln, NE (USA)
82	nanoFanca (n46)	Minimal pSMARt backbone with CMV::FancA-IRES-Luc-S/MAR. No bacterial backbone	Nature Technology (NTC), Lincoln, NE (USA)
83	nanoEF1a (n53)	Minimal pSMARt backbone with EF1a::GFP-2A-Puro-S/MAR. No bacterial backbone	Nature Technology (NTC), Lincoln, NE (USA)
84	nanoPOP2 (n65)	Minimal pSMARt backbone with SFFV::OKSMdTom-S/MAR. No bacterial backbone	Nature Technology (NTC), Lincoln, NE (USA)
85	nanoCAG (n71)	Minimal pSMARt backbone with CAG::coGFP-2A-Puro-S/MAR. No bacterial backbone	Nature Technology (NTC), Lincoln, NE (USA)
86	pSMARt_CAG::dTom-IRES_Puro	pSMARt backbone with CAG driving expression of dTom-IRES-Puro-S/MAR.	cloned in this study
89	pSMARt_SFFV::vLuc	pMAX backbone with SFFV driving expression of vLuc and S/MAR motif	cloned in this study
90	pSMARt_NFKB::GFP-2A-Luc	pMAX backbone with the NFKB activating region driving expression ofnLuc and GFP with S/MAR motif	cloned in this study
91	nanoBeast 3.2 (54)	Minimal pSMARt backbone with UbC driving GFP-2A-Puro-S/MAR and a second expression cassette with SFFV::OKSM-IRES-dTom. The reprogramming cassette is flanked by LoxP sites and the insulator UCOE separating both expression cassettes	Nature Technology (NTC), Lincoln, NE (USA)
92	nanoBeast 3.4 (56)	Minimal pSMARt backbone with EF1a driving GFP-2A-Puro-S/MAR and a second expression cassette with SFFV::OKSM-IRES-dTom. The reprogramming cassette is flanked by LoxP sites w/o insulating element separating both expression cassettes	Nature Technology (NTC), Lincoln, NE (USA)
93	nanoBeast 3.5 (57)	Minimal pSMARt backbone with EF1a driving GFP-2A-Puro-S/MAR and a second expression cassette with SFFV::OKSM-IRES-dTom. The reprogramming cassette is flanked by LoxP sites and the insulator UCOE separating both expression cassettes	Nature Technology (NTC), Lincoln, NE (USA)

94	nanoFanca (70)	Minimal pMAX backbone with PGK::FANCA-IRES-eGFPnls-S/MAR	Nature Technology (NTC), Lincoln, NE (USA)
95	pCXLE-OKSM-dTom-Puro	Integration-free (episomal) vector containing expression of human codon optimised Oct4, Klf4, Sox2 and cMyc, coupled to an IRES dTomato and IRES Puromycin	cloned in this study
96	pCAG_ΔS/MAR	SMAR deficient version of vector 71	
97	MR226105	Commercial vector encoding for murine P62	A kind gift from Prof. Tristan Mckay (MMU)
98	mP62_S024	mP62 S024 mutant generated for Prof. McKay	cloned in this study
99	mP62_S351	mP62 S351 mutant generated for Prof. McKay	cloned in this study
100	mP62_W340	mP62 W340 mutant generated for Prof. McKay	cloned in this study
101	EBNA_GFP-2A-Puro	pCXLE backbone expressing EBNA and GFP-2A-Puro. For labelling of hESC.	cloned in this study
102	pSMARt_GFP-2A-Luc	pSMARt backbone with CAG driving expression of coGFP-2A-Luciferase and IRES-Puro-S/MAR. Improved version of vector 71 for *in vivo* and *in vitro* applications	cloned in this study
103	Rep1-GFP	Gene corrective vector encoding for human Rab Escort Protein 1 (Rep1)-GFP under CMV promoter and shielded by Element 40.	Cloned by Matthias Bozza* in this study
104	nRep1-GFP	Minimal backbone devoid of bacterial sequences encoding for human Rab Escort Protein 1 (Rep1)-GFP under CMV promoter and shielded by Element 40.	Nature Technology (NTC), Lincoln, NE (USA)
105	nSMAR_Spliced	Minimal backbone containing a spliced S/MAR region, which will be removed after transcription	Nature Technology (NTC), Lincoln, NE (USA)
106	pSMARter	pSMARt backbone with a minimally sized S/MAR motif from human Apolipoprotein B	Cloned by Matthias Bozza* in this study

- Highlighted vectors are of special relevance in this study
- Marked names (*) correspond to former or current members of the DNA Vector Lab, DKFZ

3.2.2 Viral vectors

Table 8: Viral plasmids used in this study

#	Name	Description	Reference/origin
27	pRRL.PPT.SF. hOKSM-IRES-dTom	Lentiviral vector encoding for Oct4, Klf4, Sox2 and cMyc and dTomato as a reporter gene	A kind gift from Dr.Milsom (Hi-STEM, DKFZ) and Prof. Dr Axel Schambach [285]
47	pFANCA_ S11FAIEGnls	Lentiviral vector with MSCV::FANCA-IRES-eGFP-nls	A kind gift from Dr.Milsom (HiSTEM, DKFZ) [286]
	pCMV-ΔR8.91	coding for HIV Gag-Pol	A kind gift from Dr Joschka Willemsen (Binder Lab, DKFZ) [287]
	pMD.2G	coding for VSV-G glycoprotein	A kind gift from Dr Joschka Willemsen (Binder Lab, DKFZ) [287]
	pWPI-Puro	scramble vector	A kind gift from Dr Joschka Willemsen (Binder Lab, DKFZ) [287]

3.3 Cell culture

3.3.1 Cell culture components

Table 9: Cell culture components

Reagent	Company	Catalog Number
Accutase	Gibco (Life Technologies GmbH)	A11105-01
Collagenase III	Worthington	LS004183
Dimethyl sulfoxide (DMSO)	Carl Roth GmbH,	4720.1
DPBS (no calcium, no magnesium)	Gibco (Life Technologies GmbH)	14190094
Dulbecco's phosphate-buffered saline (1xPBS)	Sigma-Aldrich	D8537-500ML
Dulbecco's Modified Eagle's Medium (DMEM) - high glucose, With 4500 mg/L glucose, L-glutamine, and sodium bicarbonate, without sodium pyruvate	Sigma-Aldrich	D5796-6X500ML
DMEM:F-12 1:1 Mixture w/15mM Hepes, L-Gln 500ml	Lonza	BE12-719F
DMEM/F-12 + GlutaMAX	Gibco (Life Technologies GmbH)	31331-028
Enzyme-free dissociation buffer	Gibco (Life Technologies GmbH)	13-151-014
ESGRO LIF	Merck Millipore	ESG1106
Fetal Calf Serum (FCS) Superior	Merck Millipore	S 0615
Fetal Bovine Serum (FBS)	Sigma-Aldrich	F7524-500ml
Geneticin^{R} Selective Antibiotic (G418 Sulfate)	Life Technologies GmbH	11811-031
IMDM (with L-glutamine and 25mM EDTA)	Gibco (Life Technologies GmbH)	12440061
Knock Out DMEM	Gibco (Life Technologies GmbH)	10829018
KO Serum replacement (ESC/iPSC)	Gibco (Life Technologies GmbH)	10828010
L-Glutamine (L-Glu)	Gibco (Life Technologies GmbH)	25030-024
MEM Non-essential Amino Acid Solution (NEAA)	Gibco (Life Technologies GmbH)	11140-035
Minimum Essential Medium Eagle (with Earle's salts, L-glutamine and sodium bicarbonate)	Sigma-Aldrich	M4655-500ML
Opti-MEM Reduced Serum Medium	Life Technologies GmbH	31985062
Penicillin-Streptomycin (P/S)	Gibco (Life Technologies GmbH)	15140-122
Protein-free hybridoma medium (PFHM-II)	Gibco (Life Technologies GmbH)	12040077
Puromycin	Applichem	A2856,0010
RPMI-1640 Medium with L-glutamine and sodium bicarbonate, liquid, sterile-filtered, cell culture tested	Sigma-Aldrich	R8758-6X500ML
Trypan Blue solution 0.4%, liquid, sterile-filtered, suitable for cell culture	Biozym	T13001
TrypLE Express enzyme, no phenol red	Gibco (Life Technologies GmbH)	Gibco (Life Technologies GmbH)
Trypsin-EDTA 0,25%	Sigma-Aldrich	T4049-100ML

3.3.2 Other cell culture reagents

Table 10: Other cell culture reagents

Reagent	Company	Catalogue Number
2-Phospho-L-Ascorbic Acid	Sigma-Aldrich	49752-10G
All-Trans-Retinoic acid (ATRA)	Sigma-Aldrich	R2625-50MG
β-mercaptoethanol	Gibco (Life Technologies GmbH)	31350010
Bovine Holotransferrin	Sigma-Aldrich	T1283
CHIR99021 (GSK3 inhibitor)	Sigma-Aldrich	SML1046-5MG
Fibroblast Growing Factor (FGF2)	Peprotech	100-18B
Gelatin from porcine skin	Sigma-Aldrich	G1890-100G
Monothioglycerol (MTG)	Sigma-Aldrich	M6145
PD0325901 (MEK inhibitor)	Sigma-Aldrich	PZ0162-5MG
Plasmocin	Invivogen	ant-mpt

rhBMP-4	R&D	314-BP-050
rhFGF	R&D	233-FB-025
rhVEGF	R&D	293-VE-050
rh/m/rActivin A	R&D	ACFP338
Valproic Acid Sodium Salt (VPA)	Sigma-Aldrich	P4543-10G
Y-27632	Sigma	Y0503

3.3.3 Cell culture composition

3.3.3.1 Cell media composition

Table 11: Cell media composition

Cells	Media	Composition	Vol.
Freezing media	Fetal Bovine Serum (FBS)	90%	40 ml
	Dimethyl sulfoxide (DMSO)	10%	10 ml
HEK293T	DMEM (With 4500 mg/L glucose, L-glutamine, and sodium bicarbonate, without sodium pyruvate)	89%	500 ml
	Fetal Bovine Serum (FBS)	10%	50 ml
	Penicillin-Streptomycin	1%	5 ml
HeLa	DMEM (With 4500 mg/L glucose, L-glutamine, and sodium bicarbonate, without sodium pyruvate)	89%	500 ml
	Fetal Bovine Serum (FBS)	10%	50 ml
	Penicillin-Streptomycin	1%	5 ml
Be(2)C	RPMI-1640 (With L-glutamine and sodium bicarbonate)	88%	500 ml
	Fetal Bovine Serum (FBS)	10%	50 ml
	Penicillin-Streptomycin	1%	5 ml
	Non-Essential Amino acids (NEAA)	1%	5 ml
iMEFs pMEFS-CFL	DMEM (With 4500 mg/L glucose, L-glutamine, and sodium bicarbonate, without sodium pyruvate)	89%	500 ml
	Fetal Bovine Serum (FBS)	10%	50 ml
	Penicillin-Streptomycin	1%	5 ml
mESCs miPSC	KnockOut DMEM	82%	410 ml
	Fetal Calf Serum (FCS) Superior	15%	75 ml
	Penicillin-Streptomycin	1%	5 ml
	Non-Essential Aminoacids (NEAA)	1%	5 ml
	L-Glutamine	1%	5 ml
	B-Mercaptoethanol	0,1 mM	1 ml
	LIF	1:500 dil	500 µl
	CHIR99021 (GSK3 inhibitor)	3 µM	15 µl
	PD0325901 (MEK inhibitor)	1 µM	5 µl
	* plus additives	Other reagents	
mESCs resuspending media	IMDM	88%	500 ml
	Fetal Bovine Serum (FBS)	10%	50 ml
	Penicillin-Streptomycin	1%	5 ml
	L-Glutamine	1%	5 ml
nHDF	DMEM (With 4500 mg/L glucose, L-glutamine, and sodium bicarbonate, without sodium pyruvate)	88%	500 ml
	Fetal Bovine Serum (FBS)	10%	50 ml
	Penicillin-Streptomycin	1%	5 ml
	L-Glutamine	1%	5 ml
hESC hiPSC	DMEM/F-12 + Glutamax	80%	400 ml
	Knockout Serum Replacement (KSR)	20%	100 ml
	Penicillin-Streptomycin	1%	5 ml
	Non-Essential Aminoacids (NEAA)	1%	5 ml
	B-Mercaptoethanol	0,2%	1 ml
	* FGF2 (add fresh)	10ng/µl	50 µl

mEB Differentiation media	Knockout DMEM	83%	83 ml
	Knockout Serum Replacement	15%	15 ml
	L-Glutamine	1%	1 ml
	Non-Essential Aminoacids (NEAA)	1%	1 ml
	Penicillin/Streptomycin	1%	1 ml
	Ascorbic acid	1:1000	
mEB–HSC differentiation media	IMDM	78%	388,75 ml
	EB FCS	15%	75 ml
	Protein Free Hybridoma medium (PFHM)	5%	25 ml
	Penicillin/Streptomycin	1%	5 ml
	L-Glutamine	1%	5 ml
	Ascorbic Acid	50 µg/ml	250 µl
	Monothioglycerol (MTG)	4,5 mM	19 µl
	Holotransferrin	200 µg/ml	1 ml
	+ Cytokines (BMP-4, Activin A, VEGF and FGF2)	see Other reagents	
hEB Differentiation media	Day 0-4 hESC/hiPSC media		
	FGF2	10ng/µl	
	Y-27632	10uM	
	Day 4-15 DMEM	88%	50 ml
	Fetal Bovine Serum (FBS)	20%	10 ml
	Penicillin-Streptomycin	1%	0,5 ml
	L-Glutamine	1%	0,5 ml

3.3.3.2 Other reagents

Table 12: Other reagents

Reagent	Stock Concentration	Diluent	Final Concentration
Activin A	5µg/ml	0.1% BSA, 1% 1M HEPES in PBS	5 ng/ml
All -trans Retinoic Acid (ATRA)	10mM	DMSO	10 µM
Ascorbic acid	500mg/ml	H_2O	50 mg/ml
BMP-4	5µg/ml	0.1% BSA, 1% 1M HEPES in PBS	5 ng/ml
CHIR99021	10mM	DMSO	3 µM
Gelatin	1% w/v	H_2O	0,1%
Geneticin (G418)	50mg/ml	H_2O	1 mg/ml
Holotransferrin	100mg/ml		200 µg/ml
Monothioglycerol (MTG)			4,5 mM
PD0325901	10mM	DMSO	1 µM
Puromycin	1 mg/ml	H_2O	0.5 - 1 µg/ml
Valproic Acid (VPA)	2M	H_2O	2 mM
Fibroblast Growing Factor (FGF2)	10µg/ml	Tris/HCl pH 7.6 + 0.1% BSA	10 ng/ml
Plasmocidin	25mg/ml (treatment)		2.5 mg/ml (prophylactic)
Proteinase K	25 mg/ml	H_2O	600 µg/ml
VEGF	5µg/ml	0.1% BSA, 1% 1M HEPES in PBS	5 ng/ml
Y-27632	10mM	H_2O	10 µM

3.3.4 Transfection Reagents

3.3.4.1 Chemical Transfection

Table 13: Chemical transfection reagents

Kit/Reagent	Company	Catalogue number
Effectene Transfection Reagent	Qiagen	301425
FuGENE 6 Transfection Reagent	Promega	E2691
FuGENE HD Transfection Reagent	Promega	E2311
in vivo JetPEI	Polyplus	201-10G
jetPEI	Polyplus	101-10N
jetPRIME	Polyplus	114-01
Lipofectamine® 2000 Transfection Reagent	ThermoFischer Scientific	11668030
Lipofectamine® LTX Reagent with PLUS™ Reagent	ThermoFischer Scientific	15338030

3.3.4.2 Physical Transfection/Nucleofection

Table 14: Nucleofection reagents

Kit/Reagent	Company	Catalogue number
Mouse ES Cell Nucleofector Kit	Lonza	VPH-1001
Mouse/Rat Hepatocyte Nucleofector Kit	Lonza	VAPL-1004
NHDF Nucleofection Kit	Lonza	VPD-1001

3.4 Bacterial culture

Table 15: Bacterial culture

Media	Composition	Amount
LB media	Tryptone	100 g
	Yeast extract	50 g
	NaCl	50 g
	H2O	Up to 10L
LB agar	LB	500 ml
	Standard agar	10 g
Antibiotic	**Stock concentration**	**Working dilution**
Ampicillin	100 mg/ml in H2O	1:1000
Carbenicillin	50 mg/ml (in Ethanol)	1:1000
Kanamycin	30 mg/ml (in H2O)	1:1000

3.5 Chemicals and reagents

Table 16: Chemicals and reagents

Chemical/Reagent	Company	Catalogue number
1x Stripping Buffer	Thermo Scientific	46430
20x MOPS SDS Running Buffer	Life technologies	NP0001
6X Gel Loading Dye, Purple, no SDS	New England Biolabs (NEB)	B7025S
Acetic acid	Merck	100063
Acetone	Fischer Chemicals	A/0600/17
ACK Lysis Buffer	Sigma-Aldrich	11814389001
Agar	Roth	S210-3
Agarose	Sigma-Aldrich	A9539-500G
Bovine Serum Albumin (BSA) heat shock fraction, protease free, suitable for hybridisation, pH 7, ≥98%	Sigma-Aldrich	B4287-5G
Bromophenol Blue	Merck	108122
Chloroform	Sigma-Aldrich	3221
CloneAmp HiFi PCR Premix (2X)	Clontech	639298
Complete mini EDTA free proteinase inhibitor complex	Roche	11836170001
DNAseI	Roche	4536282001
DNAZap PCR DNA Degradation Solution	Life technologies	AM9890
Ethanol	Fischer Chemicals	E10650DF/C17
Ethylenediaminetetraacetic acid (EDTA)	Acros Organics	147850010
Glycerol	Sigma-Aldrich	15523
HEPES	Roth	9105.4
Hydrochloric acid (HCl)	Sigma-Aldrich	30721
Isopropanol	Sigma-Aldrich	33539-2
ß-Mercaptoethanol (electrophoresis grade)	Gibco	21985-023
Methanol	Sigma-Aldrich	32213-2
Mix dNTPs (25 mM each)	MP Biomedicals	NTPMX250
Paraformaldehyde (PFA)	Sigma-Aldrich	252549
Phenol:Chloroform:Isoamylalcohol (25:24:1)	Sigma-Aldrich	77617-100ML
Powdered milk	Roth	T145.2
Propidium iodide	Life technologies	P3566
Proteinase K	Roche	3115836001
Random Hexamers (dN6)	Roche	11034731001
RNAseA solution	VWR	E866-1ML
RNASeOUT cleaning solution	Life Technologies	10777019
SapphireAmp® Fast PCR Master Mix	Clontech	RR350A
Sodium acetate (3M)	Invitrogen	AM9740
Sodium chloride (NaCl)	Sigma-Aldrich	433209
Sodium deoxycholate	Serva	18330.02
Sodium Dodecyl Sulfate (SDS)	Serva	20750
Sodium hydroxide (NaOH)	Merck	106469
Tris Base	Sigma-Aldrich	T1503
Triton X100	Applichem	A4975,0500
TRIzol reagent	Cellutron Life Technologies	15596-026
Tween 20	Applichem	A1389,0500

3.6 Buffers and solutions

Table 17: Buffers and solutions

Application	Buffer	Reagent	Absolute amount	Relative amount
DNA electrophoresis	50x Electrophoresis Buffer (EB) (2L)	Tris	2M	484,6 g
		Sodium Acetate	0.25M	41 g
		EDTA	0.05M	37,2 g
		pH (acetic acid)	7.8	aprox. 160 ml
		dH2O		Up to 2 L
	1x EB (5L)	50x EB buffer		50 ml
		dH2O		4,5 L
Western Blot	1x Running Buffer	20x MOPS SDS Running buffer	commercial	200 ml
		dH2O		800 ml
	1x Stripping Buffer	Restore PLUS Stripping Buffer (Life Technologies)		
	4x Laemmli Buffer (20ml)	Tris (pH 6,8)	2M	2 ml
		SDS	8%	1,6 g
		100% Glycerol	40%	8 ml
		Bromophenol Blue	0.01%	2 µl
		ß-mercaptoethanol	2%	4 ml
		dH2O		6 ml
	1x RIPA Buffer (50ml)	Tris (pH 7,5)	1M	500 µl
		NaCl	5M	1,5 ml
		EDTA (pH 8,0)	0.5M	100 µl
		SDS	10%	500 µl
		Sodium Deoxycholate	0.5%	0,25 g
		NP-40	10%	5 ml
		Complete mini inhibitor complex		1 mini tablet
	10x TBS (1L)	Tris Base (pH 7,6)	1M	500 ml
		NaCl	5M	300 ml
		dH2O		200 ml
	1x TBS-T (1L)	10X TBS		100 ml
		dH2O		900 ml
		Tween 20		1 ml
	5% milk/TBST (w/v)	Milk powder	5%	5 g
		1x TBST	95%	100 ml
Immunofluorescence Neuroblastoma	Fixation Solution	Paraformaldehyde (PFA)	4%	
		PBS	-	
	Permeabilisation solution	Triton X100	0,2%	
		PBS	-	
	Blocking solution	FCS	2%	
		PBS	-	
	Storage solution	1x PBS		
Immunofluorescence Stem Cells/EBs	Fixation Solution	100% Methanol	-	
		Acetone	-	
	Permeabilisation solution	Tween20	0,1%	
		PBS	-	
	Blocking solution	BSA	0,5%	
		FCS	1%	
		TritonX100	0,1%	
		PBS	-	
	Storage solution	1x PBS	-	

Southern Blot	Depurination Buffer	HCl	0.25M	
	Denaturation Buffer	NaCl NaOH	1.5M 0.5M	
	Neutralisation Buffer (10L)	NaCl Tri-Sodium citrate-$2H_2O$ Tris pH (HCl)	1.5M 0.5M 7.0	1753 g 884 g 605 g approx. 400 ml
	20x SSC	NaCl Tri-Sodium citrate-$2H_2O$	3M 0.3M	
	Church Buffer (1L)	SDS	7%	70 g
		NaPi	0.5M	18,96 g NaH_2PO_4 48,56 g Na_2HPO_4
		EDTA	1mM	2 ml (0,5M EDTA)
		BSA	1%	10 g
	Wash Buffer 1 (100ml)	2x SSC		10 ml (20xSSC)
		SDS	0,1%	1 ml (10% SDS)
	Wash Buffer 2 (100ml)	0.5x SSC		2,5 ml (20xSSC)
		SDS	0,1%	1ml (10% SDS)
Phenol-Chloroform	DNA Lysis Buffer 1 (100ml)	TrisHCl pH 8.0	20mM	
		NaCl	20mM	
		EDTA	20mM	
		SDS	1%	
	DNA Lysis Buffer 2 (100ml)	NaCl	150mM	
		EDTA pH 8.0	10mM	

3.7 Kits

Table 18: Kits used in this study

Kit/Reagent	Company	Catalogue number
Alkaline Phosphatase Staining Kit II	Stemgent	00-0055
Avidin/Biotin Blocking kit	Vector Labs	SP-2001
CloneAmp HiFi PCR Premix	Clontech	639298
Dako Cytomation Target Retrieval Solution Citrate pH 6.0	Dako	S2369
Dako Real Antibody diluent	Dako	S2022
Dako Real Detection System, Peroxidase/AEC, rabbit/mouse	Dako	K5003
Dako Real Peroxidase Blocking Solution	Dako	S2023
rDNAse I	Invitrogen	AM1906
DNeasy Blood and tissue	Qiagen	69504
Dual Luciferase Reporter Assay System	Promega	E1910
EndoFree Plasmid Maxi Kit	Qiagen	12362
gDNA Clean & Concentrator	Zymo research	D4011
GenElute Gel extraction kit	Sigma-Aldrich	NA1111
High Pure RNA Isolation Kit	Roche	11828665001
HotStar Taq Polymerase mix	Qiagen	203203
Phire Tissue Direct PCR Master Mix	Thermo Fischer Scientific	F170S
Pierce™ BCA Protein Assay Kit	Thermo Fischer Scientific	23227
Prime-It II Random Primer Labeling kit	Agilent	300385
QIAGEN Plasmid Maxi Kit	Qiagen	12163
QIAprep Spin Miniprep Kit	Qiagen	27106
QIAquick PCR purification kit	Qiagen	28106
QuantiTect SYBR Green PCR Kit	Qiagen	204143
Quick-gDNA™ MiniPrep - Capped column	Zymo-Research	D3024
RNeasy Mini Kit	Qiagen	74104
SignalFire™ ECL Reagent	Cell Signalling Technology	6883P3

3.8 Markers

Table 19: Ladders and markers

Marker	Company	Catalogue number
100 bp DNA ladder	New England Biolabs (NEB)	N3231L
1 kb DNA ladder	New England Biolabs (NEB)	N3232L
GeneRuler 1 kb DNA ladder	Life Technologies GmbH	SM1331
PageRuler Plus Prestained Protein Ladder	Fermentas	26619

3.9 Enzymes

3.9.1 Restriction enzymes

Table 20: Restriction enzymes used in this study

Enzyme	Company	Catalogue number
Age I/Bsh TI	Thermo Fischer Scientific	FD1464
Ase I	New England Biolabs	R0526S
BamHI	Thermo Fischer Scientific	FD0054
BcuI / SpeI	Thermo Fischer Scientific	FD1254
Bgl I	Thermo Fischer Scientific	FD0074
Bgl II	Thermo Fischer Scientific	FD0084
BmtI / BspOI	New England Biolabs	R0658S
BspEI / Kpn2I	New England Biolabs	R0540S
BsrGI / Bsp1407I	Thermo Fischer Scientific	FD0934
ClaI/Bsu15I	Thermo Fischer Scientific	FD0143
DpnI	Thermo Fischer Scientific	10819410
EcoRI	Thermo Fischer Scientific	FD0274
EcoRV / Eco 32I	Thermo Fischer Scientific	FD0303
Hind III	Thermo Fischer Scientific	15207038
MluI	Thermo Fischer Scientific	FD0564
NdeI	Thermo Fischer Scientific	FD0583
NheI	Thermo Fischer Scientific	FD0973
NsiI/Mph1103I	Thermo Fischer Scientific	FD0734
PstI	Thermo Fischer Scientific	FD0614
SalI	Thermo Fischer Scientific	FD0644
SmaI	Thermo Fischer Scientific	FD0663
SnaBI	New England Biolabs	R0130S
XbaI	Thermo Fischer Scientific	FD0684
XhoI	Thermo Fischer Scientific	FD0694

3.9.2 Other enzymes

Table 21: Other enzymes used in this study

Enzyme	Company	Catalogue number
Alkaline Phosphatase + Buffer	Invitrogen	18011015
Antartic Phosphatase + Buffer	New England Biolabs	M0289S
T4 DNA Ligase + Buffer	New England Biolabs	M0202S
T4 polynucleotide kinase + Buffer	Invitrogen	18004010

Enzyme	Company	Catalogue number
CloneAmp™ HiFi PCR Premix	Clontech	639298
DNAseI	Roche Diagnostics	4536282001
InFusion HD cloning	Clontech	639649
M-MLV Reverse Transcriptase + Buffer	Promega	M1701
Proteinase K	Roche Diagnostics	3115836001
RNAseA	Sigma-Aldrich	10109142001

3.10 Antibodies

3.10.1 Primary Antibodies

Table 22: Primary antibodies used in this study

Antibody	Origin	Specificity	Dilution used	Company	Catalogue number
α-Tubulin	Mouse monoclonal (DM1A)	Human, mouse, rat, porc, cattle, chicken, Gebril, guinea pig, amphibian	WB (1:10.000)	ThermoFischer scientific	62204
αSMA	Mouse monoclonal	Mouse, rat, human, avian	IF (1:100)	Santa Cruz Biotechnology	SC-53142
β3-Tubulin	Mouse monoclonal (2G10)	Mouse, rat, human, bovine	IF (1:100)	Santa Cruz Biotechnology	sc-80005
cMyc	Mouse monoclonal (9E10)	Human cMyc myc Tag	IF (1:150)	Santa Cruz Biotechnology	SC-40
coGFP	Rabbit polyclonal	coGFP	WB (1:5000)	Evrogen	AB513
coGFP	Rabbit polyclonal	coGFP	IHC (1:500)	Abcam	AB290
FancA	Rabbit polyclonal	Human	WB (1:5000)	Merck	AB5063
FoxA2 (HNF3b)	Mouse monoclonal (H4)	Mouse, rat, human	IF (1:100)	Santa Cruz Biotechnology	SC-374376
Gap-43	Mouse		IF (1:500)	Sigma-Aldrich	G9264-100ul
GAPDH	Mouse monoclonal (G-9)	Mouse, rat, human	WB (1:500)	Santa Cruz Biotechnology	SC-365062
Klf4	Goat polyclonal	Mouse, human	WB (1:200)	R&D Systems	AF3158
Klf4	Rabbit polyclonal (H-180)	Human, mouse, rat GKLF	WB (1:500)	Santa Cruz Biotechnology	sc-20691
Lin-28	Mouse monoclonal	Human	WB (1:200)	Santa Cruz Biotechnology	sc-374460
Luciferase	Rabbit polyclonal	Luciferase of Photinus pyralis	WB (1:500)	Santa Cruz Biotechnology	sc-32896
Nanog	Rabbit polyclonal	human, mouse, monkey	IF (1:150)	Abcam	AB80892
Nanog	Mouse monoclonal (1E6C4)	human	IF (1:150)	Santa Cruz Biotechnology	SC293121
Ncam	Coat polyclonal		IF (1:500)	Chemicon	MAB5324
Oct-3/4	Goat polyclonal (N19)	Human, mouse and rat	WB (1:200)	Santa Cruz Biotechnology	sc-8628
Rep1	Mouse monoclonal (2F1)	Human	WB (1:1000)	Santa Cruz Biotechnology	Sc-23905
Sox2	Rabbit polyclonal	Human, mouse	WB (1:1000)	Merck	AB5603

SSEA1	Mouse monoclonal (480)	Human, mouse, rat	IF (1:150)	Santa Cruz Biotechnology	SC21702
SSEA3	Rat monoclonal (631)	Human, mouse, Rat	IF (1:100)	Santa Cruz Biotechnology	SC21703
SSEA3	mouse monoclonal (480)	Human, mouse	IF (1:100)	R&D Systems	MAB1434
SSEA4	Mouse monoclonal (813-70)	Human, mouse, Rat	IF (1:100)	Santa Cruz Biotechnology	Sc-21704
SSEA4	Rat monoclonal (MC-631)	Human, mouse	IF (1:100)	R&D Systems	MAB1435
TRA-1-60	Mouse monoclonal	Human	IF (1:100)	Santa Cruz Biotechnology	Sc-21705
TRA-1-60	Mouse monoclonal (MC-813-70)	Human	IF (1:200)	R&D Systems	MAB4770
CD45 eFluor 450	Rat monoclonal (30-F11)	Mouse, human	Flow Cytometry (0,5ug/test)	eBioscience	48-0451-82
CD11b-APC	Rat monoclonal (M1/70)	Mouse	Flow Cytometry (0,125ug/test)	eBioscience	17-0112-82
B220-APC	Rat monoclonal (RA3-6B2)	Mouse, human	Flow Cytometry (0,25ug/test)	eBioscience	17-0452-83
CD41 PE-Cy7	Rat monoclonal (GK1.5)	Mouse, human	Flow Cytometry (1:300)	eBioscience	25-0041-82
CD170 (ckit)-Alexa 780	Rat monoclonal (ACK2)	Mouse	Flow Cytometry (1:2000)	eBioscience	47-1171-82
CD144/ Vcadherin - PE	Rat monoclonal (11D4.1)	Mouse	Flow Cytometry (1:200)	BD Pharmigen	562243
CD8a PE-Cy7	Rat monoclonal (53-6.7)	Mouse	Flow Cytometry (0,5ug/test)	eBioscience	25-0081-82
B220 PE-Cy7	Rat monoclonal (RA3-6B2)	Mouse, human	Flow Cytometry (0,5ug/test)	eBioscience	25-0452-82

3.10.2 Secondary antibodies

Table 23: Secondary antibodies used in this study

Antibody	Origin	Specificity	Dilution used	Company	Catalogue number
DAPI	-	-	2 ug/ml	Sigma-Aldrich	10236276001
anti-Chicken-Alexa594	Goat polyclonal	Chicken	IF (1:400)	Invitrogen	A-11042
anti-Goat-Alexa488	Donkey polyclonal	Goat	IF (1:2000)	Invitrogen	A-11055
Anti-Goat-Alexa568	Donkey polyclonal	Goat	IF (1:5000)	Invitrogen	A-11057
anti-Goat-Alexa647	Donkey polyclonal	Goat	IF (1:2000)	Invitrogen	A-21447
anti-Mouse-Alexa488	Donkey polyclonal	Mouse	IF (1:2000)	Invitrogen	A-21202
Anti-Mouse-Alexa568	Goat polyclonal	Mouse	IF (1:500)	Invitrogen	AB-175473
anti-Mouse-Alexa594	Donkey polyclonal	Mouse	IF (1:400)	Invitrogen	A-21203
anti-mouse-Alexa647	Goat polyclonal	Mouse	IF (1:2000)	Invitrogen	A-21236
anti-Rabbit-Alexa488	Donkey polyclonal	Rabbit	IF (1:2000)	Invitrogen	A-21206
Anti-Rabbit-Alexa633	Donkey Polyclonal	Rabbit	IF (1:500)	Invitrogen	A-21070
anti-Rabbit-Alexa647	Donkey polyclonal	Rabbit	IF (1:2000)	Invitrogen	A-31573
Anti-Goat-HRP	Donkey Polyclonal	Goat	WB (1:10.000)	Life technology	A15999
Anti-Mouse-HRP	Goat Polyclonal	Mouse	WB (1:10.000)	Life technology	A31430
Anti-Rabbit-HRP	Donkey Polyclonal	Rabbit	WB (1:10.000)	Life technology	A16023

3.11 Primers

3.11.1 Cloning primers

Table 24: Cloning primers used in this study

#	Oligo name	Sequence (5'->3')
1	pMAX_smar.FOR	CTTCGCCAGATCTCGAGCTCGATGATAATGAATGTCTAAGTTAATGCAGAAACGGAGAGACA
2	pMAX_smar.REV	CCTTTTGCTCACATGTAAGATACATTGATGAGTTTGGACAAACCACAACT
3	AmpR_promoter.FOR	TTGACATGCATGGGACGCGGAACCCCTATTTGTTTATTTTTC
4	AmpR_promoter.REV	TCTCCTCCCTTTGCAAAAGCCTAGGCCTCC
5	CMV_SMAR.FOR	CCTTTTGCTCACATGTGGCATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGG
6	coGFP_BglII.REV	AAGAATATCAAGATCATGGAGAGCGACGAGAGCGG
7	preAbP.FOR	AATATTATTGAAGCATTTATCAGGGTTCGTCTC
8	PostAbP.REV	CCAATATTGATTTATGCTATATAACCAATGAATAATATGGCTAATGG
9	CMVturboGFP.FOR	CCTTTTGCTCACATGTGGCATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGG
10	CMVturboGFP.REV	CGCCGAGGCCAGATCTTTATTCTTCACCGGCATCTGCATCCG
11	turboGFP.FOR	GGGGTACCGAAGCCGCTAGCATGGAGAGCGACGAGAGC
12	turboGFP.REV	TATCATCGAGCTCGAGTTATTCTTCACCGGCATCTGCATCCGG
13	Neo/Kan.REV	TCAGAAGAACTCGTCAAGAAGGCGATAG
14	OriC_AmpNeoKanPolyA.REV	AATGTATCTTACATGTTTTCCATAGGCTCCGCCCCC

15	PciI_CMVintroncoGFP.FOR	CCTTTTGCTCACATGTGGCATTGATTATTGACTAGTTATTAATAGTAATCAATTACGGGG
16	PciI_CMVintroncoGFP.REV	GCAGGAAAGAACATGTTCATCGAGCTCGAGATCTGGCG
17	Ori. For	TCTGTCGATACCCCACGAATTCTTGAGATCCTTTTTTTCTGCGCG
18	Ori.Rev	CGCAGGAAAGAACATGATAACTTCGTATAATGTATGCTATACGAAGTTATTTTCCATAGGCTCCGCCCCC
19	SpeI_EF1a.FOR	TTGATTATTGACTAGTAAGCTTTGCAAAGATGGATAAAGTTTTAAACAGAGAGG
20	SpeI_EF1a.REV	CAATAAAGCTACTAGTGGCCGGCCAGCTTGAG
21	SpeI_UbC.FOR	TTGATTATTGACTAGTGGCCTCCGCGCC
22	SpeI_UbC.REV	CAATAAAGCTACTAGTTCTAACAAAAAAGCCAAAAACGGCCAGA
23	SpeI_PGK.FOR	TTGATTATTGACTAGTCCACGGGGTTGGGGTTG
24	SpeI_PGK.REV	CAATAAAGCTACTAGTCCTGGGGAGAGAGGTCG
25	BstAPI_Backbone.FOR	AAATTGCTAACGCAGTCAGTGCCCATTGGGGCCAATACGC
26	NheI_backbone.REV	GACCGGTAGCGCTAGCAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAACCTGAGG
27	P3noSpeI.FOR	TTGATTATTGACTAGCCCCGGGCGCGA
28	P3_SpeI.REv	CAATAAAGCTACTAGTCAGGAGCTTGTGGATCTGTGTGAC
29	Insert Seq. REV	CTCTCCGTTTCTGCATTAAC
30		
31	hPGK.FOR	TGGCATTGATTATTGACTAGTCCACGGGGTTGGGGTTG
32	hPGK.REV	TACCGCAATAAAGCTACTAGTCCTGGGGAGAGAGGTCG
33	EF1a.FOR	TGGCATTGATTATTGACTAGTAAGCTTTGCAAAGATGGATAAAGTTTTAAACAGAGAGG
34	EF1a.REV	TACCGCAATAAAGCTACTAGTGGCCGGCCAGCTTGAG
35	pMAXseq1	TATGTACATTTATATTGGCTCATGTCCAATATGACCGCC
36	pMAXseq2	GGTTACAAGACAGGTTTAAGGAGGCC
37	pMAXseq3	TGTAATGTGGCATTCTGAATGAGATCCC
38	pMAXseq4	GACCCAACACCGTGCG
39	pMAXseq5	GCCCCTGATGCTCTTCGTCC
40	coreUCOE.FOR	ATCTATATCATAATATGTACAGCCTACAGCTCAAGCCACAT
41	coreUCOE.REV	ATGAGCCAATATAAATGTACAAAGGGAATAAGAATTCCCGCCT
42	MCS.For	CCGGTCGCCACGGCCGGATATCCAAGCTTCCA
43	MCS.Rev	GATCTGGAAGCTTGGATATCCGGCCGTGGCGA
44	NLS.For	TCGAGCTCGACCAAAAAAGAAGAGAAAGGTA
45	NLS.Rev	TCGAGTCATACCTTTCTCTTCTTTTTTGG
46	[P]-3NLS.For	[P]TCGAGCTCGACCAAAAAAGAAGAGAAAGGTAGATCCAAAAAAGAAGAGAAAGGTAGATCCAAAAAAGAAGAGAAAGGTATGAC
47	[P]- 3NLS.Rev	[P]TCGAGTCATACCTTTCTCTTCTTTTTTGGATCTACCTTTCTCTTCTTTTTTGGATCTACCTTTCTCTTCTTTTTTGGTCGAGC
48	cHS4.For	GAATTCTGCAGTCGACTATATTCTCACTGAC
49	cHS4.Rev	CCGCGGTACCGTCGACGAGCTCACGGGG
50	GFP_p2A_NeoKan.For	CGCCTTCGCCAGATCTGCCACCAACTTCAGCCTGCTGAAGCAGGCCGGCGACGTGGAGGAAAACCCTGGGCCCATGATTGAACAAGATGGATTGCACGC
51	HindIII.For	GAAAAAAAGGATCTCAAGAAGATCCTTTGATAAGCTTTTCTACGGGGTCTGACGCTCAGTGGAACG
52	HindIII.Rev	CGTTCCACTGAGCGTCAGACCCCGTAGAAAAGCTTATCAAAGGATCTTCTTGAGATCCTTTTTTTC
53	Amp_Fragment.For	TGCTTCAATAATATTAAGCTTCGCGGAACCCCTATTTGTTTATTTTTCT
54	Amp_Fragment.Rev	TTTGGTCATGCCGTCCAGAAGCTTTTACCAATGCTTAATCAGTGAGGC
55	pMAX_Vector.For	AATATTATTGAAGCATTTATCAGGGTTCGTCTCG
56	pMAX_Vector.Rev	GACGGCATGACCAAAATCCCT
57	coGFP.For	CGCTAGCGCTACCGGTATGCCCGCCATGAAGATCGAGTGCC
58	stop_XhoI_3xNLS_XhoI_GFp.rev	TATCATCGAGCTCTCGAGTCATACCTTTCTCTTCTTTTTTGGATCTACCTTTCTCTTCTTTTTTGGATCTACCTTTCTCTTCTTTTTTGGCTCGAGCGAGCGAGATCTGGCGAAGGC
59	IRES_NeoKan_F1.FOR	CGCCTTCGCCAGATCTCGTGCTCGATGACGCCCCCCCCC
60	IRES_NeoKan_F1.REV	TCAATCATGATATCGCTTATCATCGTGTTTTTCAAAGGAAAACCAC

61	IRES_NeoKan_F2.FOR	ATCGAGCTCGAGATCGTCAGAAGAACTCGTCAAGAAGGCGA
62	IRES_NeoKan_F2.REV	GATAAGCGATATCATGATTGAACAAGATGGATTGCACG
63	p2A_NeoKan.FOR	CGCCTTCGCCAGATCTGGCAGCGGCGCCACCAACTTCAGCCTGCTGAAGCAGG CCGGCGACGTGGAGGAAAACCCTGGGCCCATGATTGAACAAGATGGATTGC
64	p2A_NeoKan.REV	ATCGAGCTCGAGATCTTCAGAAGAACTCGTCAAGAAGGCG
65	GFP_p2A_luc.FOR	CTTCGCCAGATCTCGCGCCCGAGCCACCAACTTCAGCCTGCTGAAGCAGGCCG GCGACGTGGAGGAAAACCCTGGGCCCATGGAAGATGCCAAAAACATTAAGAAG GG
66	GFP_Luc.REV	ACTCATCGAGCTCGACTCGAGTTACACGGCGATCTTGCCG
67	AgeI_Luc.FOR	CGCTAGCGCTACCGGTATGGAAGATGCCAAAAACATTAAGAAGGG
68	BglII_Luc.REV	ATCGAGCACGAGATCTTTACACGGCGATCTTGCCGCC
69	AgeI_coGFP.For	CGCTAGCGCTACCGGTATGCCCGCCATG
70	BamHI_cHS4.FOR	CGCGGGCCCGGGATCCTATATTCTCACTGACTCCGTCCTGG
71	BamHI_cHS4.REV	TAGATCCGGTGGATCCGAGCTCACGGGGACAG
72	AgeI_OKSMdTom.FOR	CGCTAGCGCTACCGGTATGGCCGGACACCTGG
73	BglII_OKSMdTom.REV	ATCGAGCTCGAGATCTTTACTTGTACAGCTCGTCCATGC
74	SpeI_SFFV.For	TTGATTATTGACTAGTAGCTAGCTGCAGTAACGCC
75	NheI_SFFV.Rev	ACCGGTAGCGCTAGCTCGGAGGACTGGCGCGCCGT
76	76seq_hOct34.For	GACCCCAGGCGCCGTG
77	77seq_hOct34.Rev	GGGCTTCGAATCTGCAGATGG
78	78seq_hOct34II.For	AGAACAGAGTGCGGGGC
79	79seq_hOct34II.Rev	CCAGAGGAAAGGACACGGG
80	80seq_p2A.For	TCCGGAGCCACCAACTTCTCC
81	81Seq_hKlf4.Rev	ACAGACTCAGGAGGGTGGG
82	82seq_hKlf4.For	CTTCACCTACCCCATCAGAGCCGG
83	83seq_hKlf4II.Rev	AGGTGTGTACAGCTGCTGACG
84	84seq_hKlf4II.For	GCTGTCCAGCAGAGACTGCC
85	85seq_hKlf4II.Rev	CCGGCGTAATCGCAGG
86	86Seq_T2A.For	GGCTCCGGAGAGGGCC
87	87seq_hSox2.Rev	GGTCTTTCTTCTGGGCCGG
88	88seq_hSox2.For	GCCGGCGTGAACCAGC
89	89seq_hSOX2II.Rev	GGGGCGGCAGGTTCAGGC
90	90Seq_E2A.For	CAGTGCACAAACTACGCCC
91	91seq_hmyc.rev	AAGCCGCTCCACATACAGTCC
92	92Seq_hmyc.For	CGCGCAAAGACAGCGGC
93	93seq_hmycII.Rev	CCCTCTTGGCAGCAGG
94	94seq_hMycII.For	CCTGAGACAGATCAGCAACAACCG
95	95seq_Ires.FOr	GCCCCTCTCCCTCCCCC
96	96seq_IRES.rev	CAGAGGCACCTGTCGCC
97	BglII.XhoI_Luc.Rev	ATCGAGCTCGAGATCTTTACACGGCGAT
98	SacI_IRES_Puro.FOR	GTGTAACTCGAGTCGAGCTCCGCCCCCCCCCCCT
99	BsrGI_LoxP_SFFV.FOR	TATCATAATATGTACAATAACTTCGTATAGCATACATTATACGAAGTTATAGC TAGCTGCAGTAACGCC
100	BsrGI_Tom.Rev	CCAATATAAATGTACATTACTTGTACAGCTCGTCCATGCC
101	UCOE_NsiI.For	CCCGGTCTCCTCCCACaGCCTACAGCTCAAGCC
102	UCOE_NsiI.Rev	TGGCCAATATTGACAATgcataAaGGGAATAAGAATTCCCGCCTCCGCG
103	SnaBI.LoxP.SFFV.For	TATCATAATATGTACTACGTAATAACTTCGTATAGCATACATTATACGAAGTT ATAGCTAGCTGCAGTAACGCC
104	AseI.LoxP.Tom.REV	CCAATATAAATGTACATTAATATAACTTCGTATAATGTATGCTATACGAAGTT ATTTACTTGTACAGCTCGTCCATGCC
105	SpeI_SFFV.For	TTGATTATTGACTAGTAGCTAGCTGCAGTAACGCC
106	IRES-dTOm.REV	GGGGGGGGGGCGTTACTTGTACAGCTCGTCCATGCC
107	dTom-IRES.FOR	CGAGCTGTACAAGTAACGCCCCCCCCCCCCTAACGTTACTGGCC
108	BglII_Puro.Rev	ATCGAGCTCGAGATCTTCAGGCACCGGGCTTGCGGGTC
109	Fanca.FOR	CGCTAGCGCTACCGGATGTCCGACTCGTGGGTC
110	Fanca.REV	GGGGGGGGGGCGTCAGAAGAGATGAGGCTCCTGGG
111	IRES.For	GCCTCATCTCTTCTGACGCCCCCCCCCCCTAACG
112	IRES.Rev	GTTTTTGGCATCTTCCATGATATCGCTTATCCTTATCATCGTGT

113	Luc.For	GATAAGGATAAGCGATATCATGGAAGATGCCAAAAACATTAAGAAGGG
114	Luc.Rev	ATCGAGCTCGAGATCTTACACGGCGATCTTGCC
115	115_UbC.For	TATCATAATATGTACGGCCTCCGCGCCG
116	116_EcorV.Luc.Rev	TGAGCCAATATAAATGaTAtCTTACACGGCGATCTTGCCG
117	117_UCOE.For	ATCGCCGTGTAAGATCAGCCTACAGCTCAAGCCAC
118	118_UCOE.Rev	GCCAATATAAATGATAAAGGGAATAAGAATTCCCGCCTCCG
119	119_EcoRV.IRES.For	CGCCTTCGCCAGATATCCGCCCCCCCCCC
120	120_EcorV.dTom.Rev	GGGGGGGGGGCGGATATCTTACTTGTACAGCTCGTCCATGCCGTACAGG
121	Fanca.For	TACCGAAGCCGCTAGATGTCCGACTCGTGGGTC
122	Fanca.Rev	GACCGGTAGCGCTAGTCAGAAGAGATGAGGCTCCTGGG
123	IRES-Luc.For	CTACCGGTCGCCACCCGCCCCCCCCCCCCTAACG
124	Luc.Rev	CATTATCATCGAGCTTTACACGGCGATCTTGCCG
125	125_ires.luc.for	CGCTAGCGCTACCGGCGCCCCCCCCCCCTA
126	126_Fanca.For (new)	TACCGAAGCCGCTAGATGTCCGACTCGTGGGTC
127	127_Fanca.rev (new)	CGCCGGTAGCGCTAGTCAGAAGAGATGAGGCTCCTGGG
128	GFP-2A-puro.Rev	ATCGAGCTCGAGATCTTCAGGCACCGGGCTTG
129	PciI.LoxP.For	AATGTATCTTACATGTCAATAACTTCGTATAGCATACATTATACGAAGTTATT GTGAGCAAAAGGCCAGCAAAAG
130	NsiI.Neo.Rev	TGGCCAATATTGACATGCATGGGAGGAGACC
131	SnaBI.SFFV.For	TATCATAATATGTACTACGTAAGCTAGCTGCAGTAACGCCAT
132	AgeI.Rluc.For	CGCTAGCGCTACCGGTATGACTTCGAAAGTTTATGATCCAGAACAAAG
133	BglII.Rluc.For	ATCGAGCACGAGATCTTTATTGTTCATTTTTGAGAACTCGCTCAACGAAC
134	PGK.Rluc.For	ACGAAGTTATAAGCTCCACGGGGTTGGGGTTG
135	PGK.Rluc.Rev	TGCTCACATGAAGCTTTATTGTTCATTTTTGAGAACTCGCTCAAC
136	136_Fanca1	TGGAGCCTGAAAAAATGCCGC
137	137_Fanca2	CGACCTCAAGGTTTCTATAGAAAACATGGG
138	138_Fanca3	CTCCCAGAGGTGGATGTGGG
139	139_Fanca4	CCAGCCTTCAGAGACAGAGGG
140	140_Fanca5	TTCCATGAAGACGCGGCC
141	Luc.For	AACACGATGATAAGCATGGAAGATGCCAAAAACATTAAGAAGG
142	IRES.Rev	TTTGGCATCTTCCATGCTTATCATCGTGTTTTTCAAAGGAAAACCAC
143	rLuc.PA.HindIII.Rev	CCTTTTGCTCACATGAAGCTTGAACAAACGACCCAACACCCGTGCGTTTTATT CTGTCTTTTTATTGCCGTTATTGTTCATTTTTGAGAACTCGC
144	FAnca6.Rev	GGAAAGCAGACAACCAGGGC
145	Fanca7.Rev	CCACAGCATGCATGTCGG
146	Fanca_outwards.for	CCAGGAGCCTCATCTCTTCTGA
147	Fanca_outwards.rev	GGACCCACGAGTCGGACAT
148	FancaH1.Rev	CGCCGGTAGCGCTAGCACGTCCACACATGGTCCTCACGAAGAGG
149	FancaH2.For	GTGTGGACGTGCTAGTGCTCCCTGCAGT
150	SnaBI.LoxP.Ori.For	CATCAATGTATCTTATACGTAATAACTTCGTATAGCATACATTATACGAAGTT ATGAGCAAAAGGCCAGCAAAAGGCCAGGAACCG
151	AseI.LoxP.REV	CCAATATAAATGTACTATTAATATAACTTCGTATAATGTATGCTATACGAAGT TATTGTACATATTATGATATAGATACAACGTATGCAATG
152	SFFV.For	ATCATAATATTGTACAGCTAGCTGCAGTAACGCC
153	PA.dTom.Rev	ACGAAGTTATTGTACGAACAAACGACCCAACACCCGTGCGTTTTATTCTGTCT TTTTATTGCCGTTACTTGTACAGCTCGTCCATGCC
154	NheI.Fanca.For	TATATATATATATATATATAGCTAGCATGTCCGACTCGTGGGTC
155	NheI.Fanca.rev	TATATATATATATATATATAGCTAGCTCAGAAGAGATGAGGCTCCTGGG
156	156_IRES.GFPnls.For	CGCTAGCGCTACCGGTCGCCCCCCCCCC
157	157_IRES.GFPnls.Rev	ATCGAGCTCGAGATCTTCACACCTTCCTCTTCTTCTTGGG
158	158_AgeI.Fanca.For	CGCTAGCGCTACCGGATGTCCGACTCGTGGGTC
159	159_AgeI.Fanca.Rev	GGGGGGGGCGACCGGTCAGAAGAGATGAGGCTCCTGGG
160	GFPnls.Rev	CATCGAGCTCGAGATCTTCACACCTTCCTCTTCTTCTTGGGTGGCTTGTACAG CTCGTCC
161	161_Nsi.SFFV.for	CCCGGTCTCCTCCCAAGCTAGCTGCAGTAACGCC
162	162_NsiI.TK.dTom.Rev	TGGCCAATATTGACAGAACAAACGACCCAACACCCGTGCGTTTTATTCTGTCT TTTTATTGCCGTTACTTGTACAGCTCGTCCATGCCG
163	163_AseI.EI40.For	TACGAAGTTATATTAGATCAAGAAAGCACTCCGGGC

164	164_AseI.EI40.Rev	ATAAATGTACTATTAGATCTAATGTACATCATGAGGGCTATAGTTAATAAAAA TGTATTGTCT
165	CAG.For	TTGATTATTGACTAGTCTAGTTATTAATAGTAATCAATTACGGGGTCATTAGT TCA
166	CAG.Rev	CAATAAAGCTACTAGTGCCGCCGGTCACACGCCA
167	AseI.UCOE.FOR	TACGAAGTTATATTACAGCCTACAGCTCAAGCCAC
168	AseI.UCOE.Rev	ATAAATGTACTATTAAAAGGGAATAAGAATTCCCGCCTCCGCGCCCCACTTTC ACCCC
169	dTom.For	CATCACCTCCCACAACGAGG
170	BsrGI.EI40.For	TATCATAATATGTACGATCAAGAAAGCACTCCGGGCTCCAGAAG
171	BsrGi.EI40.Re	CCAATATAAATGTACGATCTAATGTACATCATGAGGGCTATAGTTAATAAAAA TGTATTG
172	AgeI.dTOm.For	CGCTAGCGCTACCGGTATGGTGAGCAAGGGCGAGGA
173	EcoRV.dTom.Rev	GGGGGGGGGGCGGATATCTTACTTGTACAGCTCGTCCATGCCGTACAGG
174	AgeI.Kat.For	CGCTAGCGCTACCGGTATGGTGGGTGAGGATAGCGT
175	EcoRV.Kat.Rev	GGGGGGGGGGCGGATATCTCAGCTGTGCCCCAGTTTGCT
176	Fanca.outwards2.Rev	GGATTATATTTTTCCCTCTTGACCCTTCC
177	eGFP.outwards.for	GATCACTCTCGGCATGGACG
178	eGFP.outwardsII.For	CCACTACCAGCAGAACACCCCC
179	NotI.Fanca.For	CGCTAGCGCTACCGGGCGGCCGCTCGAG
180	NheI.eGFP.nls.Rev	ATCGAGCTCGAGATCGCTAGCTTGCCAAACCTACAGGT
181	SpeI.CAG.For	GACCGGCGGCTCTAGACTAGTAGCTTTATTGCGGTAGTTTATCAC
182	SPeI.CAG.Rev	TATTAATAACTAGCGACTAGTAACATGGCGGTCATATTGGACATG
183	U6.shRNAP53For	AATGTATCTTACATGGATCCGACGCCGCCA
184	U6.shRNAP53.Rev	CCTTTTGCTCACATGCCCCGGGCTGCAGGA
185	AgeI.hOCT34.For	CGCTAGCGCTACCGGTGAATTCGCCCTTCACCATGGC
186	XhoI.hOCT34.Rev	TATCATCGAGCTCGAGTCATATGACTAGTCCCCGAAGCTTGAATTCG
187	AgeI.hSox2.For	CGCTAGCGCTACCGGTGAATTCGCCCTTCACCATGTACAAC
188	XhoI.hKlf4.REV	TATCATCGAGCTCGAGTTAAAAATGTCTCTTCATGTGTAAGGCGAGGT
189	AgeI.hLMyc.For	CGCTAGCGCTACCGGTGAATTCGCCCTTCACCATGGAC
190	XhoI.hLin28.Rev	TATCATCGAGCTCGAGTCAATTCTGTGCCTCCGGGAGCA
191	Fanca-ires-GFPnls.For	GACCGGCGGCTCTAGCTAGTAGCTTTATTGCGGTAGTTTATCACAGT
192	Fanca-IRES_tGFPnls.Rev	TATTAATAACTAGCGAACATGGCGGTCATATTGGACATG
193	Lenti18.for	ATCTCGACGGTATCGATGGCATTGATTATTGACTAGTTATTAATAGTAATCAA TTACGG
194	Lenti18.rev	TCGGAATTCCCTCGAGGTCAGGCACCGGGCTTGCG
195	cPPT	GTGCAGGGGAAAGAATAGTAG
196	SFFV.For	TTGATTATTGACTAGTGTAACGCCATTTTGCAAGGCATG
197	vLuc.Rev	TATCATCGAGCTCGAGTCACTTGCACTCGTCGGGC
198	AgeI.NFKB.For	ATTGACTAGTTATTAATACAAGTTTGTACAAAAAAGCAGGCT
199	BglII.eGFP.rev	ATCGAGCTCGAGATCTTTACTTGTACAGCTCGTCCATGCC
200	attR.For	TTGATTATTGACTAGTATCAAACAAGTTTGTACAAAAAAGCTGAACGAGAA
201	EBNA_OKSM.For	TGGCAGTACATCTACAGCTAGCTGCAGTAACGCC
202	EBNA_OKSM.Rev	AGAGGGAAAAAGATCTCAGGCACCGGGCTTGCG
203	SnaBI.OKSM.For	tggcagtacatctacgtaagctagctgcagtaacgcc
204	SnaBI.OKSM.Rev	agagggaaaaagatctttacttgtacagctcgtccatgcc
205	BglII.Puro.For	gtacaagtaaagatcgccaccaacttcagcct
206	BglII.Puro.Rev	agagggaaaaagatcttcaggcaccggg
207	BsrGI.GFP.For	caaacaagtttgtacaATGCCCGCCATGAAGATCG
208	BglII.P62.For	GGTACCGAGGAGATCTATGGCGTCGTTCACG
209	MluI.P62.Rev	GCGGCCGCGTACGCGTCAATGGTGGAGGGTGCTTCGAATACTGG
210	BglII.S024.For	GGTACCGAGGAGATCTATGGCGTCGTTCACGGTGaagGCCTATCTTCTGGGCA AGGAGGAGGCGACCCGCGAGATCCGCCGCTTCGCTTTCTGCTTCAGCCCGGA
211	XbaI.W340.Rev	CTGTGAGGGGTCTAGAGAGCTTGGCCCTTCCGATTCTGGCATCTGTAGAGACT GGAGTTCACCTGTAGaTGGGTCCACTTCTTTTGAAGACAAATGTGTagcGTCA TCGTCTC
212	XbaI.S351.Rev	CTGTGAGGGGTCTAGAGAGCTTGGCCCTTCCGATTCTGGCATCTGTAGAGACT GGAGTTCACCTGTAGcTGGGTCCACTTC
213	seq.PA.For	GCTTATAATGGTTACAAATAAAGCAATAGC

214	seq.M13.For	GTTGTAAAACGACGGCCAGT
215	Seq.M13.Rev	TCTGTTCAGGAAACAGCTATGAC
216	seq.P62.For	AGCTGAAACATGGACACTTT
217	seq.P62.Rev	TTACTCTTGTCTTCTGTGCC
218	seq.T7.For	GAATTTTGTAATACGACTCACTATAGGG
219	EcoRI.GFP-2A-Puro.For	ttttggcaaagaattcATGCCCGCCATGAAGATCG
220	EcoRI.GFP-2A-Puro.Rev	cccgaagcttgaattctcaggcaccgggcttgc
221	KpnI.hP62.For	AGTCGACTGGATCCGGTACCATGGCGTCGCTC
222	EcoRI.hP62.Rev	GTGCGGCCGCGAATTCTCACAACGGCGG
223	K7A.For	AGTCGACTGGATCCGGTACCATGGCGTCGCTCACCGTGgcgGCCTACCTTCTGGG
224	ClaI.T269A.For	TTGAAGTTGATATCGATGTGGAGCACGGAGGGAAAAGAAGCCGCCTGgcgCCCGTCTCTCC
225	S403.Rev	TCAGAGAAGCCCATGGcCAGCATCTGG
226	cPPt.for	TAATAGCAACAGACATAC
227	mP62.subclone.for	ATCCGGTACCGAATTATGGCGTCGTTCACGGTGAAGGC
228	mp62.subclone.rev	GTGCGGCCGCGAATTCAGGAAACAGCTATGACCGCG
229	seq.hP62.rev	CGCTACACAAGTCGTAGTCTGGG

3.11.2 q(RT)-PCR primers

Table 25: (q)RT-PCR primers used in this study

#	Oligo name	Sequence (5'->3')	Efficiency	Binds to hOKSM (exogenous)	Binds to mOKSM (endogenous)
Q7	hOct3/4.For	CAGCATCGAGAACAGAGTGC	2,03	YES	NO
Q8	hOct3/4.Rev	ACACTCTCACGACGTCCTTT			
Q17	hKLF4.For	AGAGAGAAAACACTGCGGCA	2,03	YES	NO
Q18	hKLF4.Rev	CAGCGGCCAGATCATAAGGT			
Q29	hSOX2.For	AAAGAGCACCCCGACTACAA	2,04	YES	NO
Q30	hSOX2.Rev	ACTCCGCTGGCCATAGAATT			
Q35	hMYC.For	CTGTTGAAGCTGGCTGGAGA	2,06	YES	NO
Q36	hMYC.Rev	TAACGTTGAGGGGCATGGAG			
	hGAPDH.For	gctgcattcgccctctta	2,00	-	-
	hGAPDH.Rev	gaggctcctccagaatatgtga			
Q51	Puromycin1.For	ATCGGCAAGGTGTGGGTC	1,97	-	-
Q52	Puromycin2.Rev	CTCAACTCGGCCATGCGC			
Q53	mOCT3/4.For	TAGGTGAGCCGTCTTTCCAC	2,11	A bit	YES
Q54	mOCT3/4.Rev	GCTTAGCCAGGTTCGAGGAT			
Q55	mKLF4.For	AACATGCCCGGACTTACAAA	2,20	A bit	YES
Q56	mKLF4.Rev	TTCAAGGGAATCCTGGTCTTC			
Q57	mSOX2.For	GAAACGACAGCTGCGGAAA	2,23	A bit	YES
Q58	mSOX2.Rev	TCTAGTCGGCATCACGGTTTT			
Q59	mMYC.For	TAACTCGAGGAGGAGCTGGA	2,21	A bit	NO
Q60	mMYC.Rev	GCCAAGGTTGTGAGGTTAGG			
Q61	mNANOG.For	TTGCTTACAAGGGTCTGCTACT	2,11	NO	YES
Q62	mNANOG.Rev	ACTGGTAGAAGAATCAGGGCT			
Q63	mGAPDH1.For	AGCTTGTCATCAACGGGAAG	2,18	NO	YES
Q64	mGAPDH2.Rev	TTTGATGTTAGTGGGGTCTCG			
Q65	Puromycin3.For	TGCAAGAACTCTTCCTCACG	2,03		
Q66	Puromycin4.Rev	CGATCTCGGCGAACACC			
Q67	GFP3.For	CTTCCTGCACGCCATCAACAACG	2,22		
Q68	GFP4.Rev	GATGATCTTGTCGGTGAAGATCACG			
Q69	mGAPDH3.For	GTTGTCTCCTGCGACTTCA	1,93		
Q70	mGAPDH4.Rev	GGTGGTCCAGGGTTTCTTA			

3.11.3 Genotyping primers

Table 26: Genotyping primers used in this study

#	Oligo name	Sequence (5'->3')	Tm (°C)	Amplicon (bp)
G5	IMR_0015 *	CAAATGTTGCTTGTCTGGTG	53.2	150
G6	IMR_0016 *	GTCAGTCGAGTGCACAGTTT		
G7	Luc.For	GGACTTGGACACCGGTAAGA	59.9	439
G8	Luc.Rev	GTCCACGAACACAACACCAC		
GFP1	GFP1	GCCGCATGACCAACAAGATG	69,3	515
GFP2	GFP2	GTTGCTGTGCAGCTCCTCCA		
IC1	Sox21	AGCCCTTGGGGASTTGAATTGCTG	72,2	237
IC2	Sox21	GCACTCCAGAGGACAGCRGTGTCAATA		
G9	eGFP.For	GCAA GGGC GAGG AGCT GTTC ACC	72°C	329 bp
G10	eGFP.Rev	GGCG AGCT GCAC GCTG CCGT CGTC		

*Chr14: TCRD Region 54142899-54143104

3.12 Materials

Table 27: Materials used in this study

Material	Description	Company
Ampuwa water	distilled water for injection	Fresenius Kabi Deutschland GmbH
Bacterial spreaders	individually wrapped, sterile	VWR
Cell culture dishes	Ø 6 / 10 / 15 cm	greiner bio-one
Cell culture flasks	25/ 75 / 175 cm^2	greiner bio-one
Cell culture multidishes	6 / 12 /24/ 96 well	greiner bio-one / nunc
Cell scraper	28 cm length	greiner bio-one
Combitips	5 ml	Eppendorf
Cuvettes gene pulser	0,1 cm gap, package of 50	Bio-Rad
EDTA-Tubes	EDTA-tubes for blood collection	BD Bioscience
FACS tubes	12x7.5 ml BD FACS tubes	BD Bioscience
Falcon tubes	15 and 50 ml	Falcon
Filter Unit	0,22 / 0,45 µm	GE Healthcare/Whatman
Filtered tips	10, 20, 100, 200 and 1000 µl	Nerbe
Hybond XL nylon membrane	Nylon membrane for Southern Blot	(Amersham Biosciences
iBlot 2 Transfer Stacks, PVDF	Each iBlot® 2 Transfer Stack contains a copper coated electrode and appropriate cathode and anode buffers in the gel matrix to allow fast, reliable transfer of proteins.Pore size 0,2 um	Life Technologies
LightCycler 480 Multiwell Plate 96	white, half-skirted, white polypropylene 96 well plates supplied with a bar-code label. 96-well for reaction volumes from 10 to 50 µl.	Roche
LUNA cell counting slides	disposable precision slides for Trypan blue and fluorescence cell counting	Logos Biosystems
Lysing Matrix D	2 ml tube with grinding beads	MP biomedicals
Membrane VMWPO2500	composed cellulose membrane with 0.025 um pore size for DNA and protein dialysis	MERCK Millipore
Millex-GV	0,22 um Filter unit	MERCK Millipore
Mini Protean Gels 4-20%	4–20% precast polyacrylamide gel, 8.6 × 6.7 cm (W × L), for use with Mini-PROTEAN Electrophoresis Cells. Separation of polypeptides ~2–400 kDa. Can be used for both standard denaturing protein separations as well as native electrophoresis.	Bio-Rad
Mini Protean tetra Cell	mini gel electrophoresis system	Bio-Rad

Multipette M4	multistep pipette with a range from 1 µl to 10 ml	Eppendorf
Nalgene cryogenic vials	25 sterile tubes	ThermoFischer Scientific/Nalgene
Needles	0.9/0.8 x 40 mm and 0.5 x 25 mm	BD (Franklin Lakes, US)
Nitrocellulose membrane	PROTAN®	Whatman (Maidstone, UK)
Nylon membrane	Amersham Hybond^{TM}-N^{+}	GE Healthcare (München, Germany)
Nylon mesh	mesh size 40-70uM	Nitex
Pasteur pipettes	9 inch, 5.5 mm, Disposable, Bulk Pack, Non-Sterile, Unplugged	Corning Incorporated
PCR strips		Neolab
PCR tubes	0,2 ml thin-walled tubes with flat caps	ThermoFischer Scientific
Plastic pipettes (sterile)	5, 10, 25 and 50 ml	Corning/Costar
PVDF membrane	0,45 um, 26.5 cm x 3.75 m roll	ThermoFischer Scientific
Reagent reservoir	50ml deposits, sterile, polystyrene	Corning Incorporated
Safelock tubes	1,5 / 2,0 / 5,0 ml	Eppendorf
Scalpels	Disposable stainless steel blade with plastic handle	feather
Stericup filtering unit	HV, 0.45 µm, 500 ml	MERCK Millipore
superFROST plus slides	25 x 75 x 1,0 mm	ThermoFischer Scientific
Syringes	5 and 50 ml	Terumo
Syringes	1/ 3 / 5 / 10 / 50 ml	BD Bioscience
Transferpette multichannel	200 ul	Brand
vacushield filters	Hydrophobic, microporous PTFE membrane filter achieves 99.97% D.O.P. retention (0.3 µm particles in air). Features polypropylene housing and 6.4–12.7 mm (1/4–1/2") stepped hose barb connections (internal taper accepts male Luer fitting). Dimensions: 7.3 dia.x8.2L cm (27/8x31/4").	VWR
Whatman paper	3 mm	Whatman
µ-Plate 24 Well	µ-Plate 24 Well ibiTreat: #1.5 polymer coverslip, tissue culture treated, sterilised, black	iBidi
µ-Plate 96 Well	µ-Plate 96 Well ibiTreat: #1.5 polymer coverslip, tissue culture treated, sterilised, black	iBidi

3.13 Equipment

Table 28: Equipment used in this study

Device	Company
Microscope EVOS™ XL Core Cell Imaging System	Thermo Fischer Scientific
Gamma Cell 1000 irradiator	Gamma cell
Chemiluminiscense Imaging - Fusion SL	Analis
Bacterial Shaker Centromat SII	B. Braun Biotech International
BD LSRFortessa	BD Biosciences
BD FACSCalibur cell analyzer	BD biosciences
Luminometer Mithras LB943 multimode reader	Berthold Technologies
Electroporator MicroPulser	Bio-Rad
Gel basic power supply PowerPac™	Bio-Rad
Mini-PROTEAN® Tetra Vertical Electrophoresis Cell	Bio-Rad
IVIS lumina III	Caliper life sciences
-20˚C freezer	Comfort
Centrifuge Ependorff 5430R	Eppendorf
Centrifuge Eppendorf 5424R	Eppendorf
Centrifuge Eppendorf 5810	Eppendorf
Eppendorf® Thermomixer® R	Eppendorf
Pipettes 2, 10, 20, 200, 1000ul	Gilson
Table top spinner PCV_240	Grant-bio
Incubator In-Vitrocell 200L	ibs-tecnomara
MS2 minishaker (vortex)	IKA
Integra Vacusafe	Integra
Pipette Boy	Integra
iBlot 2	Invitrogen
Water bath	Julabo
Microscope Keyence BZ-X700 Fluorescent	Keyence
UV transluminator N90 LW366	Konrad Benda
PCR Workstation (hood)	Labcaire
+4˚C fridge	Liebherr medtime
LUNA automated Cell counter	Logos Biosystems
Amaxa 4D Nucleofector / Amaxa II Nucleofector	Lonza
Gel printer Mitsubishi P93D	Mitsubishi
Tube roller NeoLab	NeoLab
-80˚C freezer U725 innova	New Brunswick (Eppendorf),
Camera Nikon DS-Qi2	Nikon
Microscope Nikon Eclipse Ti/X-Cite120Led	Nikon
Microscope phase contrast Olympus CK40	Olympus
Packer bag sealer	Packer
Gel chambers PerfectBlue Wide Gel System ExM	Peqlab
precellys 24 lysis and homogenisation	peqlab
QIAcube robot	QIAgen
Incubator Stem Cells	Queue
Real-Time PCR System LightCycler® 96	Roche
Microwave oven	Siemens
PersonalHyb oven	Stratagene
Liquid Nitrogen cryostorage system K Series	Tec-Lab
Bacterial hood, class II type A/B3	The Baker company
Cell culture hood Safe2020	ThermoFischer scientific
Centrifuge rotor F10-6x-500y FiberLite rotor	ThermoFischer scientific
Centrifuge Sorvall RC6+	ThermoFischer scientific
Nanodrop 2000C	ThermoFischer scientific
PCR thermocycler peqSTAR 2X/96X Universal Gradient	VWR Peqlab
XGI-8 Gas anesthetic chamber	xenogen

3.14 Software

Table 29: Software used in this study

Name	Description	Origin
Ascent Software (Version 2.6)	ELISA and BCA analysis software to measure absorbance and determine concentrations	
Affinity designer	Graphic design package (similar to Adobe Photoshop)	Pantone
Fiji	Image processing package. Distribution of ImageJ, which bundles Java, Java3D and allows addition of new plugins and macros.	[288]
FlowJo 10	Platform for single-cell flow cytometry analysis	FlowJo, LLC
Primer 3	Primer design for PCR, qPCR and qRT-PCR	[289], [290]
SnapGene	Molecular biology tool for planning, visualisation and documentation of DNA cloning and PCR	GSL Biotech LLC
Papers 3.4.6	Referencing software	Labtiva, Inc
Graphic	Graphic design for Mac	Autodesk, Inc.
NIS- Elements version 4.40	Nikon Microscope imaging software	Nikon
BD CellQuest Pro	FACS software acquisition and analysis	BD Biosciences
FUSION-CAP Software	Chemiluminescence Imaging software for image acquisition and analysis	Analis
LightCycler® 96 Real-Time PCR System	Allows absolute and relative quantifications, Tm calling and Endpoint genotyping	Roche

4. METHODS

4.1 Cell culture methods

4.1.1 Standard cell culture

The cells used in this project **(Section 3.1)** were incubated under standard conditions (37°C and 5% CO_2). The media used for each cell type is described in **Section 3.3.3.**

Unless specifically required, cells were passaged 2 or 3 times per week (on demand). For that, the cells were first washed once with 1x PBS and subsequently trypsinised with 0.25% Trypsin/EDTA for 5 minutes at 37°C. Afterwards, a gentle rocking was applied to ensure complete detachment from the plate. The trypsin was inactivated with the same amount of FCS-containing media (specific for each cell type), and the detached cells were pipetted up and down to achieve single cell suspensions. Finally, single cells were re-plated at the desired density in fresh medium. Because mouse Stem Cells (mESC and iPSC), as well as human Stem Cells (hiPSC), were co-cultured with feeder-layers, they required more complex culturing techniques. The preparation of gelatinised plates, feeder-layers and the Stem Cell culture are explained below.

4.1.2 Gelatine coating of cell culture plates

The desired plate format was pre-coated with 0,1% Gelatin in dH_2O (from a 1% Gelatin stock) and incubated for 20-30min at 37°C. The gelatin was aspirated afterwards. At this point, the plates were ready to use or could be wrapped and stored at 4°C for 1-2 weeks.

4.1.3 Preparation of feeder layers

Immortalised Mouse Embryonic Fibroblasts (iMEFs) were used as feeder-layers for co-culture of both mouse and human Stem Cells. The cells were cultured as described above and grown until 90% confluency. One of the two methods described below were used to inactivate and prevent the cells from undergoing cell division:

1) **Mitomycin C treatment.** Without removing the media, Mitomycin C (Stock 100 µg/ml, final concentration 10 µg/ml) was added to the plate, and the cells were incubated for 2h at 37°C and 5% CO_2. After the incubation period, the cells were washed with PBS and trypsinised as described.

2) **γ-irradiation.** The cells were washed in PBS, trypsinised and resuspended in FCS-containing media. The single-cell suspension was transferred into a 15 ml tube, and the cells were exposed to 60 Gy of γ-irradiation (Gamma cell 1000).

Once the cells were inactivated, they were counted, and a total amount of $1{,}25x10^6$ cells were seeded per plate (regardless of the format). For each plate format, the total amount of cells was divided by the number of wells.

4.1.4 Stem Cell culture

4.1.4.1 mESC/miPSC

Murine stem cells were grown on feeder layers (in gelatin-coated plates) and in mouse Stem Cell media containing LIF and differentiation inhibitors (2i) **(Section 3.3.3).** When the stem cells reached 70-80% of confluency (the colonies were dense, but their limits did not touch with each other after 4 days of passaging), they were washed twice with PBS and trypsinised as usual. After trypsinisation, the cells were resuspended in IMDM without LIF or inhibitors. At this point, both stem cells and feeder layers were mixed in suspension. To separate them, a differential sedimentation step was performed, which relies on the separation of the cells by their density. Fibroblasts, which are bigger and denser than stem cells, will sediment and attach faster to the plate. After 20-30 min incubation time at 37°C and 5% CO_2, the supernatant containing the stem cells was collected, counted and re-seeded to the desired cell format, in Stem Cell media with LIF and 2i.

4.1.4.2 hESC/hiPSC

Human stem cells were also grown on feeder layers, and gelatin-coated plates in hSC media containing FGF2 and plasmocin **(Section 3.3.3.1)** in a prophylactic concentration against mycoplasma contaminations. The media was replaced every second day. When hSC reached 70-80% confluency (which corresponded approximately to one week after passaging), they were picked and passaged as described below in **Section 4.1.12.2.**

4.1.5 iPSC Reprogramming using defined Transcription Factors

4.1.5.1 miPSC Reprogramming (OKSM)

To reprogram murine iPSC, early passage fibroblasts (not later than passage 5) were transfected or transduced as described in **Sections 4.1.10.1** and **4.1.14.2**, with either the 4in1 Lentivirus (vector 27)

or the S/MAR versions POP (vector 35) and nanoPOP (vector 84), all containing the human codon-optimised transcription factors: Oct4, Sox2, Klf4 and cMyc. Right after nucleofection, the cells were placed in DMEM+10% FCS and devoid of antibiotics or selection, to allow the cells to recover. Two days and four days after transfection and transduction, the media was replaced with MEF media with antibiotics and selective pressure (if needed). At day six, the media was replaced by mESC media (with optional Vitamin C and VPA supplements) and changed every second day.

At day 8, the cells were washed, trypsinised and expanded into a bigger plate format (24 well to a 6 well or 6 well to a 10cm plate), which was coated with gelatin and contained feeder layers, and the media was replaced by supplemented mESC media. The media was changed every second day, and the cell morphology was also monitored every second day. By day 16, the first compact dome-shaped colonies emerged. At that point, the colonies were picked as explained in **Section 4.1.12.2.**

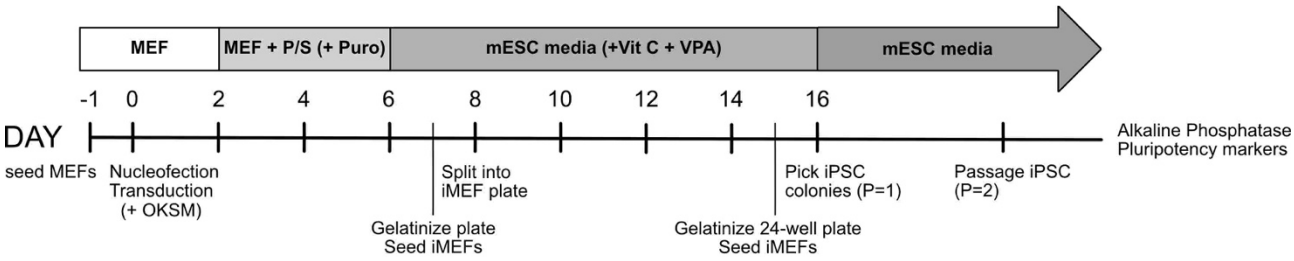


Figure 10: Murine iPSC reprogramming timeline

4.1.5.2 hiPSC Reprogramming (OKSML/shP53)

Human Dermal Fibroblasts (HDF) at early passages (p3-5) were transfected as described in **Section 4.1.10,** using either EBNA-based vectors (vectors 77-80) or S/MAR-based (vectors 73-76) containing the human transcriptions factors: Oct4, Klf4, Sox2, Lin28 and N-Myc, together with shRNA for p53.

After transfection, the cells were plated into a well of a 6-well plate in DMEM media without antibiotics (DMEM, 10%FCS, 1% L-Glut, 1% NEAA) in order to allow the cells to recover. Two days after, the media was replaced by DMEM with antibiotic and Puromycin selection (if required). At this point, the cells were split into a T25 flask, on demand, and DMEM media was changed every second day. At day seven, a 6-well plate was gelatinised and feeder layers were seeded at a confluency of $1{,}25 \times 10^6$ cells/plate. The next day, the cells were transferred to the feeder plate (30.000 cells/well of a 6-well plate or 15.000/well of a 12-well plate). The leftover cells were lysed and kept for other purposes, such as Western Blot or gDNA extraction. The fibroblast media was replaced by hESC media containing FGF2, and the cells were fed every second day.

The cell morphology was monitored using an EVOS™ XL Core Cell Imaging System (Thermo Fischer Scientific). Human iPSC colonies started to appear at day 17-21 and could be picked once the colony size was big enough (more than 100 cells) as described in **Section 4.1.12.2.**

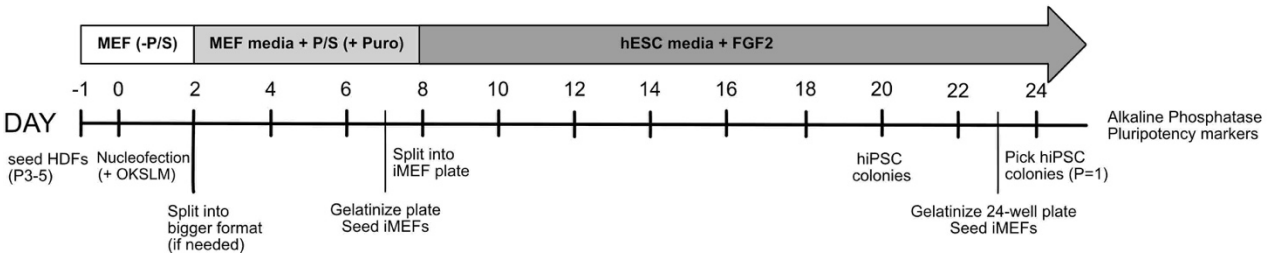


Figure 11: Human iPSC reprogramming timeline

4.1.6 Stem Cell random differentiation via Embryonic Bodies (EBs)

4.1.6.1 Murine EB formation

To differentiate mESC and miPSC into the three embryonic germ layers (ectoderm, mesoderm, endoderm), the cells were forced to form Embryonic Bodies, via formation of hanging drops as depicted in **Figure 12.** To achieve this, the cells were passaged as described above and diluted in EB medium, devoid of LIF and inhibitors, to a density of $3x10^4$ cells/ml (or 600 cells per 20µl). With the help of a multichannel pipette, 80-100 drops of 20µl were placed on the lid of a bacterial petri dish, and the dish was filled with PBS to preserve the humidity. The lid was closed quickly but carefully (to avoid collapsing of drops), and the plate was incubated at 37°C and 5%CO_2 for 3 days. During this time, the cells collapsed in the bottom of the drop and formed undifferentiated aggregates. After 3 days, the drops from one lid were harvested and pooled with 4ml of EB medium, and the aggregates were cultured in suspension in EB media for additional 3 days. After, the aggregates were transferred into a 15ml conical tube with EB media (at this point, the aggregates were visible to the naked eye) and sedimented at room temperature for about 10 minutes. They were resuspended in fresh EB media and 1-2 aggregates were transferred per well of a gelatin-coated µ-plate 24 well (iBidi). The plate was incubated for 1-2 weeks and the media was replaced by fresh EB media every second day.

At this point, the cell aggregates started adhering and spreading on the well, while slowly differentiating. The embryonic Bodies were checked regularly for differentiation and when obvious differentiated structures formed (*e.g.*: beating cardiomyocytes, neurons…), they were fixed and stained with antibodies against markers of the 3 germ layers. Protocol explained in **Section 4.3.3.3.**

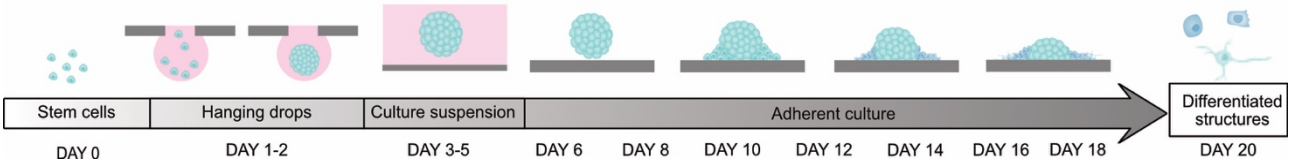


Figure 12: Random differentiation of mESC via Embryonic Bodies using hanging drops
Stem cells were forced to collapse during 3 days into drops forming undifferentiated aggregates. Then, the collected drops were transferred into adherent culture (gelatin-coated plates) and allowed to attach, spread and differentiate in the absence of LIF and differentiation inhibitors (2i). After approximately 10 days of adherent culture, differentiating structures became apparent.

4.1.6.2 Human EB formation

The hiPSC were grown on feeder cells in gelatinised plates and hiPSC media + FGF2 and were passaged two days before EB formation. To generate EBs, hiPSC colonies were scratched with a pipette tip, and the cell aggregates were transferred into an ultra-low attachment plate (Corning) and grown in hiPSC + FGF2 + 10 μM Y-27632 (Sigma). The media was changed every second day to hiPSC+FGF2. At day 4, the media was replaced by DMEM + 20% FCS and changed every second day. At day 7, the EBs were plated on black glass-bottom tissue culture treated plates (iBidi) coated with 0.1% porcine gelatin (Sigma), without feeders. The media Change media every second day until day 15 when differentiated structures appeared. At that point, the cells were fixed, permeabilised and stained as described in **Section 4.3.3.3.**

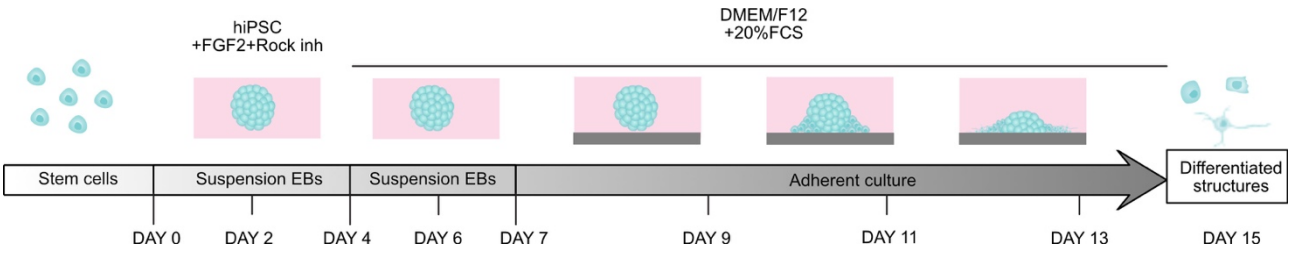


Figure 13: Random differentiation of human hiPSC via Embryonic Bodies
hiPSC were scratched to form undifferentiated aggregates, which were kept in suspension in ultra-low attachment plate and hiPSC media. At day 7, the aggregates were transferred into adherent culture (gelatin-coated plates) and allowed to attach, spread and differentiate in the absence of differentiation inhibitors. After approximately 7 days of adherent culture, differentiating structures became apparent.

4.1.7 Stem Cell directed Hematopoietic differentiation via Cytokines

Directed differentiation experiments were performed by Marleen Büchler, from Dr Michael Milsom's lab (Hi-STEM, DKFZ).

Mouse ESC were grown to confluency, washed twice in PBS, trypsinised and resuspended into single cells in EB-HSC differentiation medium. The mESC were separated from feeder layers via differential

sedimentation. Then, 75.000 cells were counted and plated in Ultra low attachment T25 flasks containing EB-HSC medium and incubated for 2,5 days (60h) under hypoxic conditions (5% CO_2 and 5% O_2). Another 500.000 cells were stained with a panel of hematopoietic differentiation antibodies and used for FACS analysis at day 0.

After 60h, the media was replaced by EB-HSC media containing 5 ng/ml of cytokines (BMP-4, Activin A, VEGF and FGF2) and incubated for 60h in hypoxia.

At day 5, the Embryonic Bodies were allowed to settle at the bottom of the flask. Then, the majority of the media containing cytokines was removed and centrifuged to pellet cell debris. Fresh EB media without cytokines was added to the Embryonic Bodies, and the centrifuged cytokine-containing media (conditioned media) was added back to the flask, which was then incubated for another 60h in hypoxic conditions.

At day 6, the EBs were collected, washed with PBS and let to be settled by gravity. Then, they were dissociated in 250 µl Enzyme-free dissociation mix in 1 ml of PBS and incubated at 37°C for 20min with occasional swirling of the tube. Then, another 8ml of enzyme-free dissociation media was added to the cells, which were then incubated for 5 minutes at room temperature. Finally, the cells were mechanically dissociated and washed with PBS, collected by centrifugation (1500rpm for 5min at 4°C) and used for downstream analysis such as FACS.

The hematopoietic differentiation panel of antibodies used, included: CD41 (1:300), CD144/VE-Cadherin (1:200) and CD117/c-Kit (1:2000).

4.1.8 Neuronal differentiation using ATRA

Neuroblastoma (Be2C) cells were forced to differentiate into neurons upon addition of 10 µM All-Trans Retinoic Acid (ATRA).

For this, 25.000 cells were counted and plated per well of a 24 well plate. Each cell line was plated in eight replicates, four of which were treated with ATRA and the other four were treated with DMSO (untreated control).

The day after, the media was replaced by RPMI media (10% FCS, 1% P/S, 1% NEAA) containing 10 µM ATRA (diluted in DMSO). The same volume of DMSO was used as 'untreated control'.

The media was changed every second day, and the cells' morphology as well as GFP expression (if any) were also monitored every second day using either a Keyence BZ-X700 or a Nikon Eclipse Ti/X-Cite120Led fluorescent microscopes.

At day 10, immunostainings for neuronal markers were performed according to **Section 4.3.3.3.** Also at that point, ATRA was withdrawn from half of the wells (2 wells per cell line), and the other half was kept under ATRA treatment for another 10 days.

4.1.9 Chemical Transfection

Several chemical transfections were performed according to the manufacturer's instructions (Refer to **Section 3.3.4.1** for more information). The most common chemical transfection protocols used in this study were using jetPEI and Effectene.

Note that for vector comparisons, the equal amounts of vector molecules (molecular ratio) were added in each transfection.

4.1.9.1 Transfection with PEI

Transfection with PEI was most commonly performed in Glioblastoma, HEK293T and HeLa cells. The cells were grown to confluency, washed with PBS, trypsinised and counted. The number of cells to be transfected varied depending on the plate format and the reagent used. **(refer to Table 30).** The day after, the cells achieved 60-80% confluency and were ready to be transfected. The amount of DNA and reagent varied depending on the plate format and the cell line **(Table 31).**

Most commonly, 2-10 µg of plasmid DNA were diluted in NaCl buffer, and the jetPEI reagent was also diluted in NaCl buffer. Both mixes were vortexed and spun. Then, the PEI mix was added to the DNA mix (in this order), and the solution was immediately vortexed and spun, followed by a 30 minutes incubation at room temperature to allow the formation of PEI-DNA complexes.

Finally, the complexes were added to the cell's media in a drop-wise manner, followed by a gentle rocking of the plate. The plate was brought back to the incubator, and the media was replaced by fresh media with 1 µg/ml Puromycin selection the day after.

Table 30: Recommended format and number of cells to seed the day before transfection

Culture vessel	Number of adherent cells to seed	Surface area per well (cm^2)	Volume of medium per well or plate (ml)
384-well	5 000 - 10 000	0.075	0.05 - 0.1
96-well	10 000 - 17 000	0.3	0.1 - 0.2
48-well	25 000 - 50 000	1	0.25 - 0.5
24-well	50 000 - 100 000	1.9	0.5 - 1
12-well	80 000 - 200 000	3.8	1 - 2
6-well/35 mm	200 000 - 400 000	9.4	2 - 4
6 cm/flask 25 cm^2	400 000 - 800 000	28	5 - 10
10 cm/flask 75 cm^2	2 000 000 - 4 000 000	78.5	10 - 15
14 cm/flask 175 cm^2	4×10^6 - 8×10^6	153 - 175	20 -30

Table 31: Complex preparation for transfection in different cell culture formats

Culture vessel	Amount of DNA (µg)	Volume of jetPEI® reagent (µl)	Volume of NaCl solution for both DNA and jetPEI® (µl)	Total volume of complexes added per well (µl)
384-well	0.1	0.2	5	10
96-well	0.25	0.5	10	20
48-well	0.5	1	25	50
24-well	1	2	50	100
12-well	2	4	50	100
6-well/35 mm	3	6	100	200
6 cm/flask 25 cm^2	5	10	250	500
10 cm/flask 75 cm^2	10 - 20	20 - 40	250	500
14 cm/flask 175 cm^2	20 - 30	40 - 60	500	1000

These tables were taken from 'jetPEI in vitro DNA transfection' manufacturer's Protocol (Polyplus)

4.1.9.2 Transfection with Effectene

Effectene was mainly used to transfect Neuroblastoma cells, according to an optimised protocol from Dr Jeannine Lacroix (DKFZ).

The cells were grown to confluency, washed with PBS, trypsinised and counted. Effectene transfections were performed in 6-well plates, and $2x10^5$ cells were used per reaction. 1 µg of plasmid DNA was diluted in 100 µl of EC-buffer, containing 2 µl of Enhancer. The mix was vortexed, spun and incubated 5 minutes at room temperature. Then, 7,5 µl of Effectene were added, and the mix was vortexed, spun and incubated 30 minutes at room temperature, to allow the formation of DNA complexes. Finally, the complexes were equilibrated with 400 µl of OptiMEM, and the complex-containing solution was delivered to the cells in a drop-wise manner. The cells were incubated overnight at 37°C, and 5% CO_2 and the media was replaced with fresh media containing 1 µg/ml Puromycin selection the day after.

4.1.10 Physical Transfection/Nucleofection

Nucleofection was used to transfect Fibroblasts (NHDF nucleofection kit) and Stem Cells (Mouse ES Cell Nucleofector kit) using the Amaxa II electroporation device and following the manufacturer's instructions. Note that for vector comparisons, the equal amounts of vector molecules (molecular ratio) were added in each transfection.

4.1.10.1 Nucleofection of Fibroblasts (MEFs and HDFs)

The desired well-plate format (usually 6-well plate) was gelatinised in advance. When the fibroblasts (either MEFs od HDFs) reached 80-90% confluency, they were washed, trypsinised and counted. Unless otherwise stated, 500.000 cells were used per nucleofection. Afterwards, such number of cells were aliquoted into a 1.5ml tube and centrifuged at 200 g for 10 minutes at room temperature. Then, the cell pellet was washed with PBS and centrifuged again. Meanwhile, 2-10 µg of plasmid DNA were diluted in 100 µl of supplemented solution (82 µl NHDF Solution and 12 µl supplement) for MEFs and 110 µl (90 µl NHDF Solution and 20 µl supplement) for HDFs. This DNA-containing mix was used to resuspend the cell pellet, which was quickly transferred into a nucleofection cuvette and nucleofected using program U-020 (MEFs) or P-022 (HDFs) and the Amaxa Nucleofector II (Lonza). Finally, the nucleofected cells were gently pipetted and transferred into a gelatine-coated 6-well plate with DMEM containing 10% FCS without selection nor antibiotics, to allow the cells to recover from

the nucleofection procedure. Two days after, the media the media was replaced by DMEM with 10% FCS, containing antibiotics and 0,5-1 µg/ml of Puromycin selection (if needed).

4.1.10.2 Nucleofection of mESC

A gelatin-coated 6-well plate with feeder layers was prepared the day before nucleofection. When the stem cells reached the desired confluency, they were washed, trypsinised and differentially sedimented as described above. Meanwhile, 2 - 10 µg of plasmid DNA (usually containing a selection marker) were diluted in 100 µl of supplemented solution (82 µl mESC Solution and 12 µl supplement).

After the feeder layers sedimented, the supernatant containing the stem cells was transferred into a new tube, and the stem cells were counted. Between $5x10^5$ - $2x10^6$ cells per reaction were aliquoted, washed with PBS and centrifuged at 200 g for 5 minutes. Then, the cell pellet was resuspended in the 100 µl DNA-containing solution, quickly transferred into a nucleofection cuvette and nucleofected using program A-013 (Amaxa Nucleofector II, Lonza). Finally, the nucleofected cells were gently pipetted and transferred into a gelatin-coated feeder containing 6 well plate with mESC media without selection, to allow the cells to recover from the nucleofection procedure. Two days after, the media was replaced by mESC media containing antibiotics and 500 ng/ml of Puromycin for selection.

4.1.11 Cell selection using antibiotics

The transfected cells were grown in their respective media, split at a low density and kept under selective pressure (1 mg/ml G418 or 0,5 - 1 µg/ml Puromycin) for approximately 2 weeks or until all the untransfected cells were dead, and all the remaining attached cells were positive for the vector's reporter gene (usually GFP or dTomato). Because of the low density passaging after transfection and the selective pressure, the cells grew to form more or less defined and separated colonies.

4.1.12 Clonal selection

4.1.12.1 Single cell-dilution

After selection, the cells were washed trypsinised as usual. Then, the cells were counted and diluted to 100.000 cells/ml while a 96 well-plate was labelled and filled with the appropriate media.

Afterwards, 1/10 serial dilutions were performed to obtain single cells in a well of a 96-well plate. The cells were kept under selection, allowed to grow until reaching a sub-confluent state and were then expanded to a bigger format.

4.1.12.2 Colony picking

4.1.12.2.1 Fibroblasts, HEK293T and HeLa cells

After selection and passaging, the remaining GFP positive colonies were observed under a fluorescent microscope (Keyence microscope). To facilitate the picking process, the plate was carefully marked, with the help of a permanent marker, around the region where the colonies of interest were.

Before starting with the picking procedure, a new plate with appropriate media was prepared. Also, Vaseline and 1000 µl pre-cut tips (cylinders corresponding to the wide diameter of the tip) were autoclaved. The cells were washed twice with PBS, and the colonies of interest were localised thanks to the labelling underneath. The cylinder was soaked in Vaseline and was positioned surrounding the desired colony. Trypsin was added to the cylinder, which was sealed thanks to the Vaseline. After trypsinisation, the colonies were resuspended and transferred to a new plate. After two days, the cells were checked for attachment, and the media was replaced every second day.

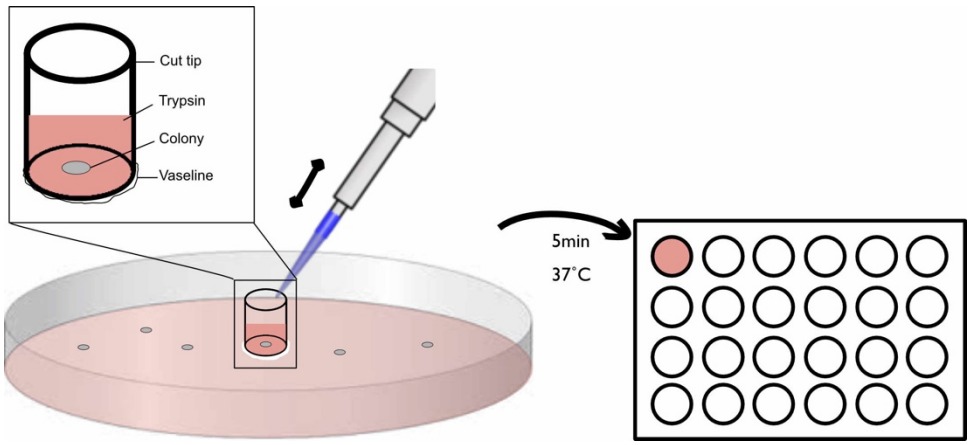


Figure 14: Non-Stem cells colony picking method
Pre-cut 1000ul tips and Vaseline were autoclaved and stored. A pre-selected colony was surrounded by the tip soaked in Vaseline, to prevent leakiness. Trypsin was added to the cylinder, and the cells were trypsinised for 5 minutes at 37°C. Then, the colony was mechanically disrupted by pipetting up and down and transferred into a 24 well-plate with appropriate media

4.1.12.2.2 Stem cells (mESC/miPSC)

The day before picking Stem Cell colonies, a gelatinised 24-well plate with feeder layers was prepared. After selection and passaging, the remaining GFP-positive colonies were observed under a fluorescent microscope (Keyence microscope). To facilitate the picking process, the plate was carefully marked, with the help of a permanent marker, around the region where the colonies of interest were. A stereoscopic microscope was carefully disinfected, placed under the cell culture laminar flow hood and sterilised under UV lamps, an hour before picking colonies.

Before starting the picking procedure, a 96 well-plate containing trypsin, 10 µl thin, sterile tips and some sterile surgical needles were prepared under the hood. The polyclonal SCs colonies, growing in 10 cm dishes or 6 well plates, were washed twice with PBS, which was then not completely removed to prevent the cells from drying. The plate was positioned under the stereomicroscope, and the pre-marked selected colonies were identified. 10 µl of trypsin was added to the desired colony, and with the help of surgical needles, the compact-round-shaped stem cell colonies were lifted from the feeder-layer cells and carefully pipetted to a trypsin-containing well of a 96-well plate. The colonies were trypsinised in the 96-well plate at 37°C for 5 - 10 minutes and pipetted to disrupt the colony and obtain single cells. Each cell suspension was transferred into a well of a gelatine-coated and feeder containing 24-well plate containing mESC media. After two days, the cells were checked for attachment, and the media was replaced every second day.

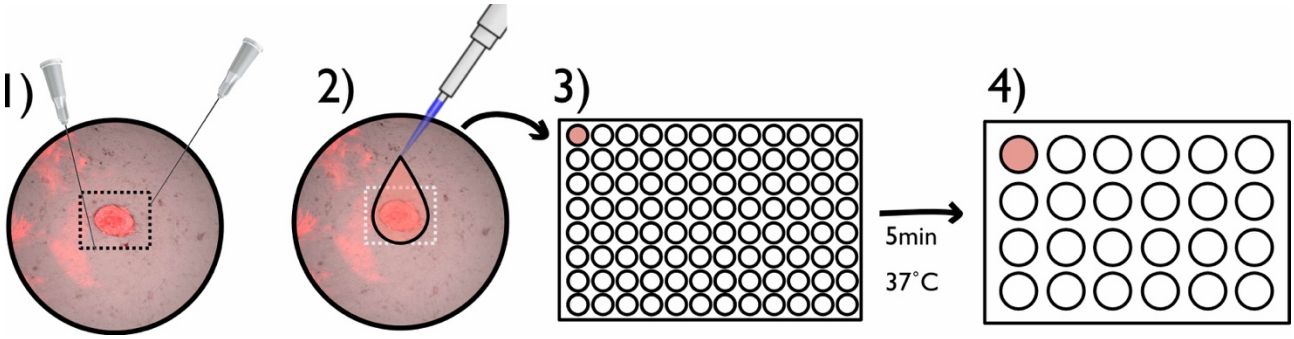


Figure 15: Murine SCs colony picking method
With the help of surgical needles, the area surrounding the colony of interest was cut. Then, a drop of trypsin was added, and the colony was aspirated and transferred to a 96-well plate containing trypsin. The colony was trypsinised for 5 minutes at 37°C and mechanically disrupted to get individual cells. Single cells were finally transferred into a gelatin-coated 24 well-plate with feeder layers.

4.1.12.2.3 Stem cells (hESC/hiPSC)

The day before picking human Stem Cell colonies, a gelatinised plate with feeder layers was prepared as described above. The day of picking, the hESC media was replaced, and the plate was visualised under the EVOS™ XL Core Cell Imaging System (Thermo Fischer Scientific). At the same time and with a 1000 µl sterile tip, hSC colonies were scratched and lifted from the plate. Then, the floating colonies were aspirated and transferred into a gelatinised plate with feeder layers, containing hSC media.

4.1.13 Luciferase Assay

4.1.13.1 *In vitro* Luciferase Assay

This method was used to measure Firefly Luciferase as a single reporter gene. The cells were washed twice with PBS and then incubated in serum-free DMEM containing Luciferin (10 µl/ml media) for 5 minutes at 37˚C. Afterwards, the plate was imaged using a FusionSL device, with exposure varying from 5 - 20 minutes. After imaging, the media was replaced with the appropriate media containing serum and antibiotics and the cells were kept in culture. Note that this was not a final measurement and the cells could be kept for other applications.

4.1.13.2 Dual Luciferase Assay (Promega)

This method was used to detect expression of Renilla and Firefly Luciferase, in dual-reporter systems. The assay was performed following the manufacturer's instructions, and the samples were analysed in triplicates. Before measuring, a working dilution (1:5) of passive lysis buffer (PLB) was diluted in water. Also, master mixes of LARII and Stop and Go (S&G) solutions were also prepared. Taking into account pipetting errors, the master mixes were prepared in excess. For the LARII master mix (50 µl/sample), LARII stock solution was diluted in Milli-Q water (1:5). For the S&G master mix (50 µl/sample), the S&G stock was diluted (1:5) in milliQ water and 1:50 of S&G substrate was added to the mix. Both master mixes and PLB working solution could be used immediately or stored at 4˚C.

The cells were then rinsed with PBS, trypsinised, counted (50.000 cells/sample) and harvested at 2000 rpm during 5 minutes. Then, the aliquoted cells were rinsed again, centrifuged and the cell pellet was re-suspended in 50 µl of PLB solution. Finally, the cells were incubated for 20 minutes at room temperature, shaking (1400 rpm) to allow cell lysis. Lysates could be analysed immediately or stored at -20˚C for further analysis.

To measure the expression of both Renilla and Firefly luciferase, 25 µl of cell's lysates (in triplicates) were loaded into a white non-transparent 96 well-plate, which was analysed in the Luminometer Mithras LB943, using the following analysis protocol: Dispense 50ul LARII, shake 2 seconds at lowest amplitude, measure 10 seconds in luminometric mode without filters set, dispense 50ul S&G solution, shake 2 seconds at the lowest amplitude and 'Measure'.

4.1.13.3 *In vivo* Luciferase Assay

A luciferin stock solution of 15mg/ml was prepared in DPBS and sterilised through a 0,2 µm filter. The aliquoted stock solutions could be used or stored at -20°C.

After the animals were anaesthetised, luciferin was injected (10 µl/g of body weight) intraperitoneally (IP), 10-15 minutes before imaging the animal, which allowed the Luciferin to reach the brain. For a 10 g mouse, 100 µL was injected to deliver 1.5 mg of Luciferin (or 150 mg luciferin/kg body weight). After 15minutes, the animals were sacrificed by an overdose of narcotic, dissected and the brains were collected and placed in unsupplemented RPMI containing luciferin (10 µL/ml). The brains were imaged as soon as possible the using Chemiluminescence Imaging system Fusion SL (exposure time 5 - 10 min). After imaging, the brains were washed in PBS and fixed overnight in 4% Paraformaldehyde at 4°C for immunohistochemical analysis.

4.1.14 Virology methods

4.1.14.1 Lentiviral production

Dr Joschka Willemsen and Dr Jamie Frankish, from the Virus Associated Carcinogenesis department (F170, DKFZ) kindly produced the lentiviral particles used in this study.

Briefly, HEK293T cells were infected with viral vectors derived from the plasmid pWPI-BLR. The lentiviral particles were produced as described elsewhere [291] by calcium phosphate transfection of the three individual plasmids at a 3:1:3 ratio: (i) pCMV-ΔR8.91, coding for HIV Gag-Pol; (ii) pMD.2G, encoding the VSV-G glycoprotein; and (iii) the lentiviral vector (v27) pRRL.PPT.SF.hOKSM-IRES-dTom or empty pWPI-backbone. (pCMV-ΔR8.91 and pMD.2G were kind gifts from Didier Trono, Lausanne [287]. Finally, the supernatants of HEK293T cells were collected, filtered and frozen for further use, without viral particle titration.

4.1.14.2 Lentiviral transduction

Mouse Embryonic Fibroblasts (MEFs) at the lowest possible passage (p3 - 5) were seeded (25.000 - 50.000 cells) in a well of gelatin-coated 24 well-plate (per transduction). Transduction was performed in three rounds. The next morning (day 1), the media was aspirated, and the cells were infected with 200 µl of viral particles suspended in DMEM media. The infection was repeated on the evening of the same day (day1) and the morning of the day after (day 2). Finally, the viral supernatant was aspirated,

and without washing steps, the cells were fed with standard DMEM supplemented with FCS and antibiotics.

Since Lentivirus (Retroviridae family) are ssRNA viruses, their genomes must be reverse transcribed and integrated into the host cell genome to be expressed, which takes approximately 72h. This means that transgene expression is expected to appear after 2 - 3 days.

4.2 Molecular biology methods

4.2.1 InFusion Cloning

InFusion cloning technology is designed for fast and directional cloning of one or more fragments of DNA into any vector. This technology relies on the homologous recombination of 15bp between the amplified insert(s) and the vector. The insert(s) are amplified by regular PCR and with primers containing 15bp of homology with the vector. The vector is digested with the desired restriction enzymes. The homologous recombination reaction occurs *in vitro* and the infused product can be then transformed into competent cells.

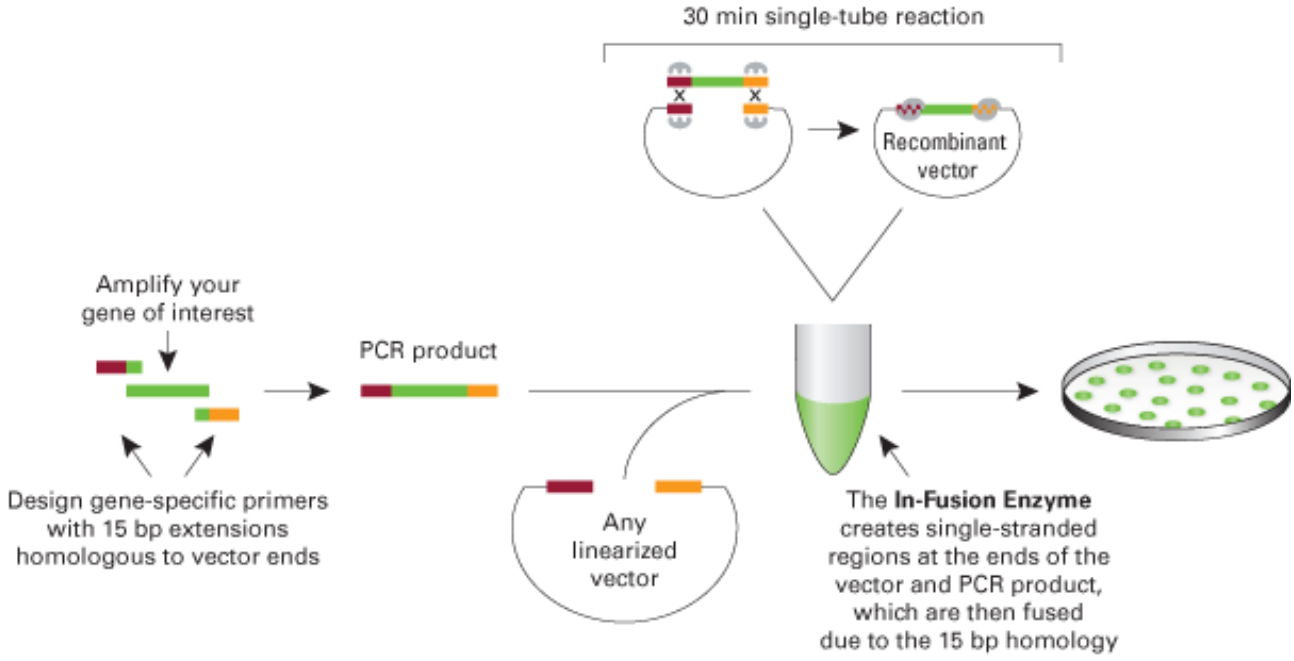


Figure 16: InFusion cloning principle
The DNA fragment(s) of interest is amplified by PCR using specific primers with 15bp overlap with the vector backbone. The backbone is linearised using specific restriction enzymes. Both insert and vector are purified from an agarose gel and mixed with the InFusion mix. During the reaction, the InFusion recombinase will recognise the 15bp of homology between vector and insert and will rearrange both fragments, generating a circular recombinant vector, which can be directly transformed into the appropriate *E.coli* strain of competent cells. *Image is taken from the InFusion HD Cloning kit user manual. TakaraBio.*

4.2.1.1 Processing of vector and insert

First, 1-3 µg of vector backbone were digested with 1-3 Units of the desired restriction enzyme(s) for 1h at 37˚C, following the example on **Table 32** (FastDigest restriction digestions could be incubated for 15 - 20min). In parallel, the insert was amplified by PCR using specific primers containing 15bp of homology with the vector as described **above.** Then, both PCR insert, and digested vector was loaded into a 1% agarose gel and run at 120V for 30 - 40 minutes. An undigested negative control (parental vector) was also run to assess complete digestion of the vector. Afterwards, the bands corresponding to both insert and digested vector were excised from the gel, and the DNA was extracted using a GeneElute Kit (Sigma), following the manufacturer's instructions. The concentration of both purified DNA fragments was assessed using spectrophotometry (Nanodrop).

Table 32: InFusion cloning: vector processing

Step	Reagent	Amout
Restriction Digestion (Vector)	Vector backbone	1-3 µg
	Restriction enzyme 1	1-3 U (1-3 µl)
	Restriction enzyme 2	1-3 U (1-3 µl)
	10x Restriction Buffer	5 µl
	dH_2O	Up to 50 µl

Table 33: InFusion cloning: insert processing

Step	Reagent			
PCR (insert)	DNA template		1-10ng	
	Forward primer		0,75 µl	
	Reverse primer		0,75 µl	
	2x HiFi mix		12,5 µl	
	dH_2O		Up to 25 µl	
	Program			
	Preheat lid	110˚C		
	Initialisation	95˚C	2 min	
	Denaturation	98˚C	10 sec	30 cycles
	Annealing	Tm-5˚C *	10 sec	
	Elongation	72˚C	X sec **	
	Final Elongation	72˚C	10 min	
	Final hold	8˚C		

* Annealing temperature depends on the primer's Tm

** CloneAmp HiFi polymerase has a processivity of 1000bp/5sec and the extension time was calculated according to the amplicon size.

Table 34: InFusion reaction

Step	Reagent	Amount
InFusion reaction	Vector (100ng) *	X µl **
	Insert (50 ng) *	Y µl **
	2x InFusion mix	2 µl
	dH_2O	Up to 10 µl

* The ratio vector: insert varied depending on their size

** The volume of vector + insert should not exceed 7 µl. If so, double the reaction volume.

4.2.1.2 Recombination 'In Fusion' reaction

To proceed with the recombination reaction, 100 ng of vector and 50 ng of the insert were mixed with water containing the 5x InFusion mix, containing the appropriate buffer and enzyme to allow homologous recombination between the 15bp of homology. It was essential that the volume of insert+vector did not exceed 7 µl. In such case, the InFusion reaction volume was doubled. The recombination reaction took place in a water bath at precisely 50°C for 15 minutes. Meanwhile, a 50 µl aliquot of *E.coli* Stellar Competent cells were thawed on ice.

4.2.1.3 Transformation of Stellar competent cells

The InFusion reaction containing the recombined plasmid product was placed on ice and 2,5 µl were used to transform the competent cells. The cells were incubated on ice for 30 minutes to allow the DNA to diffuse, subsequently heat shocked at 42°C for 45 seconds and immediately placed on ice for 2 minutes. Then, 450 µl of SOC media were added to the heat-shocked cells, which were incubated for 1h at 37°C (with gentle shaking) to allow their recovery. Passed this time, 100 µl of transformed cells were seeded in the appropriate LB-antibiotic plates and incubated overnight at 37°C.

4.2.2 Plasmid DNA preparation

The amount of pDNA and the preparation size varied depending on the downstream application. For cloning and generation of vectors, a small amount of DNA (usually few nanograms) is required and therefore, a mini preparation (miniprep) was used. For *in vitro* and *in vivo* applications, a higher amount is needed (µg or mg) and was obtained by using a maxiprep.

For animal experiments or for transfection of cells that would be later on injected into animals, an Endotoxin-free maxiprep was used to reduce bacterial products in the preparation, which could trigger undesired immune reactions.

All plasmid preparations were performed using Qiagen commercial kits (QIAprep spin miniprep kit, QIAGEN plasmid maxi kit and Endofree plasmid maxi kit) and prepared according to the manufacturer's instructions.

Regardless of the preparation format, the underlying procedure was standard for all preparations. First, the bacterial culture (5ml for minipreps and 300ml for maxipreps) was grown overnight at 37°C and shaking. Second, the bacterial cells were harvested by centrifugation. Third, the bacterial pellet was resuspended in RNAseA containing buffer and lysed under alkaline conditions (pH 12) given by the

presence of NaOH in the lysis buffer. Alkaline lysis allowed denaturation of chromosomal DNA and proteins while plasmid DNA remained stable. Successful lysis could be noticed by the colour change of the pH indicator (if contained in the buffers) or by the suspension's change of consistency and appearance of mucus, corresponding to the chromosomal DNA, intracellular proteins and solutes. Afterwards, the lysate was neutralised upon addition of acetate-containing buffer, which further precipitated large chromosomal DNA and proteins while plasmid DNA remained soluble. Once all intracellular molecules, except for the plasmid DNA, where denatured and precipitated, plasmid DNA was purified using Spin-columns, whose solid matrix bound to the negatively charged DNA and allowed other components to flow through. Plasmid DNA can be eluted upon changing the charge conditions of the column. Finally, the concentration was measured, and the plasmid DNA was either used or stored at -20°C.

4.2.3 Total DNA preparation

Isolation of total DNA (genomic, mitochondrial, viral and episomal) was performed either using a DNAeasy Blood and Tissue kit (QIAGEN) or a classical Phenol-chloroform extraction.

4.2.3.1 gDNA extraction (DNeasy Blood and tissue, QIAGEN)

For **cultured cells**, 1,5x10^6 cells were counted, washed and resuspended in 200µl PBS. Then, 20µl of Proteinase K were added per sample.

For **tissues**, 30mg of tissue was homogenised in 180µl ATL buffer and 20µl of Proteinase K in a Lysing Matrix D (MP biomedicals). The lysate was then centrifuged at 4000rpm for 5 minutes, and the supernatant was separated from tissue debris and collected in a new tube.

Afterwards, 200µl of buffer AL were added to the homogenates, followed by brief vortexing. Then, 200µl of 100% Ethanol was added and mixed with the lysates by vortexing. The mixture was transferred into a DNeasy spin column, which was then centrifuged at 8000rpm for 1 minute. The flow-through was discarded, and the column was washed with 500µl of Buffer AW1, followed by another wash with 500µl of Buffer AW2. Finally, the column was air dried by centrifugation during 3 minutes at 14000rpm, and the tDNA was eluted with two rounds of 50 µl Buffer AE. The concentration of the 100µl eluate was measured, and the tDNA was either used or stored at -20°C.

4.2.3.2 Phenol-Chloroform extraction

For this extraction protocol, three critical points were considered: 1) the tubes were not vortexed but inverted, 2) pipette tips were cut to prevent shearing of the nucleic acids and 3) pipetting was performed under a chemical hood since most chemicals used in this protocol were toxic and volatile.

To purify tDNA from cultured cells, 10^6 to $5x10^6$ cells were harvested, pelleted and resuspended in 500 µl of DNA lysis buffer 2 **(see Section 3.6)**. Proteinase K was added to a final concentration of 600 µg/ml (from a 25 mg/ml stock), and the cells were lysed upon addition of 500 µl of DNA lysis buffer 1. The tube was carefully inverted, and the digestion took place at 42°C for 1h or at room temperature overnight. Then, the Proteinase K treatment was repeated by adding another 600 µg/ml of protein and incubated overnight at room temperature, to ensure complete digestion.

Alternatively, 30 mg of tissue or 0.2 - 0.4 cm tail biopsies were lyzed in 100 µl of DNA lysis Buffer 2, followed by addition of proteinase K (600 µg/ml) and 100 µl of DNA lysis Buffer 1 and incubated overnight at 12000 rpm at 42°C.

The day after, the lysate was incubated with 50 µg/ml of RNAseA (from a 10 mg/ml stock) for 1h at 37°C. Then, one volume (1ml) of phenol:chloroform:isoamyl alcohol (25:24:1) was added, and the tube was inverted until emulsion droplets formed and the mix became cloudy. The emulsion was centrifuged at 13000 rpm during 5 minutes at room temperature and resulted in the separation of 2 phases: 1) an organic, apolar, denser-phase, containing the phenol-chloroform and proteins; and 2) an aqueous, transparent upper-phase containing soluble DNA. The upper aqueous phase was collected into a fresh tube and another volume of phenol: chloroform: isoamyl alcohol was added, to further purify the DNA from proteins. After another round of centrifugation, the upper phase was again collected and transferred into a fresh tube, in which one volume of chloroform was added. The tube was inverted until emulsion droplets formed and centrifuged at 13000 rpm during 5 minutes at room temperature. The upper phase was again collected and the DNA was precipitated by adding 1/10 volumes of 3M sodium acetate (which provided positively charged ions to neutralise the negative charges of DNA, which made it less soluble in water) and 3 volumes of 100% Ethanol (which is much less polar than water and allows stable ionic bonding between the negatively charged DNA and the positively charged ions, which allowed DNA precipitation).

The precipitation was performed overnight at -20°C or during 1h at -80°C. Next day, the precipitated DNA was pelleted for 30 minutes at 13000rpm and 4°C, washed with 75% Ethanol for 5 minutes and

centrifuged again for 1 minute. The supernatant was removed, and the pellet was air-dried to allow evaporation of ethanol traces. Finally, the DNA pellet was re-dissolved in 200 µl TE Buffer, its concentration was measured, and it was either used or stored at -20°C.

4.2.4 Genotyping of biological samples

Ear punches, tail or organ biopsies were processed using the Phire Tissue Direct PCR Master Mix (Thermo Fischer), which served for lysing and releasing the DNA as well as to use the lysed tissue for PCR analysis directly. Briefly, 20 µl of Dilution Buffer and 0,5 µl of DNA release were added to the biological samples. Brief vortexing and spinning of the samples were performed to ensure the totality of biopsy was covered in the buffer. Then, the samples were incubated 5 minutes at room temperature followed by 5 minutes at 98°C. Afterwards, 1 – 2 µl of the reaction were directly used as a template for PCR or stored at -20°C.

4.2.5 DNA/RNA electrophoresis

Electrophoresis was used to either visualise nucleic acids after extractions, to confirm successful PCR amplifications or successful restriction digestions. This technique relies on the separation of nucleic acids based on its size, and how they migrate in an agarose matrix when an electric field is applied. For that, an agarose gel was prepared in Electrophoresis Buffer **(see Section 3.6)**. The percentage of agarose (ranging from 0,8 - 2%) and the intensity of the electric field (60 – 120 V) varied depending on the nucleic acid's size and the desired degree of separation. In general, 1% agarose gels were used and run at 120 V for 30-40 min.

4.2.6 RNA extraction

4.2.6.1 RNA extraction from cultured cells

RNA was extracted from cell pellets using the High Pure RNA isolation kit (Roche) and following the manufacturer's instructions. For that, 10^6 cells were washed twice in PBS, trypsinised, counted and resuspended in 200 µl of PBS. Then, 400 µl of Lysis/Binding buffer was added to the cells, which were vortexed for 15 seconds. The lysate was then transferred to a High Pure filter, placed in a collection tube, and the nucleic acids were bound to the column by centrifugation at 8000 g for 30 seconds. After centrifugation, the flow through was discarded, and the DNA was digested in the column upon addition of 10 µl of DNAseI in 90 µl of DNaseI incubation buffer for 15 minutes at room temperature. Past the incubation time, 500 µl of Washing Buffer 1 were added to the upper reservoir followed by a

1 minute centrifugation at 8000 g. Then, two more washing steps were performed by first adding 500 µl of Washing Buffer II and centrifuging at 8000 g and then adding 200 µl and centrifuging to remove residual buffer. Finally, the RNA was eluted with 50 µl of Elution Buffer and by centrifugation at 13000g for 1 minute. The eluted RNA could be directly used for RT-PCR or stored at -80˚C for subsequent analysis.

4.2.6.2 RNA extraction from tissues

RNA from animal tissues was extracted using RNeasy Mini Kit (Qiagen), following the manufacturer's instructions. Before starting, the lysing solution was prepared by mixing (per sample) 1ml of RLT buffer with 10 µl of β-mercaptoethanol. The starting material (<30mg of frozen tissues) was first homogenised in 600 µl of lysing solution (RLT + β-mercaptoethanol) using Lysing Matrix D columns (MP biomedicals) in a homogenizer (Precellys 24, Peqlab) using program P1. The supernatant containing the homogenate was then transferred into a microcentrifuge tube and spun 3 minutes at full speed to pellet cell debris. Afterwards, the supernatant was transferred into a new microcentrifuge tube, and 1 volume of 70% RNAse-free Ethanol was added to the lysate. At this point, the lysates could be processed manually or loaded into a QIAcube robot.

The lysates were then loaded into an RNeasy column, centrifuged at >8.000g for 1 minute and the flow through was discarded. Then, successive washing steps were performed by first adding 700 µl of Buffer RW1 and spinning, followed by two washes with 500µl of RPE buffer and spinning. The membrane was then air dried by centrifugation for 1 minute, and the RNA (together with DNA) were eluted using 50 µl of RNAse-free water (DEPC-water). The RNA concentration was then measured by spectrophotometry (Nanodrop) and the extractions were stored at -80˚C for further analysis.

Note: To remove the DNA present in the samples, DNAseI treatment was performed before downstream applications (Ambion DNA-free Kit, Invitrogen).

4.2.7 Protein extraction – Whole cell lysates

4.2.7.1 Overtrypsinisation of cells

The cells were washed twice with PBS to remove medium components (especially divalent ions such as Calcium or Magnesium, which inactivate Trypsin), trypsinised and harvested with an equal volume of media. Then, they were transferred to a centrifuge tube and spun down at 1000 rpm for 5 minutes.

The supernatant was aspirated, and the pellet placed immediately on ice and resuspended in 40 µl RIPA Lysis Buffer **(See Section 3.6)**.

4.2.7.2 Cell scrapping

The cells were washed twice in PBS and plated on ice. Then, 400 µl of RIPA Lysis Buffer was added and spread through the plate with a cell lifter. After scrapping the cells from the plate, the suspension was drawn up through a 21G needle and transferred into a tube.

In both cases, the suspensions were pipetted up and down or drawn up through a 21G needle to further homogenise and shear the cells. The lysates were incubated on ice for additional 10 minutes and were subsequently centrifuged at 14000rpm for 10 – 20 min at 4˚C to pellet cell debris. The supernatant was collected and either stored at -80˚C or analysed to determine the total amount of protein using BCA assay.

4.2.8 Polymerase chain reaction (PCR)

The PCRs for cloning procedures were performed using the primers indicated in the section Cloning primers and the CloneAmp™ HiFi PCR Premix (Clontech) following manufacturer's instructions. The primers used to amplify DNA inserts for InFusion cloning contained 15bp of homology with the vector.

PCRs used for validation of rescued vectors were performed using S/MAR primers (primers 1, 2) and primers amplifying the expression cassette (primers 36, 128) and the PCR program described above.

4.2.8.1 Genotyping PCR

DNA was extracted with the Phire Tissue Direct PCR Master Mix (Thermo Fischer), and the PCR was performed according to the manufacturer's instructions, in a total volume of 20 µl per reaction. Degenerate primers amplifying a mammalian genomic DNA region of Sox21 were used as internal controls (provided with the kit). GFP primers for amplification of 515bp coGFP were designed using Primer3 and validated *in silico*. The PCR reactions for GFP and the internal controls were performed separately to get better amplification of GFP bands.

Table 35: Genotyping primers

Primer	Sequence (5'->3')	Annealing (°C)	Amplicon (bp)
coGFP1	GCCGCATGACCAACAAGATG	69,3	515
coGFP2	GTTGCTGTGCAGCTCCTCCA		
eGFP.For	GCAA GGGC GAGG AGCT GTTC ACC	72	329
eGFP.Rev	GGCG AGCT GCAC GCTG CCGT CGTC		
Sox21.For	AGCCCTTGGGGASTTGAATTGCTG	72	237
Sox21.Rev	GCACTCCAGAGGACAGCRGTGTCAATA		

Table 36: PCR reaction components

Reagent	GFP PCR Amount (µl)	Internal control PCR Amount (µl)
DNA template	1 µl	1 µl
Universal control primer mix	-	0,4 µl
Forward primer (20µM)	0,5 µl	-
Reverse primer (20µM)	0,5 µl	-
2x Phire PCR Premix	10 µl	10 µl
dH_2O	8 µl	9,6 µl

Table 37: PCR conditions

Program			
Preheat lid	110°C		
Initial denaturation	98°C	5 min	
Denaturation	98°C	5 sec	40 cycles
Annealing	69,3°C*	5 sec	
Elongation	72°C	30 sec	
Final Elongation	72°C	1 min	
Final hold	4°C	Hold	

4.2.9 Quantitative Real-Time PCR (qPCR)

Quantitative PCR was used for 1) absolute quantification of S/MAR vectors to determine their copy number in established cell lines or for 2) relative quantification of transcription factors during reprogramming of iPSC. In both cases, qPCR was performed using QuantiTect SYBR Green PCR Kit (QIAGEN) and a Real-Time PCR System LightCycler® 96 (Roche). Absolute and relative quantifications were performed using the LightCycler® 96 Software (Roche).

4.2.9.1 Absolute quantification – Copy number Assay

To quantify the number of S/MAR-vector copies present in established cell lines, gDNA was extracted as described either using Phenol-Chloroform or DNeasy Blood and Tissue kit (QIAGEN) as described

above. Two standard curves were performed to determine 1) the number of cells present in each sample and 2) the number of vector molecules present per cell.

A **standard curve to determine the number of cells per sample** was generated by performing 1:5 serial dilutions of gDNA (50, 10, 2, 0.4 and 0.08 ng/µl). GAPDH primers were used to determine how many genome copies were present in the samples, assuming that the cells were diploid and contained 2 copies of GAPDH. Then, 1µl of each serial dilution was mixed with 10µl of 2x SYBR green PCR mix, 0.1µl of each primer and 8.8µl of RNAse free water, to a total volume of 20µl per reaction. Each dilution was loaded in triplicates. A water negative control (also run in triplicates) without gDNA, was used to rule out the possibility of primer dimers or gDNA contamination.

A **standard curve to determine the number of vector molecules per sample** was generated by performing 1:10 serial dilutions of the desired vector. Briefly, a miniprep of the desired vector was diluted to 20 ng/µl and subsequently serial diluted (from $2x10^{1}$ until $2x10^{-10}$ ng/µl) to potentially achieve a single molecule dilution. A range of six serial dilutions was chosen to build the standard curve. Note that the criteria for choosing the dilutions depended on how many vector molecules were expected per cell, in other words; when higher copy numbers were expected, a higher range of dilutions was chosen and vice versa. Puromycin primers were chosen to determine the copies of the vector. Note that S/MAR primers would not have been suitable since S/MAR motifs are naturally occurring in the genome. Then, 1 µl of each serial dilution was mixed with 10 µl of 2x SYBR green PCR mix, 0.1 µl of each primer and 8.8 µl of RNAse free water, to a total volume of 20 µl per reaction. Each dilution was loaded in triplicates. A water negative control (also run in triplicates) without vector was used to rule out the possibility of primer dimers.

Finally, each sample was diluted to 5 ng/µl, and 1 µl (5 ng) per reaction was amplified with GAPDH and Puromycin primers (both run in triplicates), following the reaction conditions and program explained in **Table 38** and **Table 39**.

To minimise pipetting errors, 2 master mixes were prepared per sample: One containing the primers and the 2xSYBR green mix; and the other containing the template (gDNA or vector) diluted in water. To minimise plate-loading errors, the Primer master mix was pipetted horizontally, and the DNA mix was pipetted vertically, following the Plate loading scheme in **Table 40**. The plate was carefully sealed, centrifuged at 1000rpm for 1 minute, and run into a LightCycler® 96 (Roche) following the cycling conditions in **Table 39**.

Table 38: qPCR reaction conditions (per sample)

	Reagent	Amount	
Primer master mix	2x SYBR green PCR mix	10	µl
	Forward primer	0.1	µl
	Reverse primer	0.1	µl
Sample master mix	Sample (cDNA, gDNA, vector)	1	µl
	RNAse free-H_2O	8.8	µl
	TOTAL volume	20	µl

Note: take into account the number of replicates when preparing the master mixes. Each sample must be run in triplicates

Table 39: qPCR program conditions

	Step	Time (sec)	Temp (°C)
Heat inactivation		900	95
3-step cycling (45 cycles)	Denaturation	15	94
	Annealing	25	60*
	Extension	20**	72
Melting curve		10	95
		60	65
		1	97
Cooling			

* All qPCR primers were designed to have an annealing temperature of 60°C the free software Primer3
** The extension time varied according to the amplicon length, varying from 100-120 bp

Table 40: qPCR loading scheme

		1	2	3	4	5	6	7	8	9	10	11	12	
Standard gDNA	A	50	10	2	0.4	0.08	H_2O	H_2O	S1	S2	S3	S4	S5	Primers (GAPDH)
	B	50	10	2	0.4	0.08	H_2O	H_2O	S1	S2	S3	S4	S5	
	C	50	10	2	0.4	0.08	H_2O	H_2O	S1	S2	S3	S4	S5	
Standard vector	D	$2E^0$	$2E^{-1}$	$2E^{-2}$	$2E^{-3}$	$2E^{-4}$	$2E^{-5}$	$2E^{-6}$	S1	S2	S3	S4	S5	Primers (Puro)
	E	$2E^0$	$2E^{-1}$	$2E^{-2}$	$2E^{-3}$	$2E^{-4}$	$2E^{-5}$	$2E^{-6}$	S1	S2	S3	S4	S5	
	F	$2E^0$	$2E^{-1}$	$2E^{-2}$	$2E^{-3}$	$2E^{-4}$	$2E^{-5}$	$2E^{-6}$	S1	S2	S3	S4	S5	
	G													
	H													

Each square corresponds to a well, and each sample was run in triplicates.
The values of the standard curves correspond to the concentration of DNA and the amount (ng) of DNA loaded in 1µl. For each sample, 5ng were added per reaction, corresponding to 1µl of a 5ng/µl dilution.

Data analysis was performed using the LightCycler® 96 Software (Roche) and Microsoft Excel.

First, the **amount (ng) of gDNA for a single diploid** cell was calculated following the next equation:

Length of diploid genome= $6.47x10^9$ bp

Weight/bp = 660 Da (g/mol)

$$molecular\ weight\ of\ single\ genome\ (\frac{g}{mol}) = 6.47x10^{7} \times 660 = 4.27x10^{12}$$

Then, using Avogadro's number ($6.023x10^{23}$) we calculate the weight in (ng) for a single haploid genome, which is 7 pg or 0.007 ng

$$g\ of\ single\ genome = \frac{4.27x10^{12}}{6.023x10^{23}} = 7x10^{-12}\ \ or\ \ 7\ pg$$

Second, the **amount of cells per sample** was calculated by dividing the concentration of cells (ng of cells in 1 µl of loaded sample), which was calculated from the Cq value, by the weight of a diploid genome (0,007 ng):

$$\frac{cells}{sample} = \frac{ng\ cells}{0.007\ ng} = X\ cells$$

For example: Amplification of Sample #1 with GAPDH primers gave a Cq mean of 24.18, which corresponded to 11.90 ng of gDNA.

In this case, 11.90 ng cells/ 0.007 ng genome = **$1.7x10^{3}$ cells**

Third, the **number of vector molecules** was calculated by taking into account the vector's size (bp) and by using online resources (http://cels.uri.edu/gsc/cndna.html).

For example, for a 11921bp vector, there are $1.55x10^{9}$ molecules in 20 ng. This relationship should be kept constant for all the points of the standard curve.

Fourth, the **number of vector copies** was calculated by multiplying the concentration of vector, which was calculated from the Cq value, for the number of known copies from the standard curve and divided by the number of ng loaded in the standard curve:

$$\frac{vectors}{sample} = \frac{ng\ vector\ (in\ 1ul) * molecules\ of\ vector(from\ standard)}{ng\ vector\ (standard)} = Y\ copies$$

For example, amplification of Sample #1 with Puromycin primers gave a Cq mean of 21.19, which corresponds to $4.61x10^{-5}$ ng of vector.

In this case, ($4.61x10^{-5}$ ng vector x $1.55x10^{9}$ molecules of vector)/20 ng vector = **$6.94x10^{3}$ copies**

Finally, the number of **vector copies per cell** was determined by dividing the Y number of vector molecules by the X number of cells.

$$\frac{vectors}{sample} = \frac{Y\ copies}{X\ cells}$$

For the stated example, $6.94x10^{3}$ copies/$1.7x10^{3}$ cells = **4.1 copies/cell**

These calculations were applied to each sample and their errors to represent the values of copies/cell in a column chart with error bars.

4.2.10 Reverse Transcription PCR (RT-PCR)

RT-PCR was used to confirm expression of pluripotency markers in fibroblasts, mESC and reprogrammed iPSC. For that purpose, the cells growing in a 6-well plate were harvested. Fibroblasts were directly harvested, and Stem Cells were harvested after differential sedimentation. Then, total RNA was extracted as described **above,** using the High Pure RNA isolation kit (Roche) with on-column DNAseI treatment and following the manufacturer's instructions.

The reverse transcription reaction was performed using M-MLV Reverse Transcriptase (Promega) in its supplied buffer, and all the reactions were performed in triplicates. A negative control (RT- control) without Reverse Transcriptase was also performed in triplicates to rule out the possibility of gDNA contamination after DNAseI treatment.

The cDNA synthesis had a total volume of 20µl per reaction. First, 1µg of RNA was mixed with 2µg (1µl from a 2µg/µl aliquot) of random hexamer primers (dN6) and RNAse free-H_2O to a total volume of 10µl.The mixture was pre-incubated at 70°C for 10 minutes. Then, 4 µl of 5x M-MLV buffer, 2 µl of 100mM DTT (pH=7), 0,5µl of dNTPs (25mM each), 0,5 µl of RNAseOUT, 1µl of M-MLV reverse transcriptase and 2 µl of RNAse free-H_2O were added to the mix, to reach a final volume of 20µl. The reactions were incubated in a thermocycler using the conditions described in **Table 42.** Specific RT-PCR parameters are detailed below.

Table 41: Reverse Transcription reaction conditions

Reagent	Amount	
Total RNA	1000	ng
Random hexamers (2µg/µl)	1	µl
RNAse free-H_2O	Up to 10	µl
Then add per reaction…		
RNAse free-H_2O	2	µl
5x M-MLV Buffer	4	µl
DTT (100mM)	2	µl
dNTPs (25mM)	0,5	µl
MMLV-RT	1	µl
RNAseOUT	0,5	µl
TOTAL	20	µl

Table 42: Reverse Transcription cycling conditions

Step	Temperature	Time
Pre-incubation	70 °C	10 minutes
Reverse Transcription	37 °C	10 minutes
	42 °C	60 minutes
	46 °C	20 minutes
	52 °C	20 minutes
	56 °C	20 minutes
	95 °C	5 minutes
	4 °C	storage

The generated complementary DNA (cDNA) had an estimated concentration of 50 ng/µl, assuming that all RNA (1000 ng) was reverse transcribed into cDNA in a 20 µl reaction volume. Such cDNA was either stored at -20°C or used for quantitative PCR analysis.

4.2.10.1 Relative quantification – Expression of reprogramming factors

In order to distinguish between endogenous pluripotency factors, switched on when the cell becomes pluripotent, and exogenous reprogramming factors (OKSM) delivered either with the 4in1 Lentivirus or the S/MAR vectors (POP and nPOP); a relative quantification by qRT-PCR was performed. This was only possible because the exogenous reprogramming factors are human codon optimised and differ from the endogenous mouse factors.

RNA from 1) wild-type ESC (containing endogenous factors), 2) ESC transduced with a Lentivirus (Endogenous and exogenous factors), 3) wildtype MEFs (no factors) and 4) MEFs transduced with a Lentivirus (exogenous factors) was used. The expected primer amplifications are shown below.

	1)mESC ∅	2)mESC + Lenti OKSM	3)MEFs ∅	4)MEFs + Lenti OKSM
Endogenous (murine) OKSM	YES	YES	NO	NO
Exogenous (human codon-optimized OKSM)	NO	YES	NO	YES

Primer sets targeted to amplify endogenous murine or exogenous human codon-optimized Oct4, Klf4, Sox2 and cMyc (**Section 3.11.2**) were designed and tested *in silico*.

RNA was extracted, and cDNA was synthesised at a concentration of 50 ng/µl, as explained above. Then, cDNA serial dilutions of each sample (50 ng/µl, 5 ng/µl, 0,5 ng/µl and 0,05 ng/µl) were performed to build a standard curve and determine the primer pair efficiency.

The exogenous primers were tested in cDNA from HEK cells transfected with exogenous OKSM, and 1µl of each serial dilution as well as water controls were run in duplicates.

The endogenous primers were tested in cDNA from mESC and 1µl of each serial dilution as well as water controls were run in duplicates.

Also, 1 µl of the 5 ng/µl dilution of each cDNA (1 - 4) was used as a template and run in triplicates for each primer set. A more comprehensive overview of the experiment is shown in the following tables:

Table 43: Standard curve for exogenous (human) primers and endogenous (murine) primers

		1	2	3	4	5	6	7	8	9	10	11	12	
A	*cDNA HEK 35*	50	5	0,5	0,05	H2O	50	5	0,5	0,05	H2O	50	50	*cDNA HEK 35*
B	*cDNA HEK 35*	50	5	0,5	0,05	H2O	50	5	0,5	0,05	H2O	5	5	*cDNA HEK 35*
C	*cDNA HEK 35*	50	5	0,5	0,05	H2O	**50**	**5**	**0,5**	**0,05**	**H2O**	**0,5**	**0,5**	***cDNA mESC***
E	*cDNA HEK 35*	50	5	0,5	0,05	H2O	**50**	**5**	**0,5**	**0,05**	**H2O**	**0,05**	**0,05**	***cDNA mESC***
F	*cDNA HEK 35*	50	5	0,5	0,05	H2O	**50**	**5**	**0,5**	**0,05**	**H2O**	**H2O**	**H2O**	***cDNA mESC***
G	*cDNA HEK 35*	50	5	0,5	0,05	H2O	**50**	**5**	**0,5**	**0,05**	**H2O**			***cDNA mESC***
H	*cDNA HEK 35*	50	5	0,5	0,05	H2O	**50**	**5**	**0,5**	**0,05**	**H2O**			***cDNA mESC***
I	*cDNA HEK 35*	50	5	0,5	0,05	H2O	**50**	**5**	**0,5**	**0,05**	**H2O**			***cDNA mESC***

Table 44: Test of exogenous (human) primers in murine cells

		1	2	3	4	5	6	7	8	9	10	11	12	
A	*MEF ø*	*hOct4*	*hKlf4*	*hSox2*	*hMyc*	mGAPDH	*hOct4*	*hKlf4*	*hSox2*	*hMyc*	mGAPDH			***mESC***
B	*MEF ø*	*hOct4*	*hKlf4*	*hSox2*	*hMyc*	mGAPDH	*hOct4*	*hKlf4*	*hSox2*	*hMyc*	mGAPDH			***mESC***
C	*MEF ø*	*hOct4*	*hKlf4*	*hSox2*	*hMyc*	mGAPDH	*hOct4*	*hKlf4*	*hSox2*	*hMyc*	mGAPDH			***mESC***
E	*MEF ø*	RT-	RT-	RT-	RT-	RT-	RT-	RT-	RT-	RT-	RT-			***mESC***
F	*MEF Lenti*	*hOct4*	*hKlf4*	*hSox2*	*hMyc*	mGAPDH	**5 hOCT4**	**5 hKLF4**	**5 hSOX2**	**5 hMYC**	**5 mGAPDH**			***cDNA HEK35***
G	*MEF Lenti*	*hOct4*	*hKlf4*	*hSox2*	*hMyc*	mGAPDH	**5 hOCT4**	**5 hKLF4**	**5 hSOX2**	**5 hMYC**	**5 mGAPDH**			***cDNA HEK35***
H	*MEF Lenti*	*hOct4*	*hKlf4*	*hSox2*	*hMyc*	mGAPDH								***cDNA ESC***
I	*MEF Lenti*	RT-	RT-	RT-	RT-	RT-								

Table 45: Test of endogenous primers in murine cells

		1	2	3	4	5	6	7	8	9	10	11	12	
A	*MEF ø*	mOct4	mKlf4	mSox2	mMyc	mNanog	mGAPDH	mOct4	mKlf4	mSox2	mMyc	mNanog	mGAPDH	*mESC*
B	*MEF ø*	mOct4	mKlf4	mSox2	mMyc	mNanog	mGAPDH	mOct4	mKlf4	mSox2	mMyc	mNanog	mGAPDH	*mESC*
C	*MEF ø*	mOct4	mKlf4	mSox2	mMyc	mNanog	mGAPDH	mOct4	mKlf4	mSox2	mMyc	mNanog	mGAPDH	*mESC*
E	*MEF ø*	RT-	RT-	RT-	RT-	RT-	RT -	RT-	RT-	RT-	RT-	RT-	RT -	*mESC*
F	*MEF Lenti*	mOct4	mKlf4	mSox2	mMyc	mNanog	mGAPDH	5 mOCT4	5 mKLF4	5 mSOX2	5 mMYC	5 mNANOG	mGAPDH	*retrofit*
G	*MEF Lenti*	mOct4	mKlf4	mSox2	mMyc	mNanog	mGAPDH	5 mOCT4	5 mKLF4	5 mSOX2	5 mMYC	5 mNANOG	mGAPDH	
H	*MEF Lenti*	mOct4	mKlf4	mSox2	mMyc	mNanog	mGAPDH							
I	*MEF Lenti*	RT-	RT-	RT-	RT-	RT-	RT -							

It was necessary to include one point of the standard curve on each plate so that the software could apply the primer's efficiency to each sample and analyse their relative quantification.

The PCR reaction set-up and cycling conditions are those cited above in **Table 38** and **Table 39**, as the primers were designed to all have the same annealing temperature and extension time.

After the standard curves were built, the primer efficiency and the relative quantification of the exogenous and endogenous genes were determined using the LightCycler 96 software.

4.2.11 Plasmid Rescue

Rescuing of plasmid DNA is one way to prove that a vector is not integrated and that it remains in its episomal state. This protocol relies on the ability of circular DNA to re-transform bacteria after its recovery from mammalian cells; in other words, an episomal circular DNA that was used to transfect mammalian cells should replicate and segregate with the cells keeping its integrity and can, therefore, be recovered from gDNA preparations and used to retransform bacteria.

4.2.11.1 Plasmid Rescue Protocol 1

A protocol for successful plasmid rescue was designed and optimised in this study.

Initial attempts to rescue plasmids were performed, in which the gDNA directly extracted from mammalian cells was transformed into *E.coli* competent cells. However, no colonies were grown, which suggested that the vector could not be rescued. This result did not necessarily mean that the vector was 'unrescuable' or not episomal, but rather that some intermediate steps were required to enrich the episome and facilitate its entry into bacteria. For example, after extraction of DNA, all DNA forms were present and mixed together. Since gDNA is the largest and most abundant of all, it might create a 'DNA matrix' in which all DNA molecules are tangled to each other. In this case, although a circular DNA might not be integrated into the genome, it will probably not retransform bacteria, simply because

it is tangled and trapped with gDNA and because the bacteria's pore size is limited and will not allow molecules larger than a certain size to enter, therefore the plasmid DNA was not able to retransform bacteria and could not be rescued.

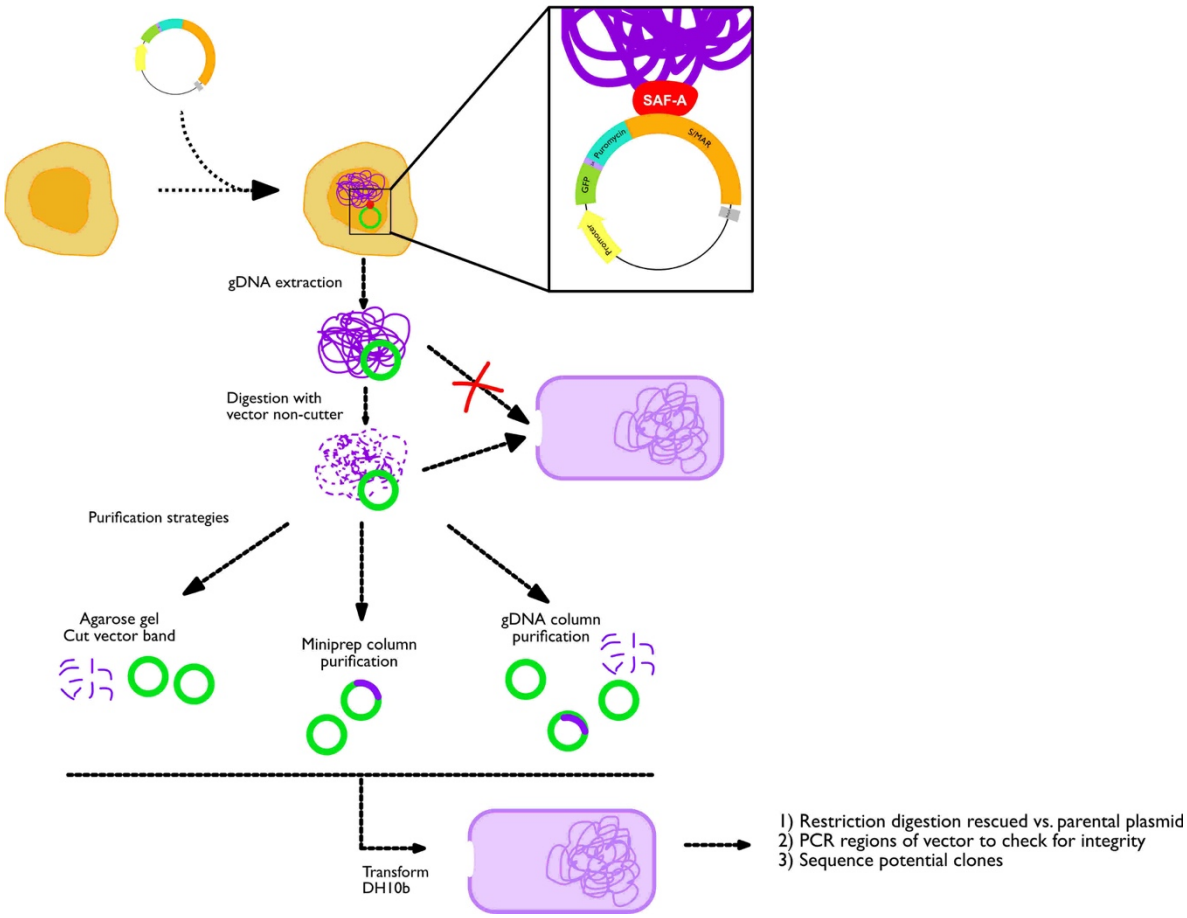


Figure 17: Cartoon depicting a Plasmid rescue protocol
Eukaryotic cells (HEK293T, Be2C, MEFs or mESC were labelled with GFP-SMAR episomal vectors. After generating stable cell lines with established vectors, the cells were lysed, and total DNA (genomic, mitochondrial and episomal) was digested with Proteinase K, to release the vector from proteins, and extracted using Phenol/Chloroform and precipitated using ethanol. The large and tangled gDNA was digested with a plasmid's non-cutter enzyme, and the circular DNA was purified either using i) agarose gel electrophoresis, ii) miniprep spin column or iii) gDNA spin column. The enriched and purified circular DNA was then electroporated into *E.coli* DH10β competent cells. The clones where then miniprepped and the DNA was digested and subjected to PCR to check for vector rearrangements and integrity. Finally, successfully rescued clones were confirmed by sequencing.

First, total DNA was extracted using Phenol-Chloroform extraction and ethanol precipitation as described above. Then, 10 µg of total DNA (containing genomic, mitochondrial and episomal DNA) was digested for 1h at 37 °C (**Table 32**) with a vector non-cutter, in other words, an enzyme that digested the genomic DNA but kept the plasmid intact. This was a crucial step to cut and reduce the gDNA in size and to 'untangle' and release the circular episomal DNA. An undigested total DNA control was used to assess correct DNA digestion.

At this point, one of the three purification strategies could be used to enrich the episomal DNA:

- **DNA Electrophoresis and cutting of the approximate DNA band**

The digested and undigested control DNA was run on a 1% agarose gel, at 120V for 45 minutes and the presence of a genomic DNA smear confirmed the successful digestion. To purify as much episomal DNA as possible, the agarose gel was cut around the expected episome size (*e.g.*: 6000bp) ± 1000bp to avoid losing different conformations of the circle that might have run differently in the gel. One must keep in mind that this purification step is 'targeting' or forcing the purification of the 'correct' or parental episomal vectors and other rearranged forms of different sizes (if any) are not selected and therefore not considered.

The circular DNA together with linear gDNA fragments of similar size were recovered from the gel band using the GeneElute Kit (Sigma).

- **'Miniprep' column**

Digested DNA was treated as if it was bacteria by following the miniprep kit manufacturer's protocol. With this purification step, the circular DNA was selectively purified in a Spin column while the linear gDNA was washed away. Note that the purpose of doing so and of using the 'unnecessary' buffers for bacterial lysis is to provide the appropriate ionic charge for the spin column in order to avoid premature elution of the circular DNA, which is diluted in water.

First, 250 µl of resuspension buffer containing RNAseA followed by 250 µl of Lysis buffer were added to the digested DNA. Then, the solution was neutralised upon addition of 350 µl of Neutralisation buffer containing sodium acetate and subsequently bound to the spin column. The plasmid DNA was washed twice with 500 µl of isopropanol based buffer followed by 750 µl of Ethanol based buffer. Finally, the circular DNA was eluted with 30 µl of pre-warmed elution buffer (EB) provided by the kit.

- **'gDNA clean and concentrator' column**

The digested DNA was purified using a commercial gDNA purification kit and following the manufacturer's instructions, with the intention of purifying also circular DNA. For this, 2 volumes of CHIP binding buffer were added to the DNA digestion product, which was then loaded into the spin column and centrifuged at 13000 rpm for 30 seconds. Upon binding of the DNA to the column, 200 µl of washing buffer were added and the column was again centrifuged. Finally, the DNA was eluted with 30 µl of pre-warmed elution buffer.

Once the DNA was purified using one of these methods, 100-200 ng (no more than 5 µl) of purified episomal plasmid were electroporated into *E.coli* DH10β competent cells. For this, an aliquot of 20 µl of competent cells was thawed on ice. An electroporation cuvette with 0,1 cm gap and the DNA was aliquoted in a fresh tube, were also placed on ice. Once the cells were thawed, they were added to the DNA and the mix was transferred to the cooled electroporation cuvette, which was quickly electroporated with the BioRad micropulser's 'Bacterial program'. Immediately after, 950 µl of SOC media were added to the cells and they were transferred into a 15ml Falcon tube, where they were incubated for 1 - 2h at 37˚C, shaking. Past the recovery time, the cells were harvested at 4000rpm for 5minutes, the supernatant was discarded and the bacterial pellet was re-suspended in the leftover media and plated in the appropriate LB-antibiotic plates, which were grown overnight at 37˚C.

The day after, colonies were picked and grown in a 5 ml LB- antibiotic pre-culture, overnight at 37˚C.

The following day, bacteria were harvested and minipreped as described above. The integrity of rescued DNA was confirmed by restriction digestion and compared to the digested parental vector. In addition, PCR amplification of the vector's expression cassette and S/MAR motif was performed in order to confirm their unaltered presence in the vector. Correct rescued clones were confirmed by sequencing.

4.2.11.2 Plasmid Rescue Protocol 2

100 µl of Total gDNA extracted with DNeasy Blood and Tissue kit (Qiagen) were digested using a restriction enzyme that would digest the genomic DNA but not the vector to be rescued, usually EcoRV-HF (NEB). The amount (units) of enzyme was adjusted depending on the gDNA extraction yield, considering that 10 U of enzyme are required to digest 1 µg of DNA in 1h at 37˚C. The digestion volume was adjusted to 150 µl and the gDNA was digested overnight at 37˚C, shaking. The next day, total gDNA (digested) containing intact plasmids were extracted using phenol: chloroform. For that, one volume (150 µl) of Phenol:Chloroform were added to the restriction reaction, followed by vigorous vortexing and a centrifugation step for 10 minutes at 4˚C and maximum speed. Then, the aqueous phase was recovered and 1/10 volume of Sodium Acetate as well as 3 volumes of 100% Ethanol were added. The DNA was allowed to precipitate at -80˚C for 30 minutes or overnight at -20˚C. After precipitation, the DNA was pelleted by centrifugation at 4˚C for 30 minutes at maximum speed. The DNA pellet was washed with 70% Ethanol and centrifuged again for additional 10 minutes. Finally, the ethanol was removed, the DNA pellet was allowed to air-dry and it was resuspended in 10 µl of TE buffer. The DNA concentration was measured using spectrophotometry (Nanodrop).

Afterwards, between 400 ng and 1 µg of digested gDNA were electroporated into *E.coli* DH10β. Briefly, an aliquot of 20 µl of competent cells was thaw on ice. An electroporation cuvette with 0,1 cm gap was placed on ice and the DNA was aliquoted in a fresh tube, also placed on ice. Once the cells were thaw, they were added to the DNA and the mix was transferred to the cooled electroporation cuvette, which was quickly electroporated with the BioRad micropulser's 'Bacterial program'. Immediately after, 950 µl of SOC media were added to the cells and they were transferred into a 15 ml Falcon tube, where they were incubated for 1 - 2h at 37˚C, shaking. Past the recovery time, the cells were harvested at 4000rpm for 5minutes, the supernatant was discarded and the bacterial pellet was re-suspended in the leftover media and plated in the appropriate LB-antibiotic plates, which were grown overnight at 37˚C.

The day after, colonies were picked and grown in a 5 ml LB-antibiotic pre-culture, overnight at 37˚C. The following day, bacteria were harvested and miniprepped as described above. The integrity of rescued DNA was confirmed by restriction digestion and compared to the digested parental vector. In addition, PCR amplification of the vector's expression cassette and S/MAR motif was performed in order to confirm their unaltered presence in the vector. Correct rescued clones were confirmed by sequencing.

4.2.12 Southern Blotting

4.2.12.1 DNA Probe synthesis

DNA probes used for southern blot analysis were radioactively labelled with [α32P]dATP (3000 Ci/mmol) or [α32P]dCTP (3000 Ci/mmol) using Prime-it II Random Primer Labeling Kit (Agilent) and following the manufacturer's instructions.

For that, the DNA fragment to be served as probe, usually against the gene corresponding to coGFP or Puromycin, was digested with an appropriate restriction enzyme or amplified by PCR and isolated via separation in a 1% agarose gel. The DNA was recovered using a GeneElute Kit (Sigma), following the manufacturer's instructions. The concentration of the purified DNA fragment was assessed using spectrophotometry (Nanodrop). Then 25 ng of DNA fragment or control DNA were mixed with 10µl random oligonucleotide primers, in a total volume of 34 µl. Then, the reaction tubes were boiled in a water bath for 5 minutes and centrifuged briefly at room temperature. After, the following reagents were added to the reaction tubes: 10 µl of 5x dCTP or dATP Primer Buffer, 5 µl of labelled nucleotide [α32P]dATP (3000 Ci/mmol) or [α32P]dCTP (3000 Ci/mmol) and 1 µl Exo(-) Klenow

Enzyme (5U/µl). The content was mixed thoroughly by pipetting and incubated at 37˚C for 10 minutes. The reaction was then stopped upon addition of 2µl of Stop mix. Finally, the probe was used for hybridisation.

4.2.12.2 Restriction and gel electrophoresis

100 µl (7-12 µg) of gDNA extracted with DNeasy Blood and Tissue kit (Qiagen) were used to perform Southern Blot experiments. For that, the total gDNA including the circular vector was digested with an enzyme that would linearise the vector, as well as digest the gDNA (usually BamHI linearises vectors with pSMARt backbone). The reaction volume was adjusted to 50 µl and the reaction took at 37˚C for at least 4h at 37˚C, shaking. In addition, 2 – 5 ng of parental plasmids were also digested as positive controls. Past the digestion time, the digested gDNA and linearised vector were loaded in a 0,8% Agarose gel, which was run overnight at 15 - 25V. After imaging the gel to check for correct gDNA digestion by presence of a smear, the gel was immersed in Depurination buffer (250mM HCl) for 10 minutes, in gentle agitation, followed by two washed with milliQ water. Then, the gel was placed in Denaturation Buffer (1.5M NaCl and 500mM NaOH) and incubated twice for 15 minutes, in agitation. Then, the gel was Neutralised (1.5M NaCl, 0.5M Tri-sodium citrate, pH 7.0) and incubated twice for 15 minutes, in agitation. Immediately after, the gel was equilibrated twice in 20xSSC Buffer, for 10 minutes.

4.2.12.3 Transfer into membrane

Meanwhile, a Hybond-XL nylon membrane (Amersham Biosciences) was first soaked in milliQ water and then pre-equilibrated in 10x SSC buffer. In order to transfer the ssDNA into the membrane, a Transfer apparatus was set up as depicted in **Figure 18.** The gel was supported on a layer of Whatmann paper, which was soaked in a tank containing 10xSSC buffer. The nylon membrane was placed on top of the gel and the air bubbles were removed by carefully rolling a glass pipette. Then, a couple of layers of Whatmann paper (soaked in 10xSSC buffer) were placed on top of the membrane and the air bubbles were also removed. In order to promote the linear movement of water by capillarity, the remaining unused area of the plate was covered in parafilm. Finally, a stack of paper towels was positioned and a weight was balanced on top. The transfer occurred overnight.

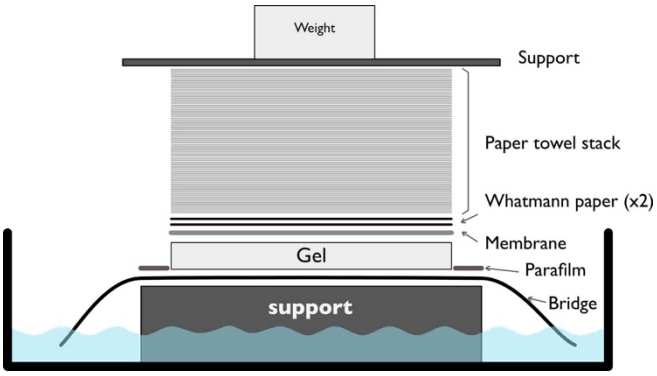


Figure 18: Representation of a Southern Blot apparatus

The next day, the apparatus was disassembled and the membrane was exposed to UV-B radiation in order to covalently and permanently cross-link the DNA to the membrane, which could then be stored in a sealed plastic bag at 4°C or hybridised.

The membrane was then pre-hybridised in Church Buffer (7% SDS, 0.5M NaPi, 1mM EDTA, 1% BSA) for at least 30 minutes at 65°C in a rolling glass bottle.

Meanwhile, 100 µl of radioactively-labelled DNA probe was denatured at 95°C for 10 minutes, chilled on ice and immediately transferred into an ice bath for 5 minutes. Afterwards, the DNA probe was diluted in 1 ml of Church buffer, mixed and added to the rolling membrane. The hybridisation was carried out overnight at 65°C.

The next day, the hybridisation solution was removed, and the membrane was washed four times. First, two washes with 100 ml Wash buffer 1 (2xSSC, 0.1% SDS), first for 5 minutes and then for 15 min at 65°C. Then, another two washes with 100 ml of Wash buffer 2 (0.5xSSC, 0.1% SDS), for 15 minutes at 65°C.

Finally, the membrane was placed in a plastic bag and developed for short exposures (3-4h) to overnight or up to a week.

4.3 Biochemical methods

4.3.1 Bicinchoninic acid (BCA) Assay

To prepare the samples for Western Blotting, whole cell lysates from cells were obtained as described in **Section 4.2.7**. Briefly, cells growing in a 6-well plate were washed, trypsinised and lysed in 40 µl of RIPA buffer (for 1 well of a 6 well plate).

After lysing the cells, the total amount of proteins was determined using the Pierce BCA Protein Assay kit and following manufacturer's instructions. Briefly, this colourimetric assay measured the total protein concentration in a given sample as compared to a known protein standard (BSA) by the formation of a purple-coloured product, which had an absorbance at 562 nm.

4.3.2 Sodium Dodecyl Sulphate (SDS) Polyacrylamide Gel Electrophoresis (PAGE) and Western Blotting

After the total protein concentration was determined, 30 µg of total protein were mixed with 4x Laemmli Buffer, diluted in RIPA buffer to a final volume of 20 µl and denatured for 5 minutes at 95˚C. Then, pre-casted Mini-PROTEAN Gels 4 - 20% (Bio-Rad) were assembled into the electrophoresis chamber (Mini-PROTEAN Tetra Cell). The chamber was filled with 1x Running Buffer and the samples together with the pre-stained Protein Ladder (PageRuler Plus), were loaded into the acrylamide gel. The electrophoresis was performed at 100 – 120 V for 1 - 2h. For buffer composition, refer to **Section 3.6.**

After the SDS-PAGE gel was run and the proteins were separated by size, a dry transfer of proteins into a PVDF membrane was performed. For that, the gel was separated from the plastic plates, the stacking gel was cut off and the resolving gel was soaked in running buffer to facilitate its handling. Then, the gel was placed in between the commercial iBlot 2 Transfer Stacks, containing: filters, Whatman paper, the PVDF membrane and the electrodes. The dry transfer was performed at 20 V for 7 minutes using an iBlot 2 device (Invitrogen). The membrane was then blocked in blocking buffer (5% powdered milk in 1x TBS) for 1h at room temperature and incubated with the correspondent primary antibody (diluted in blocking buffer) overnight at 4˚C in rotation.

The day after, the membrane was washed three times with 1x TBST buffer (each wash lasted for 10 minutes, in shaking conditions) and incubated for 1h at room temperature with the appropriate secondary antibody coupled to HRP (diluted in blocking buffer). For primary and secondary antibody

specifications and dilutions, refer to **Section 3.10.** After the secondary antibody incubation, the membrane was again washed three times with 1x TBST and finally developed using Chemiluminescence (SignalFire™ ECL Reagent) and exposed using a FusionSL machine.

4.3.3 Immunostainings and microscopy

4.3.3.1 Immunohistochemistry

Immunohistochemistry stainings for coGFP were performed in collaboration with Vanessa Vogel and Ornella Kossi (Hi-STEM, DKFZ).

Tissues were fixed, deparaffinised and sectioned into slices by Andrea Pohl-Arnold (DKFZ, W420) as described in **Section 4.6.4.** Some slides were stained using Haematoxylin and Eosin to identify and visualise the tissue architecture while others were subjected to immunohistochemistry staining for GFP (AB290, Abcam) using biotinylated secondary antibodies and a Peroxidase/AEC detection system (DAKO Real Detection system, K5003).

First, the tissue sections were deparaffinised in Xylol (twice for 10 minutes each time and an extra rinse in Xylol). Then, the tissues were dehydrated in decreasing Ethanol concentrations: two times in 100% Ethanol for 5 minutes each, followed by a 5-minute bath in 96% ethanol, a rinse with the same ethanol concentration and a final dehydration step in 70% ethanol for 5 minutes. The slides were rinsed in distilled water before retrieving the epitope.

Second, a damp heat-induced epitope retrieving step was performed to unmask and expose the antigens. This was achieved by boiling the samples for 15 minutes in a steam pot with citrate buffer at pH 6.0, followed by a 30-minute cooling-off period in the same buffer. The slides were rinsed with distilled water.

Third, the tissues were subjected to blocking with Avidin/Biotin (Avidin/Biotin blocking kit, SP-2001, Vector Laboratories). For that, the slides were rinsed in 1xPBS-T, followed by a 10-minute incubation at room temperature in Avidin Block. Then, the slides were rinsed again in PBST and incubated with Biotin block for another 10 minutes at room temperature. A final rinse in PBST was performed to remove residual blocking solution.

Forth, the samples were incubated with the primary antibody Rabbit anti-GFP (AB290, Abcam), diluted 1:500 in Dako Real antibody diluent (Dako, S2022). For that, 200µl of antibody dilution was added per slide and incubated for 30 minutes at room temperature. Then, the slides were rinsed in PBST.

Fifth, secondary antibody incubation was performed by adding 200µl per slide of multilink secondary antibody (Goat anti-Rabbit) provided in the Dako Peroxidase/AEC detection system (DAKO Real Detection system, K5003). The secondary antibody was incubated for 20 minutes at room temperature, and the slides were then washed in PBST.

Sixth, the activity of endogenous tissue peroxidase was depleted by adding 200µl of Dako Real Peroxidase blocking solution (Dako, S2023) and incubating for 5 minutes at room temperature, followed by a rinse in PBST.

Seventh, 200µl of streptavidin-peroxidase (HRP) was added and incubated for 20 minutes at Room Temperature, followed by a rinse in PBST.

Eighth, 200µl of AEC (reddish) chromogen was added to the specimen, and the samples were incubated and monitored under a microscope until a red product appeared (approximately 5 minutes). The reaction was stopped by rinsing the slides twice in PBST and finally in distilled water.

Finally, the tissue sections were counterstained for 1 minute in Haematoxylin, rinsed in tap water and mounted in Aquatex mounting media.

4.3.3.2 Alkaline Phosphatase staining

Alkaline Phosphatase stainings were performed in either black µ-Plate 96 Well (Ibidi) coated with gelatin (together with the immunostainings to detect pluripotency markers) or in gelatin-coated regular 24 well plates. The staining was performed using Alkaline Phosphatase Staining Kit II (Stemgent) and following the manufacturer's instructions. Briefly, the cells were washed once with PBS with 0.05% Tween20. Then, the cells were fixed with the fixative solution provided with the kit for 2-5 minutes and washed again with PBST afterwards. Finally, a staining mix containing equal volumes of solution A, B and C were added to the cells, which were incubated for 5-15 minutes in the dark. When the cells turned purple, the reaction was stopped by washing with PBS. The images were taken with a Keyence microscope (Keyence).

4.3.3.3 Immunofluorescence stainings

4.3.3.3.1 Neuroblastoma cells

Immunostainings to detect neuronal markers (GAP43, NCAM, β3-tubulin) were performed in a black µ-Plate 24 Well (Ibidi). Neuroblastoma cells were seeded in the staining plate, and they were treated with 10µM ATRA or DMSO and differentiated into a neuron-like phenotype, as explained in **Section 4.1.8.** After 10-12 days, the cells exhibited a neuron-like morphology, with elongated neurites and extensions. At this point, the cells were washed twice with PBS, fixed with 4% Paraformaldehyde (4% PFA) for 30 minutes at room temperature and washed again twice with PBS. Then, the cells were permeabilised using 0,2% Triton X-100 in PBS during 10 minutes at room temperature, followed by another two washes with PBS. At this point, the plate could be stored at 4°C overnight.

The day after, the cells were blocked with Blocking solution (2% FCS in PBS) for 1h at room temperature and incubated with the primary antibody (diluted in blocking solution) for another 1h at room temperature. Then, the cells were washed five times with PBS and incubated with the secondary antibody conjugated with Alexa-Fluor dyes and 2 µg/ml DAPI as a nuclear counterstaining for 1h at room temperature and protected from the light. Afterwards, five washes with PBS followed by another five washes with 100% Ethanol were performed. Finally, the wells were covered with PBS and the plate was either stored at 4°C (in the dark) or imaged using a Nikon Eclipse Ti/X-Cite120Led microscope. For primary and secondary antibody specifications and dilutions, refer to **Section 3.10.**

4.3.3.3.2 Stem Cells

Immunostainings to detect pluripotency or differentiation markers were performed on Stem Cells cultured on black µ-Plate 96 Well, glass bottom (Ibidi). Three days before staining, the cells were washed, trypsinised and sedimented as described before and 10-20µl of a confluent 6-well plate were transferred per well of a gelatinised and feeder containing µ-Plate 96. Three days after, the cells were washed twice with cold PBS and fixed with 100% Methanol for 7 minutes at -20°C followed by a rinse with acetone for 20 seconds also at -20°C. When plastic plates were used, the acetone-rinsing step was avoided to prevent the plastic from degrading. Then, the cells were permeabilised with 0,1% Tween20 in PBS (PBST) for 5 minutes at room temperature and washed three times with cold PBST for 5 minutes at room temperature. At this point, the plate could be stored at 4°C overnight or blocked to proceed with the staining. The next day, the cells were blocked with Blocking solution (1% FCS, 0,5% BSA, 0,1%TritonX100 in PBS) for 30 minutes at room temperature and incubated with the primary antibody

(Murine SCs: *Oct4, Nanog, SSEA1*; Human SCs: *Tra160, SSEA3, SSEA4 and Nanog*) diluted in blocking solution, overnight at 4°C. The day after, the cells were washed three times with blocking solution and then incubated with the secondary antibody conjugated with Alexa-Fluor dyes and 2 µg/ml DAPI as a nuclear counterstaining for 1h at room temperature and protected from the light.

After that, the cells were washed with cold PBST followed by two other washes with cold PBS and the wells were covered with PBS to prevent drying of the samples. Finally, the plate was either stored at 4°C (protected from the light) or imaged using a Nikon Eclipse Ti/X-Cite120Led microscope. For primary and secondary antibody specifications and dilutions, refer to **Section 3.10.**

4.3.3.3.3 Embryonic Bodies

Immunostainings to detect differentiation markers corresponding to the 3 germ-layers (β3-Tubulin for ectoderm, α-Smooth Muscle Actin for mesoderm and FoxA2 for endoderm) were performed on EBs cultured on black µ-Plate 24 Well (Ibidi) as described on **Section 4.1.6** and following the same protocol as described above in **Section 4.3.3.3.2.**

The EBs were washed twice with cold PBS and fixed with 100% Methanol for 7 minutes at -20°C followed by a rinse with acetone for 20 seconds also at -20°C. Then, they were permeabilised with 0,1% Tween20 in PBS (PBST) for 5 minutes at room temperature and washed three times with cold PBST for 5 minutes at room temperature. At this point, the plate could be stored at 4°C overnight or blocked to proceed with the staining. The next day, the cells were blocked with Blocking solution (1% FCS, 0,5% BSA, 0,1%TritonX100 in PBS) for 30 minutes at room temperature and incubated with the primary antibodies (β3-Tubulin, αSMA and FoxA2) diluted in blocking solution, overnight at 4°C. The day after, the cells were washed three times with blocking solution and then incubated with the secondary antibody conjugated with Alexa-Fluor dyes and 2 µg/ml DAPI as a nuclear counterstaining for 1h at room temperature and protected from the light.

After that, the cells were washed with cold PBST followed by two other washes with cold PBS and the wells were covered with PBS to prevent drying of the samples. Finally, the plate was either stored at 4°C (protected from the light) or imaged using a Nikon Eclipse Ti/X-Cite120Led microscope.

4.4 Flow cytometry

4.4.1 Quantification of transfection efficiency and vector performance

Flow cytometry was used to determine transfection efficiency of S/MAR vectors as well as to monitor their GFP expression over time. To prepare the cells for analysis, the media was removed, and the cells were washed twice in PBS, trypsinised for 5 minutes and re-suspended in FCS containing media. Then, the cells were harvested at 1000 - 1500rpm for 5 minutes, washed, re-suspended in PBS containing Propidium Iodide (1:100) and filtered through a 15 ml filtered BD FACS tube, placed on ice to prevent cell clumping.

4.4.2 Hematopoietic Stainings

The analysis of hematopoietic tissues was performed in collaboration with Sina Stäble and Marleen Büchler, from Dr Milsom Lab (Hi-STEM/DKFZ).

4.4.2.1 Blood staining

The animals were anaesthetised and the blood was collected by terminal bleeding in EDTA-tubes.

Then, 30 µl of each blood sample was transferred into a tube (in duplicates) to stain for a panel of hematopoietic cell surface markers or check for GFP expression.

The unstained blood cells were diluted in 70 µl of 2% FCS/PBS. For the panel staining, a master mix containing CD45.1, CD45.1, CD11b, B220, CD4 and CD8a antibodies was prepared in 2% FCS/PBS and 70 µl were added to the blood sample and incubated for 30 minutes at 4°C.

Table 46: Blood staining panel for FACS analysis

CD45	Pan blood surface marker
CD11b	Macrophage surface marker
B220	B cells surface marker
CD4	T cells surface marker
CD8a	T cells surface marker

Then, 1 ml of ACK Lysis buffer was added to each sample and incubated for 10 minutes at room temperature with occasional shaking, to lyse the red blood cells. Then, the cells were pelleted at 845g for 5 minutes at 4°C, washed with 1 ml of 2% FCS/PBS and re-suspended in 200 µl of 7AAD/2%FCS/PBS (5µl/ml, 1:200 dilution). Finally, the cells were filtered through polystyrene FACS tubes and analysed using an LSRII cytometer.

4.4.2.2 Spleen Staining

After sacrificing the animals, the spleens were placed in IMDM media, kept at 4°C and processed as soon as possible. To disrupt the tissue, the spleen was smashed through a filter (40 µm EASYstrainer, Greiner bio-one) and collected in 2% FCS/PBS. The cells were then stained with CD45-Pacific Blue antibody for 30 minutes at 4°C, and the red blood cells were lysed as described above. The splenocytes were finally pelleted, washed, filtered, and FACS analysed using an LSRII cytometer.

4.4.2.3 Bone Marrow staining

After sacrificing the animals, the femur and/or tibia from back legs were placed in IMDM media, kept at 4°C and processed as soon as possible. The bones were then cleaned from muscle, and the bone marrow was flushed out in 2 ml PBS/FCS with the help of a syringe. The bone marrow cells were then stained with CD45-Pacific Blue antibody for 30 minutes at 4°C, and the red blood cells were lysed as described above. The cells were finally pelleted, washed, filtered, and FACS analysed using an LSRII cytometer.

4.4.3 Data acquisition

The data was acquired with the BD FACS Fortessa™ or LSRII analysers™ analysers (BD Biosciences) and the FACSDiva (BD Biosciences) acquisition software.

For vector transfections, population distributions were analysed by gating the single cell population (SSC-A vs SSC-H). Then, the viable cell population was gated on single cells by plotting the PE-Texas red-A vs side scatter. GFP fluorescence intensity (Alexa Fluor 488-A) was gated on the viable population of GFP fluorescence intensity (Alexa Fluor 488-A). For hematopoietic stainings, CD45-positive cells were gated on living cells (7-AAD-negative population), and GFP was gated on CD45-positive cells.

4.4.4 Data analysis

The following parameters: Number of events, % Gated, % Total, Mean, Geometric Mean, SD, CV and Median, were analysed per gated population. Usually, a histogram of GFP-positive cells recorded at different timepoints served as a measure for vector transfections and performance.

For hematopoietic stainings, a histogram of GFP-positive cells was used as a measure of chimerism and degree of transgenesis. For hematopoietic differentiation experiments, a histogram of GFP-positive cells was used to assess survival of GFP-expressing vectors during differentiation.

4.5 DNA Sanger sequencing

Sequencing of DNA plasmids was performed by GATC Biotech AG. For this, 5 µl of >100 ng/µl plasmid DNA was mixed with 5 µl of primer at a final concentration of 5 µM.

4.6 Animal experiments

4.6.1 Subcutaneous Neuroblastoma injections

Animal experiments involving the subcutaneous injection of human Neuroblastoma cells were performed in collaboration with Dr MD Jeannine Desiree Lacroix and Domenic Hartmann, from the Pediatric haematology and Oncology department of the Heidelberg University Clinic and the Tumor Virology department (F010) from DKFZ. Domenic Hartmann performed the subcutaneous injections.

A total of 25 SCID mice were injected subcutaneously and distributed in groups as represented in **Table 47.** Of these 25 mice, ten were injected with Be2C-GFP (labelled with vector 36): five of which were injected with 10^6 cells and the other five were injected with $5x10^6$ million cells. The other mice were injected with Be2C-Luciferase (labelled with vector 50): five of which were injected with 10^6 cells and the other five were injected with $5x10^6$ million cells. Also, five other mice were injected with 10^6 parental Be2C (unlabelled) cells.

Table 47: Neuroblastoma subcutaneous injections

Number of mice			Cell line	Amount (cells)
25	10	5	Be2C-GFP	$1x10^6$
		5	Be2C-GFP	$5x10^6$
	10	5	Be2C-Luciferase	$1x10^6$
		5	Be2C-Luciferase	$5x10^6$
	5	5	Be2C (parental)	$1x10^6$

To prepare the cells, Be2C Neuroblastoma cells labelled with GFP (vector 36) or Luciferase (vector 50) were cultured into three T125 flasks (per cell line). When the cells reached 80% confluency, they were washed once with cold PBS+, followed by another wash with ice-cold PBS- and trypsinised for 5 minutes at 37°C. After that, trypsin was inactivated upon addition of PBS- containing 10% FCS and the cells were re-suspended vigorously to disrupt clumps. They were then washed twice with cold PBS+ and finally re-suspended in the same buffer. Afterwards, 10^6 and $5x10^6$ cells (per injection) were counted for Be2C-GFP and Be2C-Luciferase labelled cells, and 10^6 cells (per injection) were counted for Be2C parental cells. For each injection, such amount of cells was then re-suspended in 120 µl of ice-cold PBS.

Also, an extra tube per injection type was prepared to have a cell backup in case of unexpected events during the injections, such as clumping or clothing of the needles.

Before injecting, the cells were mixed with Matrigel and subsequently injected subcutaneously into severe combined immune deficient (SCID) mice. Cell engraftment was assessed by the formation of tumours and the monitoring their growth over time until they reached a critical diameter of 15 mm and the animals had to be sacrificed.

For each time point, all injected mice were checked for the presence of tumours and, if present, their length, height and width were measured using a calliper. Afterwards, the tumour volume (mm^3) was calculated as follows:

$$tumor\ volume = \frac{(L \times W \times H) \times \pi}{3}$$

It should be taken into consideration that the actual tumour volume might be overestimated when using this formula due to the inability to properly address the infiltration of inflammatory tissues, infiltration of blood, necrosis or calcification.

Once the tumours reached 10% of the animal's body weight or exceeded 15 mm in diameter, the animals were sacrificed by cervical dislocation Thereafter, the animals were dissected and each tumour was removed and cut in three pieces: one piece was fixed in 4% PFA overnight and paraffin embedded to perform immunohistochemistry stainings, another piece was snap frozen in liquid nitrogen and the last piece was trypsinised, mechanically homogenised and cultured in Be2C media. The cultured primary cells from the tumours were finally checked for GFP or Luciferase expression.

4.6.2 Stereotactic Neuroblastoma injections

Animal experiments involving stereotactic injection of Neuroblastoma cells, were performed in collaboration with Prof. Dr Ana Martin-Villalba and Sascha Dehler, from the Molecular Neurobiology Department (A290) from DKFZ. Sascha Dehler performed the stereotactic injections.

For intracranial injections of S/MAR-labelled Be2C cells, 8-week old female SCID BEIGE (CB17.Cg-PrkdcscidLystbg-J/Crl) mice from Charles River were used. A total of 7 females were injected intracraneously in the striatum (AP=0 /ML=2.5/DV=3.5) via stereotactic surgery. Before injection, GFP or Luciferase S/MAR cells were washed with PBS – (without Calcium and Magnesium),

trypsinised and resuspended in FCS containing RPMI media to inactivate the trypsin. Then, the cells were centrifuged at 200g for 5 minutes and washed twice in PBS – containing 1 mM EDTA and 25 mM HEPES in order to prevent cell clumping. Afterwards, $1{,}25x10^6$ cells were counted, aliquoted in PBS (+ 1 mM EDTA, 25 mM HEPES) and placed on ice during transportation. While preparing the stereotactic apparatus and the mice, the cells were kept in the water bath at 37°C and were only centrifuged and resuspended in 3 µl PBS- right before injection.

Injected mice were checked regularly for behavioural changes, such as impaired motility or balance. Upon appearance of visual tumour signs, the mice were sacrificed and the brains were removed and analysed via Luciferase Assay. For this, 150 mg Luciferin/kg of weight (10 µl/g from a 15 mg/ml stock, filter sterilised) were injected intraperitoneally 10 - 15 minutes before sacrificing of the animals, to allow the Luciferin to reach the brain tissue. After, the animals were administrated with an overdose of narcotics and the brains were removed and placed in unsupplemented RPMI, containing 10 µl/ml of Luciferin. The brains were quickly imaged with a 10 minutes' exposure time using Fusion SL - Chemiluminescence Imaging (Analis). After imaging the tumours, the brains were fixed in 4% PFA overnight at 4°C. The day after, the PFA was replaced by PBS and the brains were fixed and cut into 50µm coronal sections by using a Leica VT1200 Vibratome. The brain slices were rinsed in PBS and were subsequently mounted with Fluoromount-G/DAPI (eBioscience) on glass slides, which were imaged using a Nikon Eclipse Ti/X-Cite120Led Microscope.

4.6.3 Generation of Transgenic mice

4.6.3.1 Indirect approach: Stem Cell injections into Morulae/Blastocyst

Animal experiments involving the generation of Chimeric mice were performed in collaboration with Franciscus A. van der Hoeven and Ulrich Kloz, from the Transgenic Service (W450) from DKFZ.

In the first chimaera generation attempt, which was performed under the experimental License number: G-97/12, CD1 albino mice were mated, and embryos at a *morula* stage (E3.0) or late blastocyst (E4.5) were collected after surgery.

Black 6 Mouse embryonic stem cells (mESC BL6) labelled with pSMARt_GFP (vector 71) were cultured, washed and trypsinised as described above. Then, *Zona Pellucida* was laser ablated, and 6 - 12 mESC were injected into either *morulae* or late blastocysts. The day after injection, the embryos were transferred into pseudo-pregnant females that were previously mated with sterile males. The embryos were brought to term, and the pups were checked for coat chimerism as well as for GFP expression as depicted in **Figure 19.**

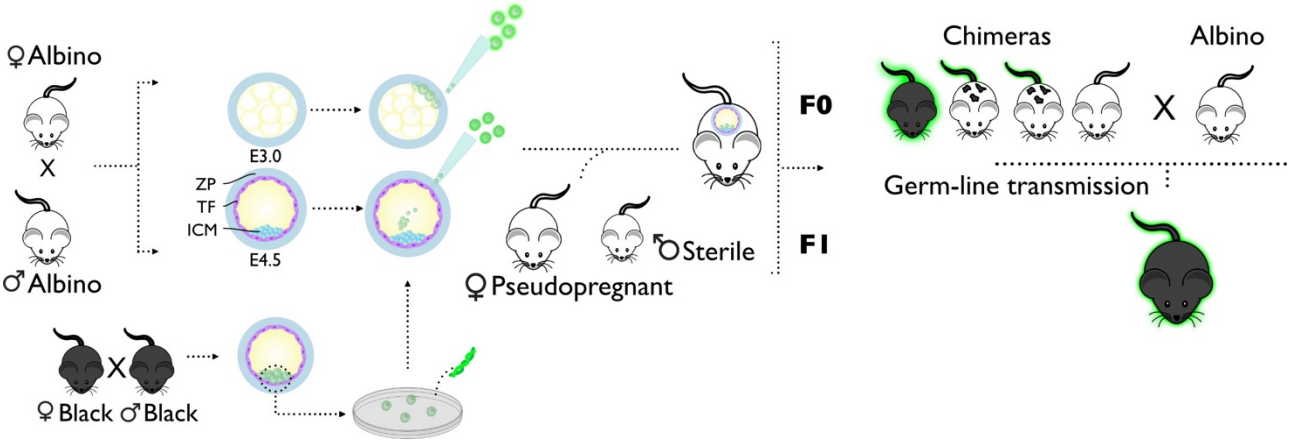


Figure 19: First attempt to generate chimeric mice via morulae/blastocyst injections of BL6 mESC labelled with pSMARt vectors.

The second attempt to generate chimeric mice was performed under the License Number G-148/13. This time, 129Ola E14-1 Mouse Embryonic Stem Cells (E14) were labelled either with pSMARt_GFP (pCAG), vector 71; or nSMARt_GFP (nCAG) vector 85. Transfected cells were cultured as described and between 6 - 12 stem cells were injected into C57BL/6N x B6D2F1 embryos as described above, following the strategy shown in **Figure 20.**

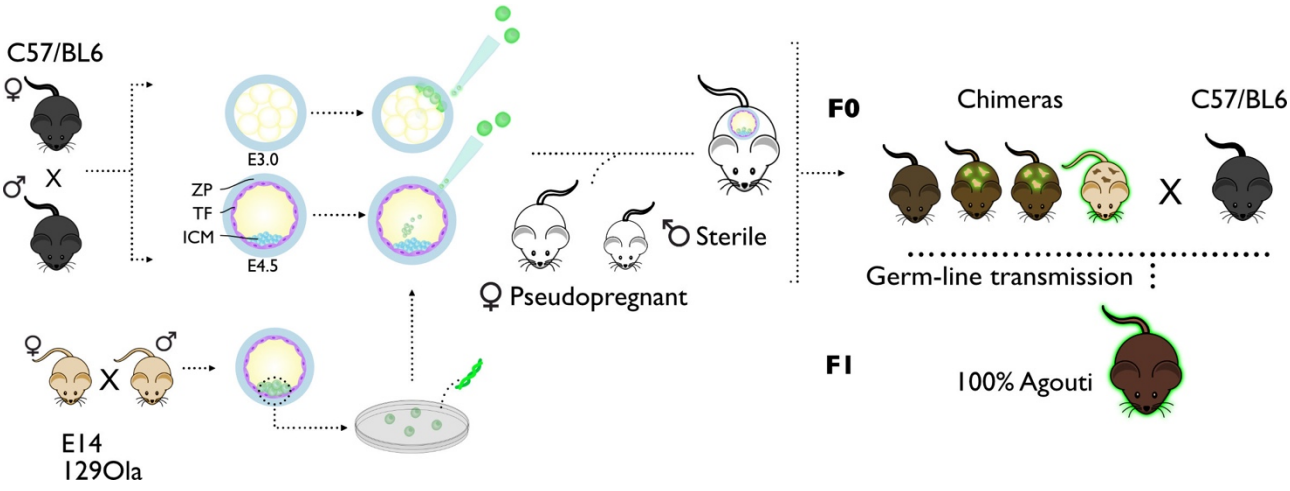


Figure 20: Second attempt to generate chimeric mice via morulae/blastocyst injections of E14 mESC labelled with pCAG and nCAG vectors.

4.6.3.2 Direct approach: DNA Pronuclear injections

Plasmid DNA pronuclear injections were performed in collaboration with Dr Franciscus van der Hoeven, from the Transgenic Service (W450) from DKFZ, under the License number: G-97/12.

Before injection, approximately 100 µl of plasmid DNA (vector 71) at a concentration 1000 ng/µl were filtered through a Millipore Millex-GV 0.22 µm by using a disposable 1ml syringe and with an air-bubble to fill up the rest of the volume. Then, 50 µl of filtered DNA was placed on top of a floating Millipore membrane VMWPO2500 (0.025 µm pore size) and desalted via dialysis in 50 - 100 ml of Dialysis Buffer (100 ml Ampuwa ddH_2O, 10 mM Tris and 0.1 mM EDTA) for 3 hours at 4°C to avoid evaporation and DNA loss. After 3h, the DNA drop was carefully recovered, and both quality and quantity of DNA were assessed via spectrophotometry (Nanodrop) and Agarose Gel electrophoresis. DNA recoveries ranging from 70 - 80% were achieved. The DNA solution was stored at -20°C or used for downstream applications.

The day of injection, 1 - 2 picoliters of plasmid DNA at a concentration of 1 - 3ng/µl were injected into the pronucleus of E0.5 murine C57BL/6N zygotes. After, the zygotes were transferred into a foster mother and brought to term as shown in **Figure 21.**

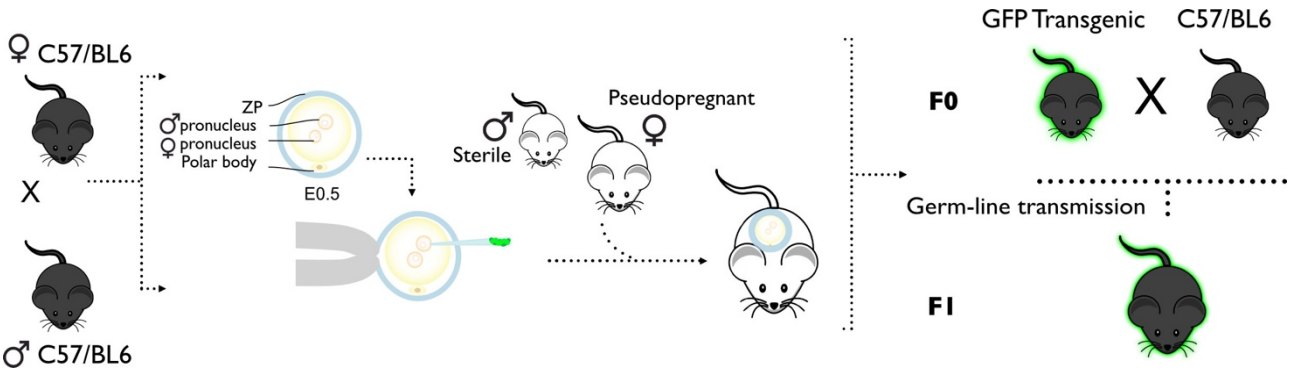


Figure 21: Generation of transgenic mice via pronuclear injection

4.6.4 Processing and analysis of murine tissues and samples

The processing and analysis of transgenic animals was performed in collaboration with Dr Karin Müller-Decker, Stephanie Laier and Andrea Pohl-Arnold, from the Tumor model department (W420) from DKFZ, under the killing number (VV number DKFZ345).

The mice were sacrificed by cervical dislocation or anaesthetised with Isofluorane and terminally bled. The blood was collected in EDTA-tubes and analysed via Flow Cytometry. The animals were then

dissected, and internal organs from different embryonic derivatives such as liver, kidney, heart, skeletal muscle, skin and testes were taken and cut into four pieces. One piece was fixed in 4% PFA and used for immunohistochemistry, another piece was used for fluorescent imaging, and two pieces were snap frozen for further molecular analysis.

For quantification of GFP, the spleen and bone marrow from the femurs of sacrificed animals were placed in IMDM media and stained with a panel of hematopoietic markers for FACS analysis as described in **Section 4.4.2**.

4.6.5 Sperm collection

The collection of sperm from transgenic animals was carried out by Heinrich Steinbauer and Andrea Rausch, from the cryopreservation service (DKFZ W430). The males were euthanised with CO2 inhalation and cervical dislocation. After an abdominal incision and a peritoneal cut, the testes and cauda epididymis were exposed, cut and separately placed in PBS. With the aid of a stereomicroscope and tweezers, the *cauda epididymis* was cut, and the sperm was released into the buffer.

For imaging, 10 µl of sperm in PBS were placed on a slide and covered with a coverslip. Then recordings of motility or fluorescence images were taken using a Nikon Ti microscope and a 20x objective. When the sperm could not be freshly imaged, it could be kept alive and motile at room temperature for up to 2 hours. The remaining sperm was pelleted and frozen for total DNA extraction and downstream applications. Modified protocol from [292].

4.6.6 Isolation of seminiferous epithelium

The protocol from the isolation of seminiferous epithelium and depletion of interstitial cells was adapted from [293]. Murine testes were placed in Digestion media (DMEM/F12 supplemented with 25 mM HEPES and 1% Penicillin-Streptomycin). Under a stereomicroscope and with the help of tweezers, the testes were immobilised, and the *Tunica albuginea* was perforated and removed, allowing the release of the seminiferous tubules, which were then mechanically separated with tweezers. Then, the tubules were subjected to sequential digestion. First, they were transferred into a 15 ml Falcon tube containing 5ml of Digestion media and 12,5 µl of Collagenase III; and incubated for 30 minutes at 37˚C with occasional shaking, which allowed the tubules to break into smaller pieces and release the interstitial cells. The tubules were allowed to settle at the bottom of the tube, and the supernatant containing the interstitial cells was removed. Then, they were washed with PBS and allowed to settle

by gravity. Second, the tubules were disaggregated, and the seminiferous epithelial cells were released during a 5-minute incubation with trypsin. Finally, the cells were filtered through a cell strainer to remove cell clumps and processed for downstream applications, such as Flow Cytometry or DNA extraction.

5. RESULTS

5.1 Vector development

Several studies demonstrated that S/MAR vectors were able to transfect all cell lines tested [126], including CD34+ HSC cells [165]. However, the use of S/MAR vectors for the genetic modification of pluripotent stem cells, including mESCs and iPSCs, has never been shown before.

In this chapter, the evolution and development of DNA episomal vectors; from the originally described pEPI vector to the latest improved nano plasmid versions, is described. Most of the vector developing work has been performed by Dr Matthias Bozza (DNA vector Lab) and has been performed in collaboration with Nature Technology Corporation, who generated the nano vector versions of our plasmids. Finally, the suitability of the newly generated vectors is evaluated in mouse embryonic stem cells (mESCs) and mouse embryonic fibroblasts (MEFs).

5.1.1 Evolution of S/MAR vectors

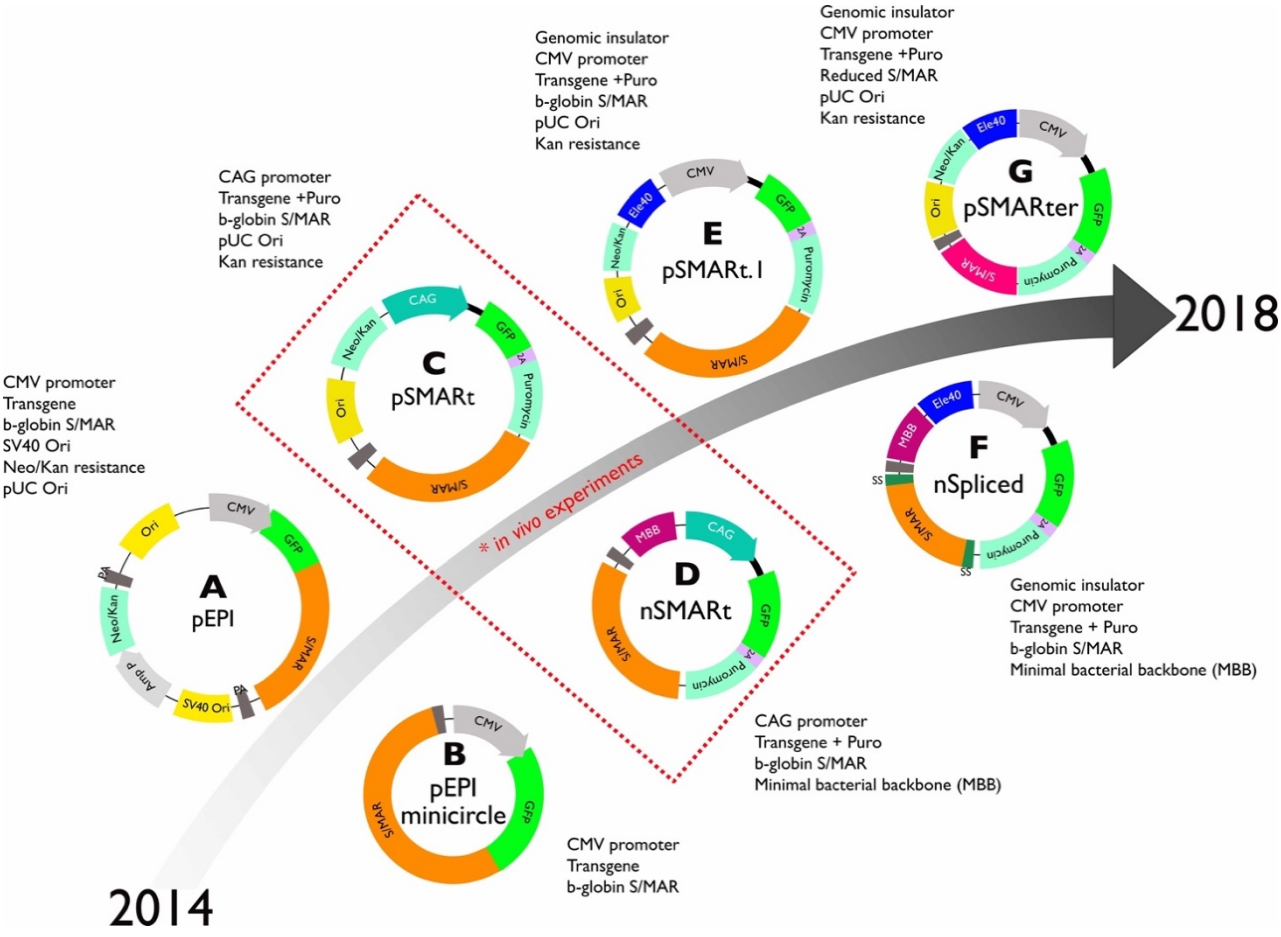


Figure 22: Cartoon depicting the evolution of S/MAR DNA vectors

The critical vector development points are listed below, represented in **Figure 22** and explained in detail in the coming sections.

1. Changing of the selection marker (Neomycin/G418 to Puromycin) and the configuration of the vector by forcing the selection marker to be an active part of the transcriptional unit resulted in increased levels of transgene expression (pSMARt generation, **Figure 22c)**

2. Addition of insulating elements, such as UCOE or Element40, enhanced expression and improved the vector establishment (pSMARt.1 generation, **Figure 22e)**

3. Reduction of the bacterial backbone resulted in enhanced expression and vector establishment (nSMARt generation, **Figure 22d).**

4. Replacement of the original S/MAR motif by a smaller one substantially increased the performance of the vector (pSMARter generation, **Figure 22g).**

5. Splicing the S/MAR sequence from the mRNA transcript resulted in better expression and establishment of the vector (nSpliced generation, **Figure 22f).**

5.1.2 Suitability of pEPI minicircle for stem cell work

The use of S/MAR vectors for the genetic modification of murine embryonic stem cells has not previously been shown. Therefore, we started by evaluating the suitability of the described pEPI minicircle vector for stem cell work.

In collaboration with Dr Michael Milsom's Lab (Hi-STEM/DKFZ), mESC were transfected with the pEPI minicircle, which contained an eGFP reporter gene driven by the CMV promoter and the S/MAR motif. Transfected mESC were picked and 18 mESC clones were expanded and evaluated for GFP expression via Flow Cytometry as shown in **Figure 23.** The number of GFP-positive clones obtained as well as the levels of GFP expression were not very promising. Only 7 out of 18 selected clones showed some degree of GFP expression. It is worth noticing the heterogeneity of GFP expression, especially in clones 3, 16 and 17.

Although the transfection efficiency, as well as the transgene expression levels, were low, this test experiment suggested that S/MAR vectors have the potential to transfect mESC and hence, the pEPI vector served as a starting point to improve the vector's features and performance.

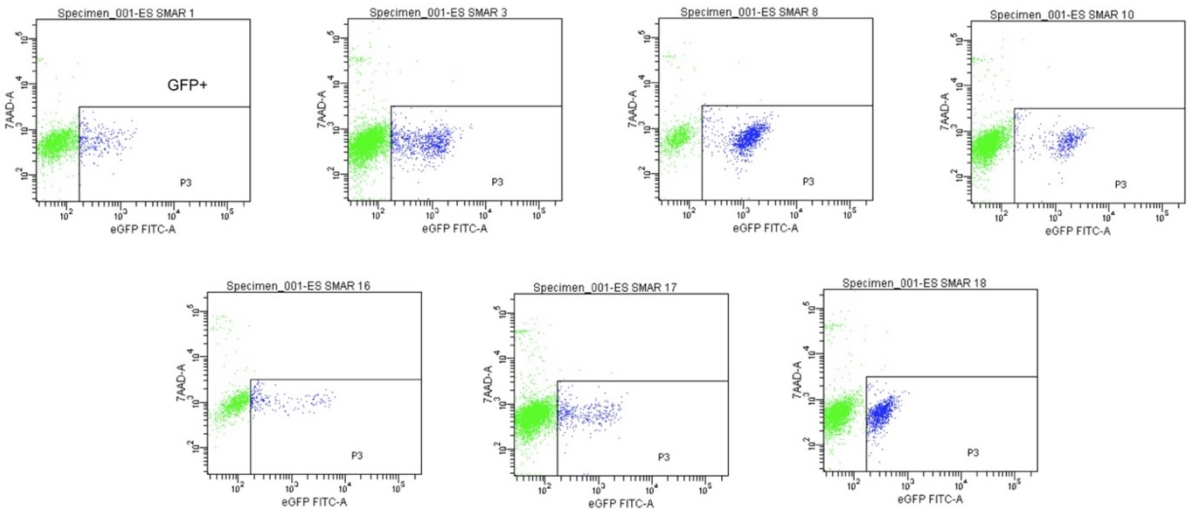


Figure 23: pEPI minicircles are unable to sustain homogeneous and high levels of transgene expression
The ability of S/MAR vectors to transfect murine embryonic stem cells (mESC) was evaluated by transfecting a pEPI minicircle, containing GFP After transfection, eighteen mESC clones were analysed for GFP expression using flow cytometry. Only seven clones heterogeneously expressed GFP. 7AAD was used as live-dead staining to define the viable cell population.

5.1.3 Rearranging the vector – the birth of pSMARt

5.1.3.1 Change and relocation of the selection marker

The original pEPI vector contains a dual Kanamycin/G418 selection cassette **(Figure 22a).** Establishing cell lines using G418 as selection marker was not only inefficient but a lengthy process, taking up to a month to obtain clones (data not shown). Also, spontaneously G418 resistant clones were commonly detected. This problem was overcome by changing and repositioning the selection marker. This was accomplished by directly coupling the gene of interest (GOI), in this case, GFP, to the selection marker Puromycin (instead of G418) and the S/MAR motif. This small change allowed to make the establishment of the vector part of an active process that can be directly controlled by the amount of selective pressure applied to the cells.

When neuroblastoma cells were transfected with S/MAR vectors with a CMV promoter driving expression of firefly Luciferase and G418, the cells showed low levels of Luciferase expression. Also, we observed the presence of Luciferase negative spontaneously resistant clones, which contributed to the dilution of the vector (and therefore luciferase) in the population cell population **(Figure 24a)**. When the selection marker was replaced by puromycin the reporter's gene expression was 5-fold higher as compared to cells selected with G418 **(Figure 24b).** Also, when the anti-repressive Element40 was used, the luciferase expression increased by almost 17-fold as compared to G418 selection and 3,4-fold as compared to puromycin marker **(Figure 24c).**

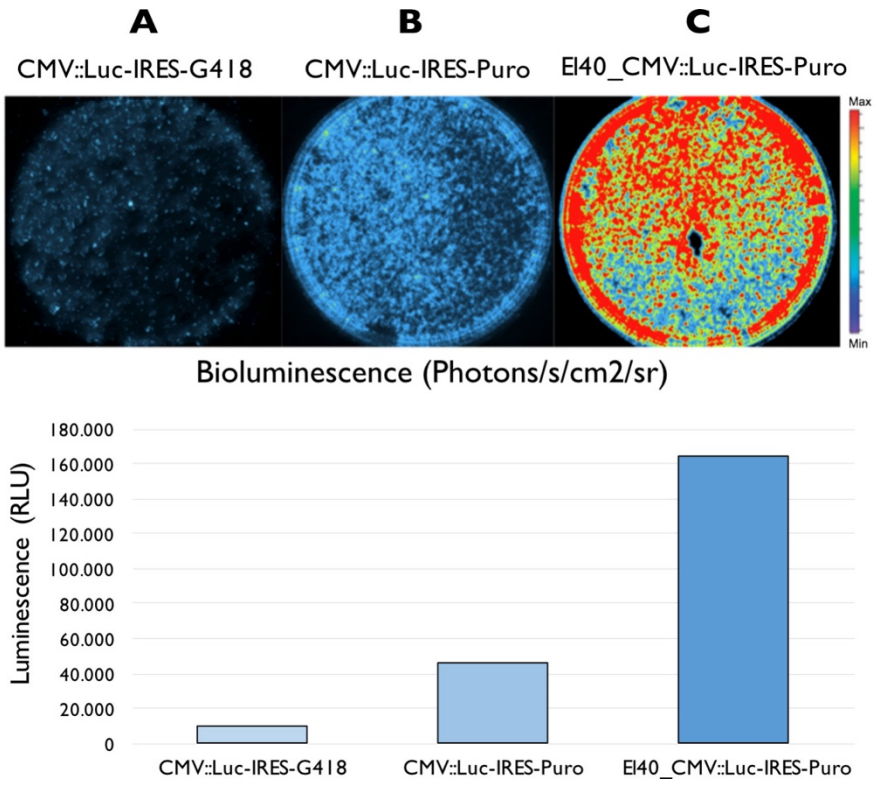


Figure 24: Replacing and repositioning the selection marker improves the transgene expression
Neuroblastoma cells were transfected with CMV::Luciferase constructs coupled to either G418 (vector 23) or Puromycin (vector 28) and selected for 2 weeks. Also, a third construct (vector 50) containing the chromatin shielding element (Element40) was also used. Afterwards, the Luciferase expression was measured using a chemiluminescent reader (FusionSL) and quantified with ImageJ. A notorious difference in Luciferase expression was achieved with minimal modification and change of selection marker and its position on the vector. The expression increased even further when insulating elements, such as Element40, were included.

5.1.3.2 Addition of chromosomal elements enhances the vector's expression

We then built a library of different vectors with different arrangements; including different selection markers (Puromycin or G418), different molecular linkers (IRES or 2A sequences) or different anti-repressive elements (UCOE or Element40), and compared them to each other and to the originally described pEPI vector. This vectors including shielding elements and based on pSMARt backbone were named pSMARt.1 generation **(Figure 22e).** The addition of the insulating Element40 resulted in increased transgene expression in Be2C cells **(Figure 24c)**.

Different S/MAR prototype vectors were transfected into HEK293T cells, which were then sorted. 100 GFP- positive cells were plated and kept under selection for four weeks. After this period, cell colonies were stained and counted to determine the efficiency of vector's establishment in a colony forming assay as shown in **Figure 25.**

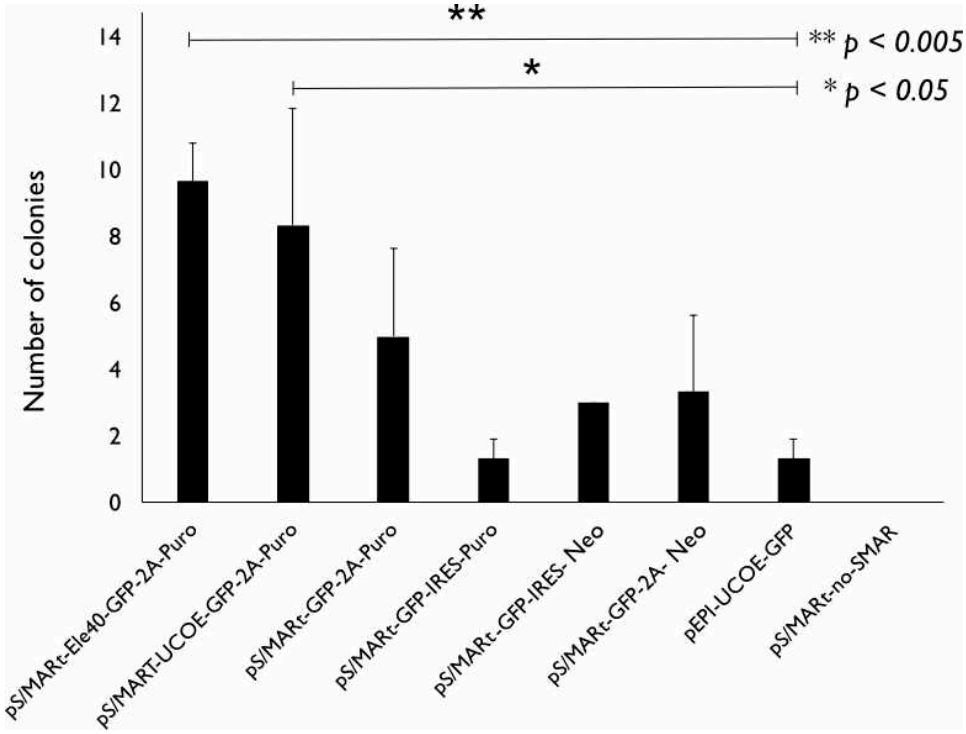


Figure 25: Evaluation of the establishment efficiency of different S/MAR vectors in a colony forming assay. Different versions of GFP-S/MAR vectors containing different chromosomal elements and selection markers in different combinations were tested in HEK293T cells. A non-S/MAR construct was used as negative control. After transfection, the GFP positive cells were FACS sorted, and 100 cells were plated and grown for 4 weeks in the presence of the respective antibiotic (0.5ug/ml Puromycin or 0,75mg/ml G418). After a month, the colonies were fixed, stained and quantified. The cells containing pSMARt vectors shielded with genomic elements (UCOE or Element40) generated a more substantial number of colonies as compared to their non-shielded counterparts and the original pEPI vector prototype. (The statistical comparison between vectors was performed using a T-test **p <0.005, * p< 0.05). *Credit: Matthias Bozza, DNA Vector Lab*

From this experiment, we observed that by introducing anti-repressive elements before the mammalian expression cassette, we could improve the vector establishment efficacy. The colony forming assay also revealed that by linking the reporter gene and the selection marker with the self-cleaving P2A peptide was more efficient than using internal ribosome entry site (IRES) sequences.

The use of a P2A sequence creates a direct link at the transcriptional and translational level between transgene and selection marker. This configuration was used for further studies.

Since the insulator sequences significantly improved the number of the resistant colonies, the anti-repressive Element 40 was selected for other applications. The UCOE element, although efficient, was dropped because its intellectual property is protected (Patent number: WO 2002024930 A2) [294].

All vectors containing the Puromycin selection coupled to the reporter gene via 2A self-cleaving peptides were named 'pSMARt' generation **(Figure 22c).**

5.1.4 Less is more – the birth of nSMARt

Some cells, especially primary cells, are refractory to transfection with plasmids carrying large conventional bacterial backbones including prokaryotic selection markers and pUC origins of replication. It has also been suggested that plasmids with large bacterial sequences and CpG islands are prone to silencing [295]. To overcome these limitations, we generated minimally sized DNA vectors devoid of bacterial backbone based on the Nanoplasmid™ technology in collaboration with Nature Technology Corporation (NTX) [296].

This novel class of vectors can be propagated and produced at high yield in an engineered *E.coli* strain, taking advantage of the antibiotic-free system developed by Nature Technology. This system is based on small interfering 70bp antisense RNA present in the plasmid, that binds to the chromosomally encoded constitutively expressed levansucrase (*sacB*) mRNA, which prevents its expression, allowing growth on sucrose media. In the absence of nanoplasmids, bacteria produce levansucrase, which results in cell death in the presence of sucrose.

5.1.4.1 nSMARt vectors show better establishment and transgene expression

'Nano' versions of the previously generated pSMARt vectors, which were named nano-SMARt (nSMARt) vectors (**Figure 22d),** were generated and compared to their pSMARt counterparts. The expression levels were not only higher (as shown in **Figure 30** to **Figure 32**) but also their establishment efficiency was significantly increased **Figure 26.** The colony forming assay showed the detrimental effect of the bacterial backbone in the plasmid expression and establishment.

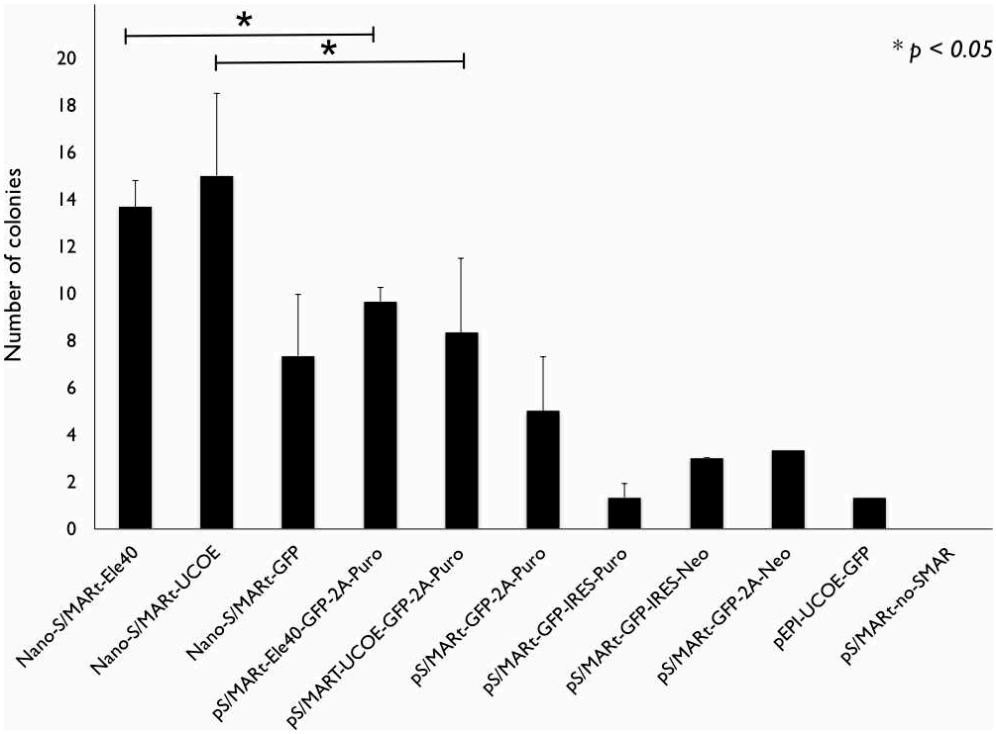


Figure 26: Evaluation of establishment efficiency of pSMARt and nSMARt vectors in a colony forming assay. Different versions of pSMARt vectors and their minimally sized nSMARt counterparts were transfected into HEK293T cells. A non-S/MAR construct was used as negative control. After transfection, the GFP positive cells were FACS sorted, and 100 cells were plated and grown for 4 weeks in the presence of the respective antibiotic (0.5ug/ml Puromycin or 0,75mg/ml G418). After a month, the colonies were fixed, stained and quantified. The vectors were more stable, and their efficiency was significantly increased when the bacterial sequences were removed from the vector backbone. The statistical comparison between vectors was performed using a T-test ($p < 0.05$) and the transfection was standardised to the number of DNA molecules. *Credit: Matthias Bozza, DNA Vector Lab.*

5.1.4.2 pSMARt and nSMARt vectors' effects on cell transcription

We then investigated the effects of pSMARt vectors on cell transcription. The human pancreatic adenocarcinoma cell line Capan-1 was used and stably transfected with the pSMARt_Luciferase vector. After selection, RNA was isolated, and a microarray was performed to unravel how many genes were dysregulated due to the presence of the S/MAR and the bacterial backbone. We found that only 1% of the cell's transcriptome was changed, which corresponded to 252 downregulated and 190 upregulated genes **(Figure 27a).**

In a similar experiment, the effect of nSMARt vectors was investigated in more challenging pancreatic cancer cells. Patient-derived Pancreatic Cancer Cells (Paco-2) were stably transfected with a nSMARt vector expressing the reporter gene GFP, (nSMARt_GFP). The RNA profiles of stably modified Paco-2 cells were compared to parental untransfected cells. The absence of bacterial backbone resulted in only a 0,05% change in RNA transcription, which corresponded to only 5 upregulated genes, mostly related to inflammatory pathways **(Figure 27b).**

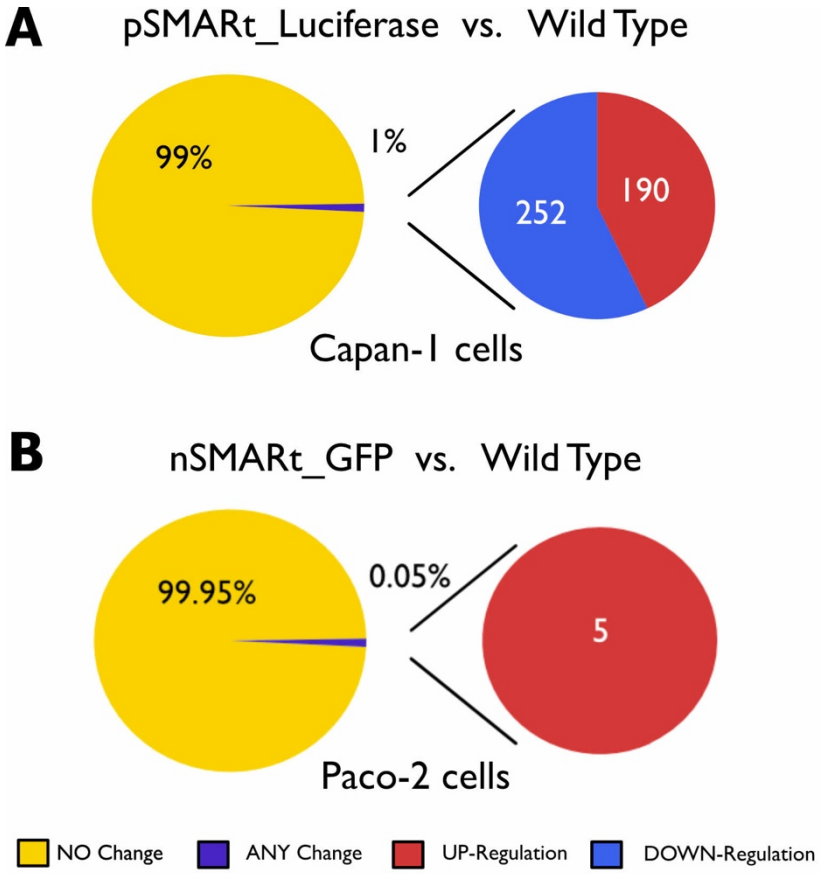


Figure 27: Genome-wide RNA profiles from pSMARt and nSMARt labelled pancreatic cell lines.
RNA from pancreatic tumours was extracted and used for microarray analysis on an IlluminaHuman.12 chip. The array was performed by Matthias Bozza with the help of the DKFZ Genomic and Proteomic Core Facility. The normalisation of samples was also performed there. For the analysis, pancreatic tumours from 3 mice were analysed per cell line, and the gene expression was quantified using Partek Genomic Suite Software (Thermo Fischer). The data analysis was performed by Matthias Bozza with the help of Mattia Falcone and Dr Elisa Espinet (HiStem, DKFZ). A cut off of > 2 fold and <-2 fold, FDR = 0.1 was applied to study the expression differences. The figure shows the number of genes up- and down-regulated in pSMARt_luciferase modified cell lines when compared to the parental cells **(A)** and between nSMARt_GFP and parental cells **(B).**

5.1.5 Replacing the S/MAR motif – The birth of pSMARter

After rearranging the conformation of the vector by coupling the selection marker to the expression cassette and reducing the size of the vector and its immunotoxicity by eliminating the bacterial backbone, we replaced the most important feature of this vector system. The S/MAR motif was first cloned in 1999 from the human ß-interferon gene [162], and it has been considered irreplaceable until now. However, we challenged the original vector sequence and replaced the human ß-interferon S/MAR motif by a smaller and more efficient sequence. This change gives rise to the new generation of smaller pSMARter vectors **(Figure 22g).**

This new vector prototype was tested and compared to the previously generated vectors in HEK293T cells. Although no nano version of pSMARter has been generated yet, this vector performed better in

HEK293T cells than any other type of vector, including the nano vectors, devoid of bacterial backbone **(Figure 28).**

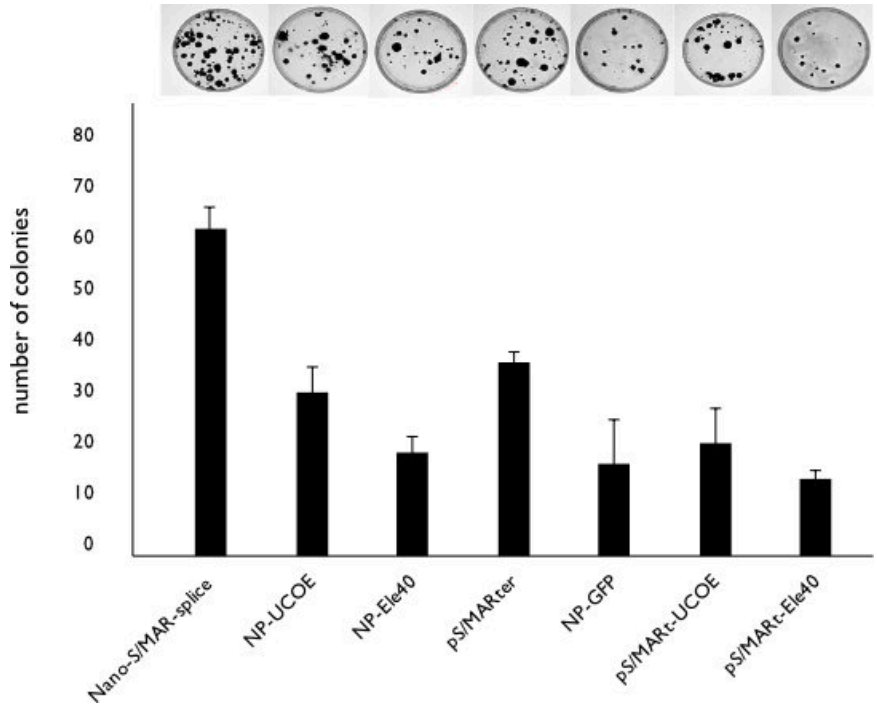


Figure 28: Comparison of establishment efficiencies between all DNA vector series.
This figure illustrates the colony forming assay and establishing the efficiency of all the series of vectors generated in the lab in the past three years. As described before, the DNA was transfected into HEK293T cells, which were sorted for GFP expression suing a FACS Aria II. Then, 100 GFP positive cells were plated and selected with 0.5ug/ml Puromycin for 4 weeks. After colonies were formed, those were fixed with PFA and fixed with crystal violet. The results are the average of 3 independent experiments, and a picture of colonies from a representative experiment is shown above the graphs. pSMARter, although carrying a bacterial backbone, performed better than the other nano vectors and the nano-SMAR-splice vector produced the highest amount of colonies. *Credit: Matthias Bozza, DNA Vector Lab*

5.1.6 Splicing the S/MAR out of the transcript – the birth of nSpliced

A spliced version of S/MAR vectors was designed and gave rise to the nSpliced vector generation **(Figure 22f),** which contained an expression cassette that would mimic a human endogenous gene in which the transgene contains intron-exon like structures. We compared a nano S/MAR spliced vector (nSpliced) to the previous vector series in a colony forming assay **(Figure 28).** We also compared the levels of transcript expression performing RT-PCR and showed that the nano-SMAR-spliced had the highest and most stable mRNA expression (data not shown).

5.1.7 Perfecting the system - Finding the right promoter

Previously described vector improvements were performed in vectors containing a CMV promoter. Although yielding high levels of transgene expression, viral promoters tend to get silenced already in pluripotent cells or during differentiation [225], making them unsuitable for stem cell work. Therefore, a library of vectors based on the pSMARt generation was cloned and tested in HEK293T cells. Vectors with different promoters from eukaryotic origins, such as UbC, PGK, EF1, or the chimeric promoter CAG; were transfected and its expression was monitored over time **(Figure 29).** The cells were imaged 24h post transfection, and no visual differences were observed regarding transfection efficiency. However, two weeks after selection with Puromycin, a significant drop in GFP expression was observed in all vectors except for pSMARt_CAG.

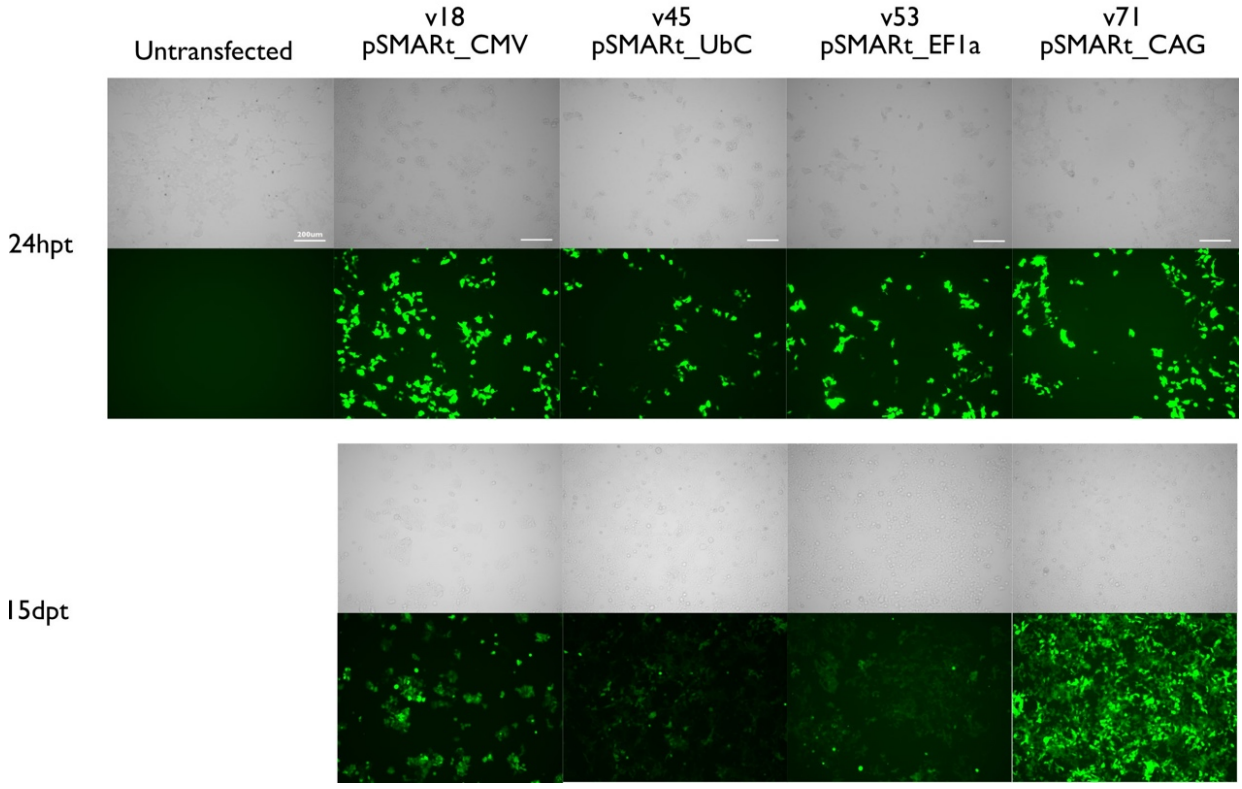


Figure 29: Testing a library of pSMARt vectors with different promoters in HEK293T cells.
A range of vectors with different promoters (CMV, UbC, EF1a and CAG) were used to transfect HEK293T cells. The cells were selected with Puromycin and imaged over time. No differences in the transfection efficiency were observed at 24h. However, only the pSMARt_CAG vector was able to sustain the expression of the transgene after two weeks of selection.

5.1.7.1 Testing of different S/MAR prototypes in mouse embryonic fibroblasts

The use of S/MAR vectors for the derivation of iPSC has never been shown before. For this, the vector has to be expressed in differentiated cells, such as fibroblasts, but remain expressed through reprogramming. Therefore, the transfection efficiency and expression levels of pSMARt_CAG (pCAG) and nSMARt_CAG (nCAG) were tested in immortalised MEFs. Taking into account the vector establishment performance of nSpliced_CMV and pSMARter_CMV they were also included in this comparative experiment.

MEFs were electroporated and kept under Puromycin selection for 4 weeks. After that period, the selective pressure was then removed. The GFP expression was monitored via fluorescent microscopy and quantified using flow cytometry. Representative fluorescent images and FACS histograms are shown in **Figure 30.**

In this comparison, both pSMARt_CAG and its minimally sized counterpart nSMARt_CAG were found to be suitable for transfecting and labelling MEFs as well as to express high levels of the transgene (GFP) over time, even in the absence of selection. nSMARt_CAG showed better transfection efficiency than pSMARt_CAG (21,8% vs 16,5%). Also, nSMARt_CAG transfected cells had slightly higher GFP expression levels, and the population was more homogeneous as compared to the pSMARt_CAG population. The spliced version nSMAR_Spliced showed the best transfection efficiency (33,6%) amongst all vectors tested, and its GFP intensity was comparable to that from nSMAR_spliced. Although pSMARter showed similar transfection efficiencies, its GFP expression levels were almost one log scale lower than nSMAR_spliced.

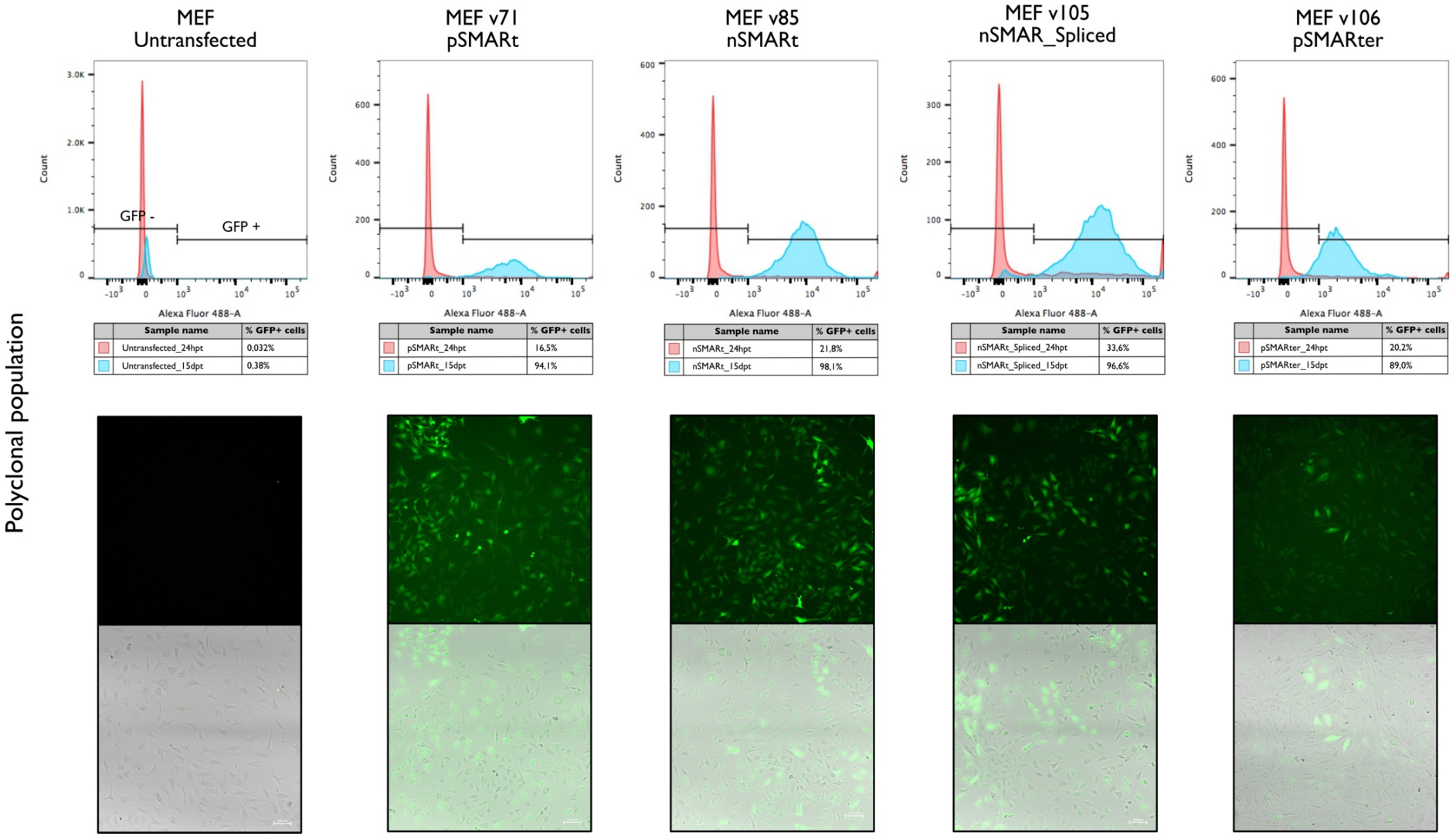


Figure 30: Evaluation of transfection efficiency and expression levels of different S/MAR vectors in murine fibroblasts

pSMARt_CAG-GFP (pSMARt) and nSMARt_CAG-GFP (nSMARt) vectors were tested in Mouse Embryonic Fibroblasts (MEFs). A spliced vector version in which the S/MAR motif was edited and removed after transcription (nSpliced) and a vector with a minimal S/MAR motif from the human Apolipoprotein B (pSMARter) was also included. The fibroblasts were kept under Puromycin selection for 30 days and the selection pressure was withdrawn afterwards. The nSpliced vector showed the best transfection efficiency amongst all vectors (33,6%). However, after selection with Puromycin, the nSMARt vector showed higher percentage of GFP expressing cells.

5.1.7.2 Testing of different S/MAR prototypes in murine stem cells (mESC)

A principal aim of this work is to understand the S/MAR vector behaviour during differentiation. As described above, S/MAR vectors such as the original pEPI minicircle, were able to transfect mESC; although their expression levels were heterogeneous and dramatically dropped during the experiment. After improving and testing the aforementioned vector prototypes in HEK293T, their transfection efficiency was tested, and their expression levels were monitored in the mESC-line E14 129/Ola.

The cells were electroporated and kept under Puromycin selection for 4 weeks. After that period, the selective pressure was then removed. The GFP expression was monitored microscopically and quantified using flow cytometry. Representative fluorescent images and FACS histograms for polyclonal and clonal populations are shown below in **Figure 31** and **Figure 32.**

Both pSMARt_CAG and its minimally sized counterpart nSMARt_CAG were suitable for transfecting and labelling mESC and both vectors were able to express high levels of the transgene (GFP) over time, even in the absence of selection. nSMARt_CAG showed better transfection efficiency than pSMARt_CAG (45,1% vs 39,2%). Even more obvious than in MEFs, the nSMARt_CAG transfected cells had the highest GFP expression levels, and the population was more homogeneous as compared to the other vectors. The spliced version nSMAR_Spliced and pSMARter, showed similar performance compared to pSMARt_CAG, albeit containing a CMV promoter with an insulating element.

When clones from S/MAR-labelled mESC were picked after one week of Puromycin selection, we could observe that nSMAR_CAG was the highest expressing vector in mESC. Also, pSMARt_CAG and nSMAR_spliced showed similar expression levels, whereas pSMARter performed the worst amongst all vectors tested.

Note: this experiment was chronologically performed after deciding in favour of pSMARt_CAG and nSMARt_CAG vectors as best performing vectors at the time. Later in the course of this thesis, other vectors such as nSMAR_Spliced and pSMARter, showed to have similar performance in transfecting and maintaining expression in MEFs. However, no differentiation or *in vivo* study was performed with them, and their suitability should be further tested. Also, the subsequent molecular analysis was performed mostly on pSMARt_CAG but not on nSMARt_CAG labelled cells. The reason being the absence of bacterial backbone and the inability of nano vectors to be rescued and retransformed in bacteria.

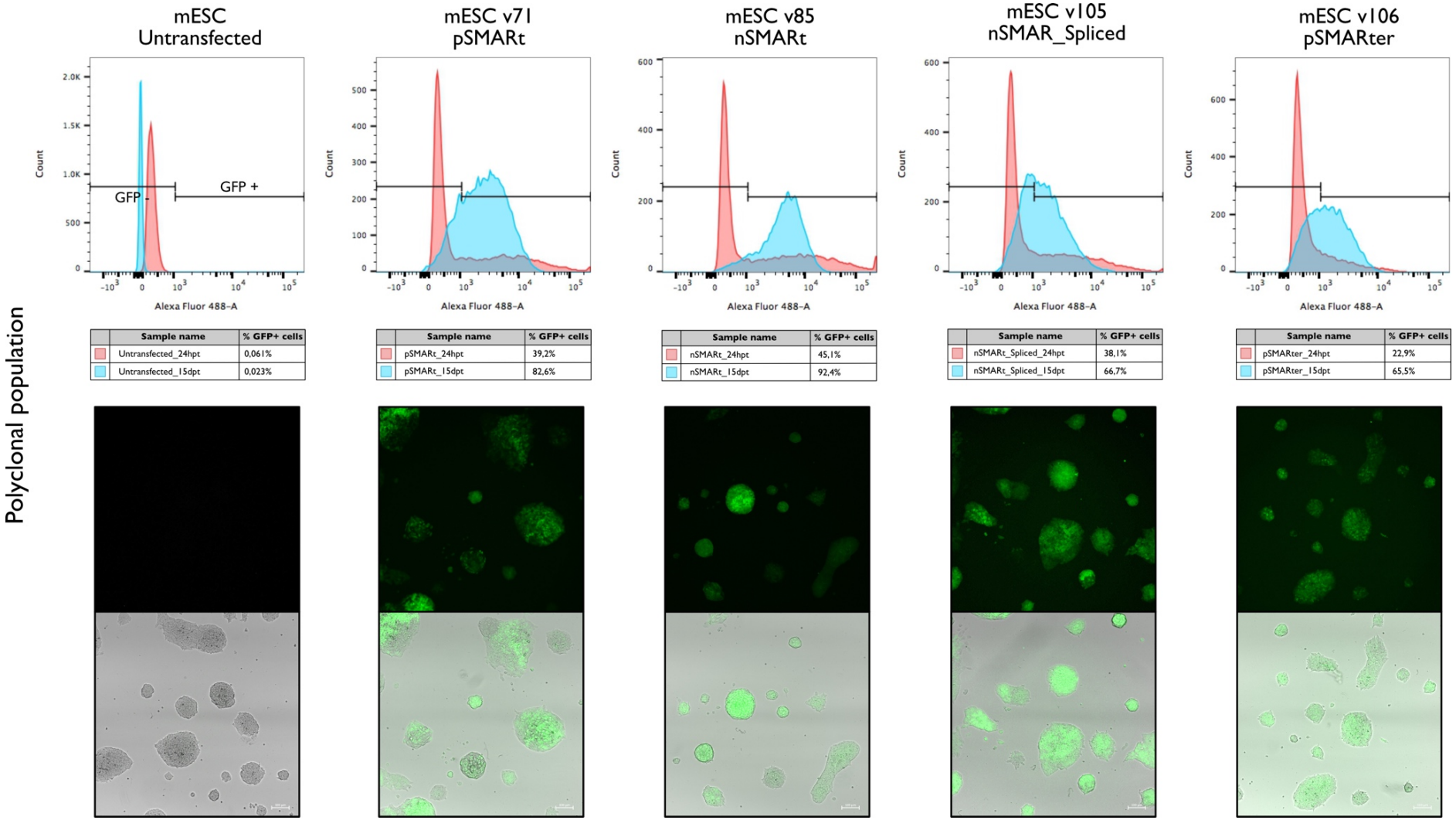


Figure 31: Evaluation of transfection efficiency and expression levels of different S/MAR vectors in polyclonal pupulations of mESC

pSMARt_CAG and nSMARt_CAG vectors were tested in Mouse Embryonic Stem Cells (mESC). We also included a nSMAR_Spliced version in which the S/MAR motif was edited and removed after transcription and a pSMARter vector with a minimal S/MAR motif from the human Apolipoprotein B. The mESC were kept under Puromycin selection for 30 days and the selection pressure was withdrawn afterwards. The nSMARt vector showed the best transfection efficiency amongst all vector tested. The nSMARt transfected cells have higher percentage of GFP expressing cells (92,4% vs 82,6%) and the population is more homogeneous in GFP expression (narrower peak) as compared to the pSMARt population. The nSMARt_Spliced vector performed better than pSMARter in both transfection efficiency and GFP expression.

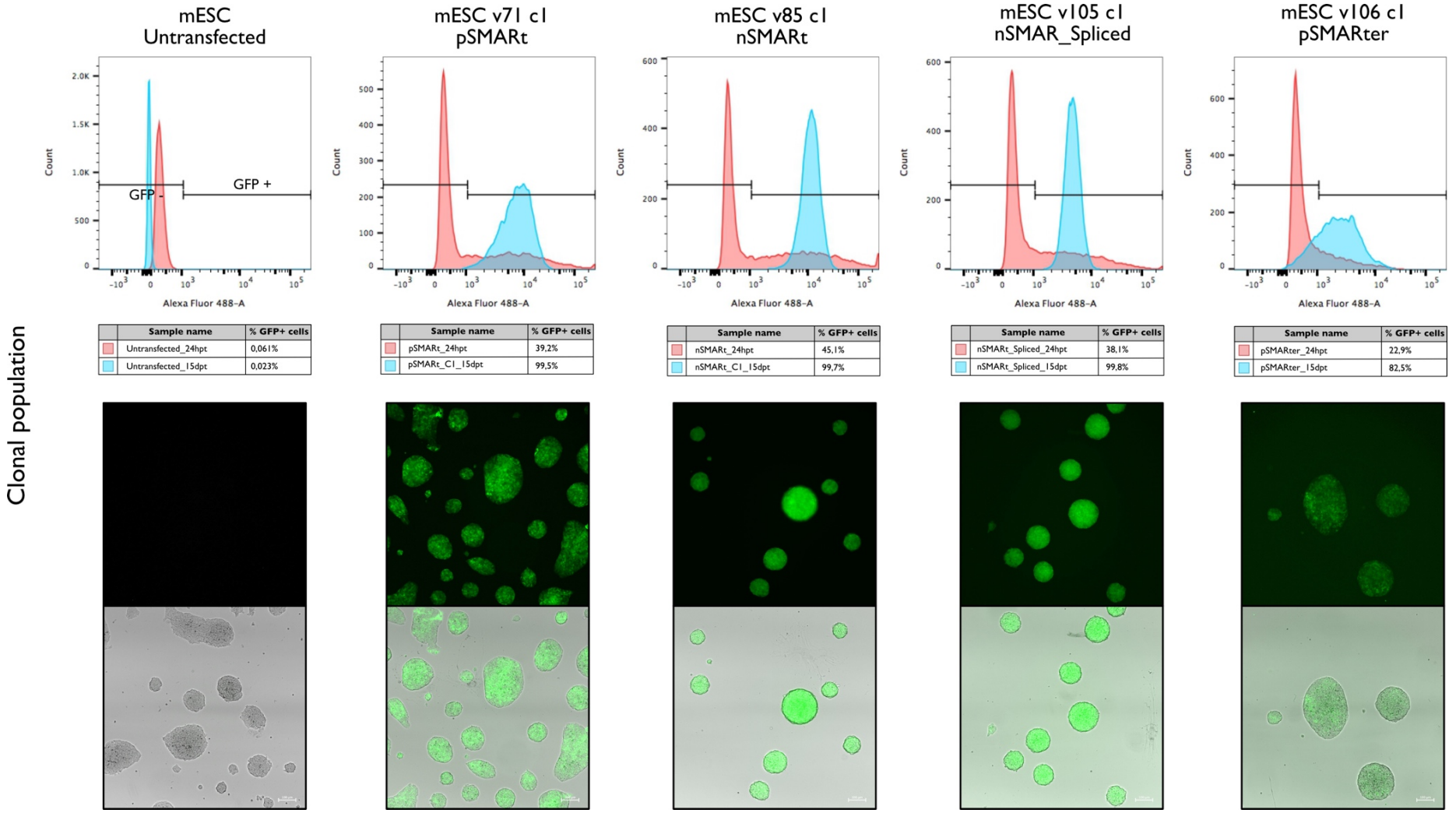


Figure 32: Evaluation of transfection efficiency and expression levels of different S/MAR vectors in mESC clones

After one week of selection, mESC colonies were picked, expanded and FACS analysed. pSMARt, nSMARt and nSMARt_Spliced clones were almost homogeneous with 99% of GFP+ cells. On the contrary, the pSMARter population was heterogeneous and the levels of GFP expression were lower than those from the other clones.

5.1.8 Introducing the main characters

This section describes the relevant vectors used in this thesis for *in vitro* and *in vivo* studies and that are based on the aforementioned vector prototypes. Here the reader can get a better understanding of the vector configuration as well as to the vectors' functionality, which is tested by assessing their protein and reporter gene expression.

5.1.8.1 Labelling vectors: pSMARt_CAG and nSMARt_CAG

pSMARt_CAG and nSMARt_CAG were chosen for *in vitro* and *in vivo* experiments. Differently, from the original pEPI vector, both vectors were capable of stably transfecting MEFs and mESC as shown in previous sections. Their expression levels were high and kept for as long as the cells were cultured, and their establishment efficiency was significantly improved.

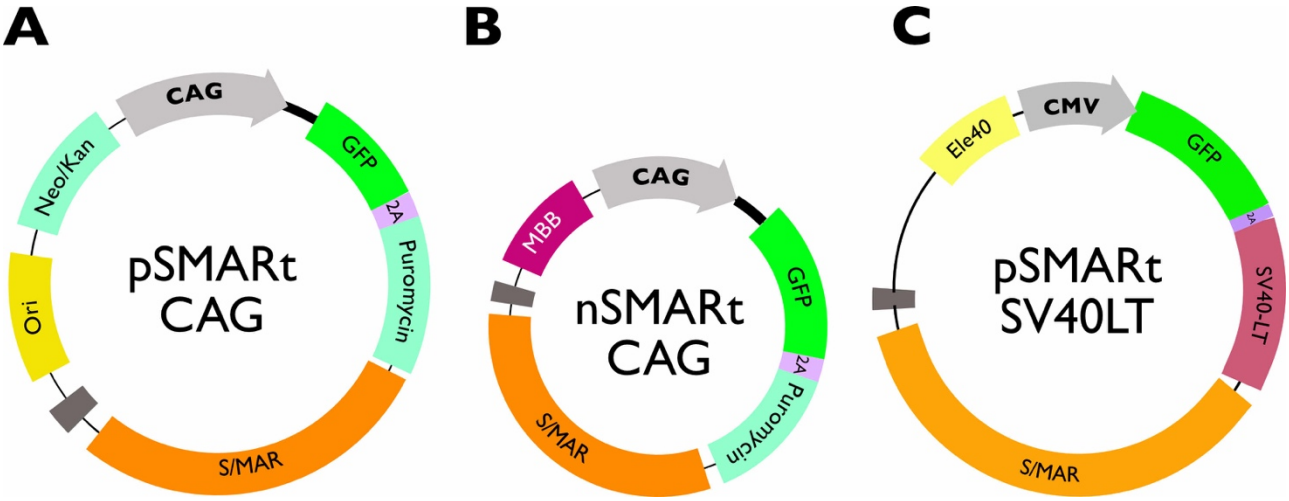


Figure 33: pSMARt (A) and nSMARt (B) labelling vectors. Immortalizing and labelling vector (C).

Both vectors are based on the chimeric CAG promoter, which comprises the cytomegalovirus (CMV) enhancer fused to the promoter and first intron of chicken ß-actin and the splice acceptor of rabbit ß-globin gene. A chimeric intron, which is meant to enhance the mRNA transcript stability, is placed directly downstream of the promoter. The transcription unit consists of the reporter gene coGFP and the selective marker Puromycin, separated by a self-cleavage 2A sequence, and runs through the S/MAR motif until the polyadenylation tail (polyA).

pSMARt_CAG contains approximately 1.5kbp of bacterial backbone, comprising a bacterial origin of replication and a dual Neomycin/Kanamycin resistance, whereas nSMARt_CAG contains a minimal

bacterial backbone (MBB) consisting of approximately 400 bp. However, the reduction of its bacterial backbone limits downstream applications, such as plasmid rescue experiments.

Both pSMARt_CAG and nSMARt_CAG were used to transfect MEFs and mESC as described above stably, and their expression after establishment is shown in **Figure 34.**

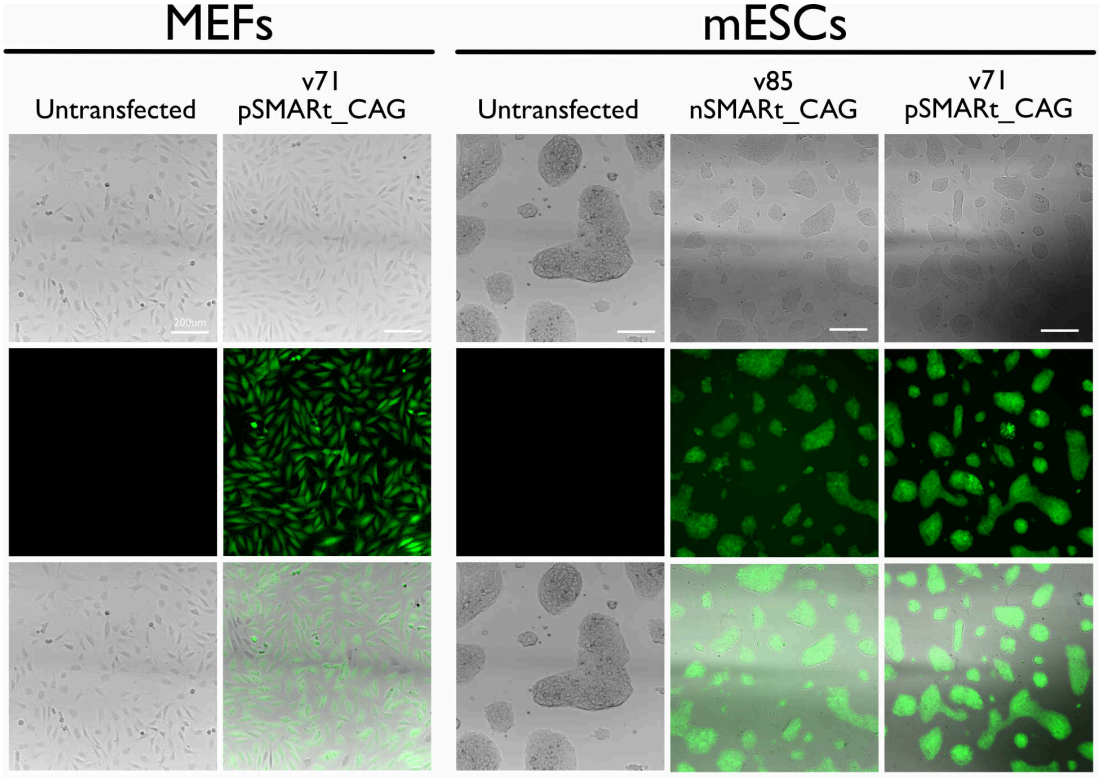


Figure 34: pSMARt and nSMARt drive persistent and high levels of transgene expression in MEF and mESC
MEFs and mESCs were nucleofected using the Amaxa II nucleofector device and the NHDF nucleofector kit or Mouse ES Cell nucleofector kit respectively. The cells were kept under selection with 500ng/ul of Puromycin for 4 weeks, and their GFP expression was checked using a Nikon Ti microscope (Exposure 1 second, amp gain x2.1). Scale bar=200um.

5.1.9 Labelling and immortalising vector

A bifunctional labelling and immortalising S/MAR DNA vector was generated by Matthias Bozza (DNA vector lab). This vector is based on the pSMARt backbone, but the puromycin was replaced by the SV40 Tag, linked to GFP through a 2A self-cleavage peptide. The CMV promoter drives the expression of the cassette, which is further protected from silencing by the anti-repressive Element40. A cartoon illustrating the vector is shown in **Figure 33*c.***

5.1.10 Reprogramming vectors derived from the 4in1 Lentiviral vector: POP and nPOP

Different versions of S/MAR vectors for mice and human reprogramming were generated in this study. The episomal vectors are based on the **polycistronic lentiviral reprogramming vector '4in1'** described by Warlich *et al.* [285] and Kuehle *et al.* [297], which codes for the human codon optimised sequence of the Yamanaka reprogramming factors Oct4, Klf4, Sox2 and cMyc, separated by different 2A self-cleavage sequences and driven by the viral SFFV promoter. The red reporter gene dTomato (dTom) is separated by an internal ribosome entry site (IRES). The different 2A sequences, which have different self-cleavage efficiencies, generate a decreasing amount of protein encoded in the mRNA transcript (expression levels Oct4>Klf4>Sox2>cMyc). Also, the viral SFFV promoter is chosen due to its silencing after reprogramming; when the cell switches on its internal pluripotency machinery and the external reprogramming factors are not desirable anymore [225].

Two reprogramming vector series were generated in this study and are depicted in **Figure 35.** Both are based on the Lentiviral 4in1 vector but featuring different elements, which will allow them to serve different purposes.

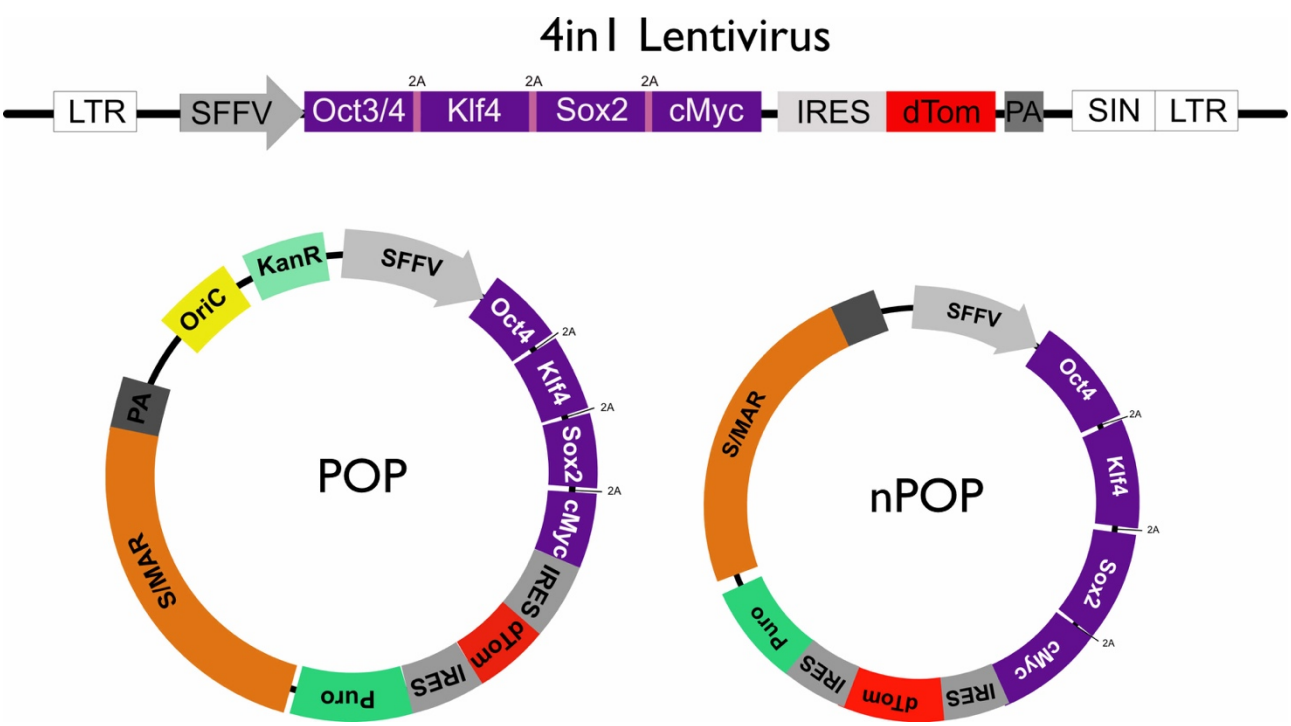


Figure 35: Cartoon depicting different versions of reprograming vectors

The **Proof-of-Principle (POP)** vector was generated by subcloning the transcription unit of the Lentiviral 4in1 vector into a pSMARt vector, already containing an IRES-Puro linked to the S/MAR. This resulted in a hexacistronic cassette controlled by the SFFV promoter driving the expression of OKSM, the reporter gene dTomato and the selective marker puromycin. This vector not only allows to reprogram cells into iPSC but also serves as a proof-of-principle that the S/MAR motif needs to be part of a transcriptionally active unit to be kept episomal and functioning. Upon transfection, the POP vector will express the OKSM reprogramming factors, together with the dTom reporter gene and can also be selected with Puromycin. However, when the transfected cells reach pluripotency, the viral SFFV promoter will get silenced, which will result in loss of the vector containing the reprogramming factors, which are not only useless at that stage but also undesirable.

The **nano-Proof-of-Principle (nPOP)** vector is a minimally sized version of POP, in which the bacterial backbone is reduced and replaced by a minimal sequence. Its reduced size allows better transfection efficiency and the minimal bacterial sequence makes it less prone to be methylated and silenced, yielding higher and longer expression of reprogramming factors, and overall making the reprogramming process more efficient. It is important to highlight that both POP and nPOP vectors should lose expression of their reporter genes, and hence the vector, after reprogramming.

The reprogramming vectors were tested and transfected using jetPEI into HEK293T cells to check for fluorescent reporter expression as well as protein expression by Western Blot. Untransfected cells, which were used as negative control, did not show fluorescent expression of the reporter genes. The lentiviral vector 4in1, which was used as positive control, expressed the reporter gene dTom as well as pPOP **(Figure 36a).** The expression of reprogramming factors was further tested by Western Blot, as shown in **Figure 36b.** nPOP was generated at a later timepoint, and its expression analysis is not included in this experiment.

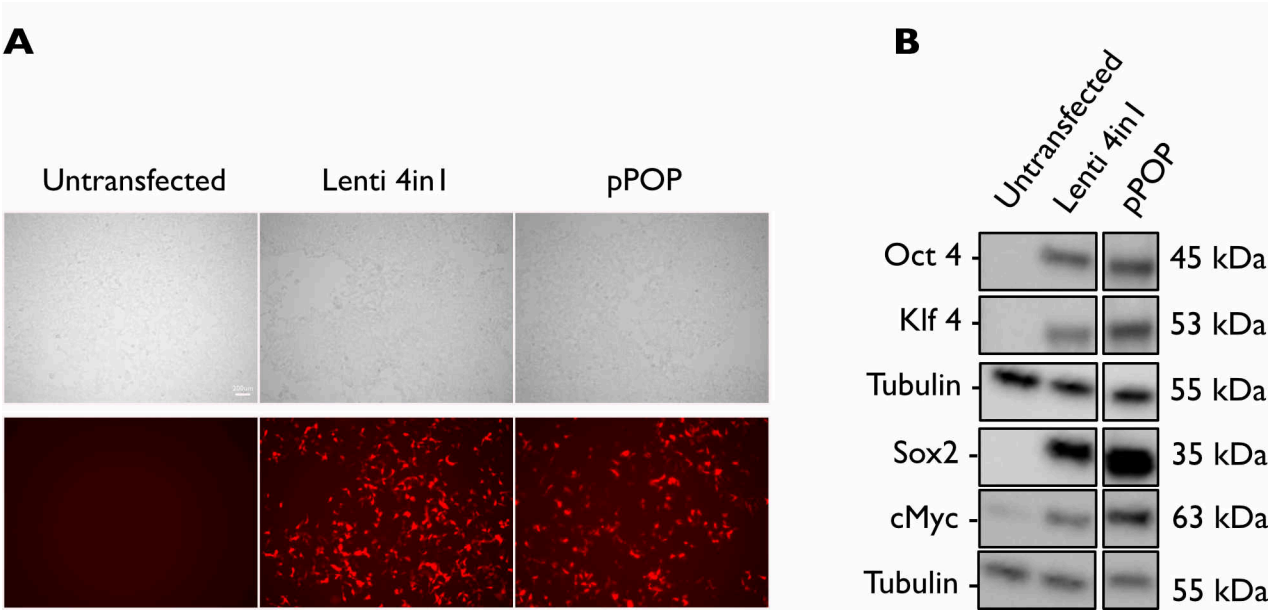


Figure 36: Assessment of reprogramming vector functionality in HEK293T cells.
HEK293T cells were transfected with the reprogramming vectors Lenti4in1 and POP using jetPEI. The cells were then imaged 24hpt with a Keyence fluorescent microscope. Scale bar = 200um. All vectors showed expression of the reprogramming cassette (dTom) **(A).** Whole cell lysates from transfected HEK293T cells were obtained 24hpt. The protein concentration was determined using a BCA and 30ug of total protein were loaded per lane. αTubulin (55kDa) was used as a loading control. Both vectors showed expression of the Yamanaka reprogramming factors Oct4, Sox2, Klf4 and cMyc **(B).**

5.1.11 Reprogramming vectors derived from the EBNA-1/OriP system

Previously, lots of effort has been directed into developing safer and non-integrative (episomal) alternatives for the genetic modification of cells, for example by using EBNA-1 based vectors. Although non-integrative, EBNA-based vectors rely on continuous expression of the oncoprotein EBNA-1, which might affect the cells' behaviour by remodelling the chromatin or altering the cells' transcription profile [253]. By replacing the EBNA-1 by S/MAR motifs, we hoped to lower the impact of the vector on cell transcription as well as to eliminate the oncogenicity of the vectors. Prof. Dr Mckay from Manchester Metropolitan University (MMU) in the United Kingdom, kindly provided the EBNA-1 vectors containing the reprogramming factors Oct4, Sox2, Klf4, L-Myc and Lin28 described in [298].

The **EBNA-1 reprogramming system** comprises 4 vectors, two of them containing a CAG promoter driving expression of a bicistronic cassette containing either Sox2 and Klf4 or Lin28 and L-Myc, separated by a 2A self-cleavage peptide, and the OriP/EBNA-1 sequences. A third vector with the same configuration codes for Oct4 and a shRNA to knockdown P53, which enhances the reprogramming efficiency by preventing the cell cycle arrest and therefore, increase cell proliferation [299 - 301]. A fourth vector containing only EBNA-1 is used to boost the expression of this protein and aid the replication of the other three vectors **(Figure 37a).**

The **S/MAR reprogramming system** is composed of 3 vectors, which are derived from the aforementioned EBNA-1 vectors. To generate them, the reprogramming factors were amplified and subcloned into pSMARt_CAG, which was digested with *AgeI* and *XhoI* to remove the GFP-2A-Puro sequence **(Figure 37b).** In the 3-S/MAR vector system, there is no need for a 4th 'S/MAR enhancing' vector since the S/MAR motif does not need to be translated to be active. An intermediate pSMARt_CAG vector, including the shRNA to knockdown P53, was also generated. This vector was used as a labelling vector for co-transfection with the reprogramming vectors POP and nPOP.

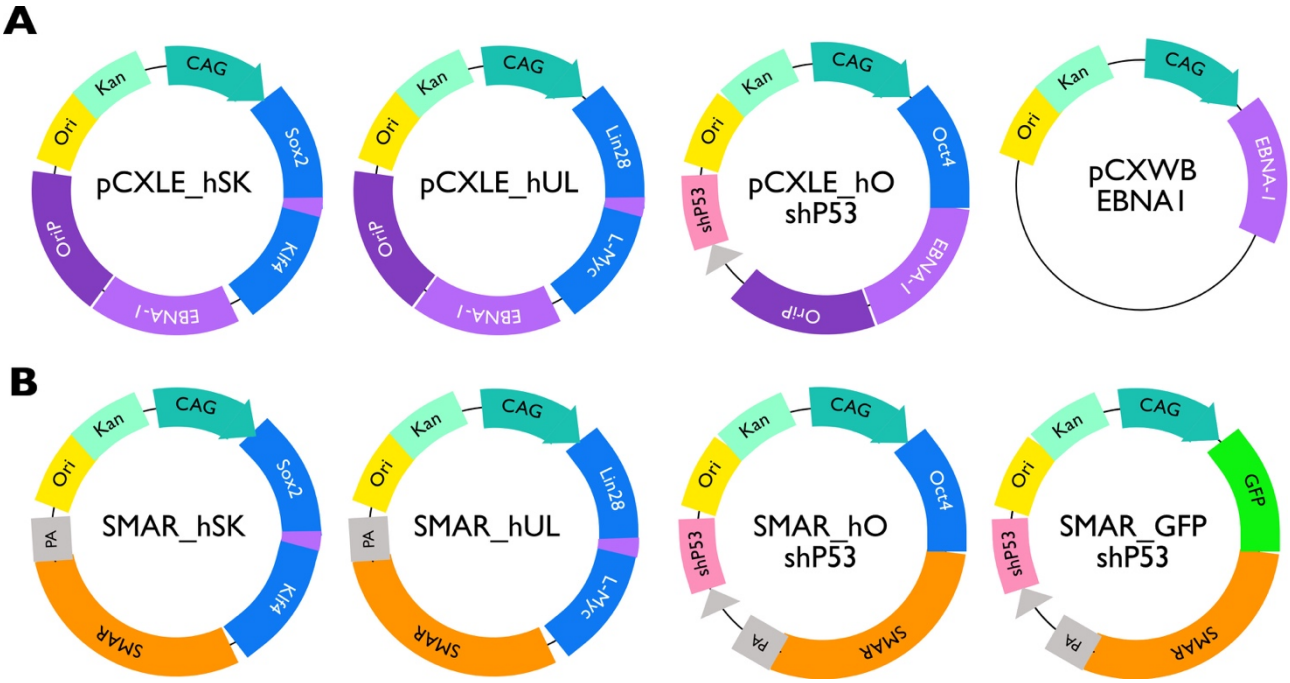


Figure 37: Cartoon depicting the EBNA-1 reprogramming vectors and their S/MAR counterparts

Since both EBNA and S/MAR vectors lacked the presence of a reporter gene, the expression and functionality of the vectors were tested by assessing the presence of the reprogramming factors via Western Blot, as shown in **Figure 38.** For this, the three reprogramming EBNA vectors, as well as the three S/MAR vectors, were transfected separately into HEK293T cells. Untransfected cells were used as negative control, and Lentivirally transduced cells with the 4in1 vector were used as positive control. However, the total protein concentration from the 4in1 sample was low, and the expression of the reprogramming factors was almost undetectable. All EBNA and S/MAR vectors showed expression of the reprogramming factors, which proved successful cloning and demonstrated the functionality of the vectors. Due to the lack of L-Myc antibody, its expression could not be determined. However, expression of Lin28, which is transcribed together with L-Myc, could be detected and suggested that L-Myc was also expressed.

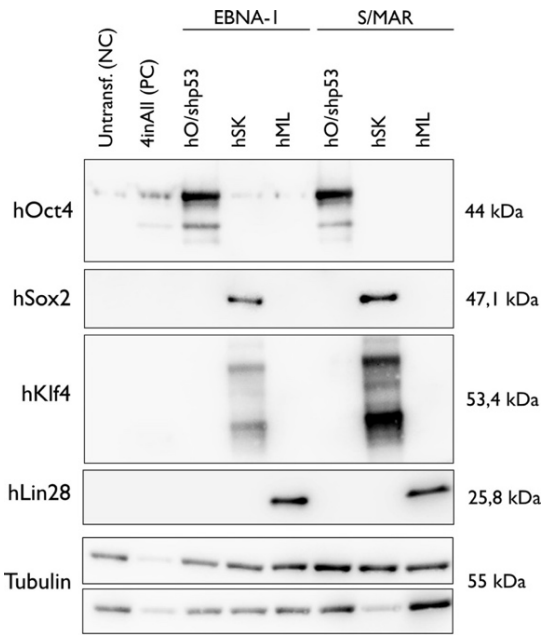


Figure 38: Yamanaka factors' expression (OKSML) from EBNA-1 and S/MAR reprogramming vectors.
Whole cell lysates from HEK293T cells transfected individually with the EBNA and S/MAR reprogramming vectors were obtained at 24hpt. The protein concentration was determined using a BCA and 30ug of total protein were loaded per lane. αTubulin (55kDa) was used as a loading control. All vectors showed expression of the reprogramming factors Oct4, Sox2, Klf4 and Lin28. Although expression of L-Myc could not be assessed due to lack of antibody, its expression was presumed since it belonged to the same transcription unit as Lin28.

5.1.12 Therapeutic vectors

Two candidate diseases were chosen to be targeted and corrected using our S/MAR vector technology: Fanconi Anemia (FA) and Choroideremia. Although in principle very different, both diseases share their monogenic recessive nature and can be corrected by supplementation of the defective gene, which makes them an ideal candidate for gene therapy.

5.1.12.1 FANC-A vectors

The reference **S11FAIEGnls vector** (FancA_GFP) described in [275, 286] was kindly provided by Dr Milsom (Hi-STEM, DKFZ). This lentiviral vector contains a CMV promoter driving expression of the FancA cDNA and the reporter gene eGFP with a nuclear localisation signal (NLS), separated by an IRES sequence. The FancA cDNA was subcloned into a pSMARt backbone containing firefly luciferase, to generate the **pSMARt_FancA-Luc vector**.

Both vectors were tested for reporter gene and FancA expression in HEK293T cells. For this, the same amount of molecules of each plasmid were transfected using jetPEI, and whole cell lysates were obtained after 24h. The reporter gene expression was determined by delivering Luciferin (10 µl/ml) into the cultured cells and measuring chemiluminescence with a Fusion SL device. The eGFP expression was measured via conventional fluorescent microscopy **(Figure 39a).** To detect the expression of FancA, 30 µg of total protein extracts were used in a Western Blot. Untransfected HEK293T cells were used as a negative control and αTubulin was used as a loading control. FancA expression was quantified by

measuring the intensity of the protein band in the blot and using the FusionSL analysis software **(Figure 39b).** Both vectors were functional and expressed both the therapeutic FancA gene and their respective reporter genes. Although having equalled the number of molecules of FancA-GFP and pSMARt_FancA-Luc, the latest vector showed higher expression of FancA. It is important to note that HEK203T cells are of human origin and they endogenously express FancA. However, the overexpression of exogenous FancA proceeding from the vector transfection can be readily observed by comparing the intensity of the band between the negative control and FancA overexpressing cells.

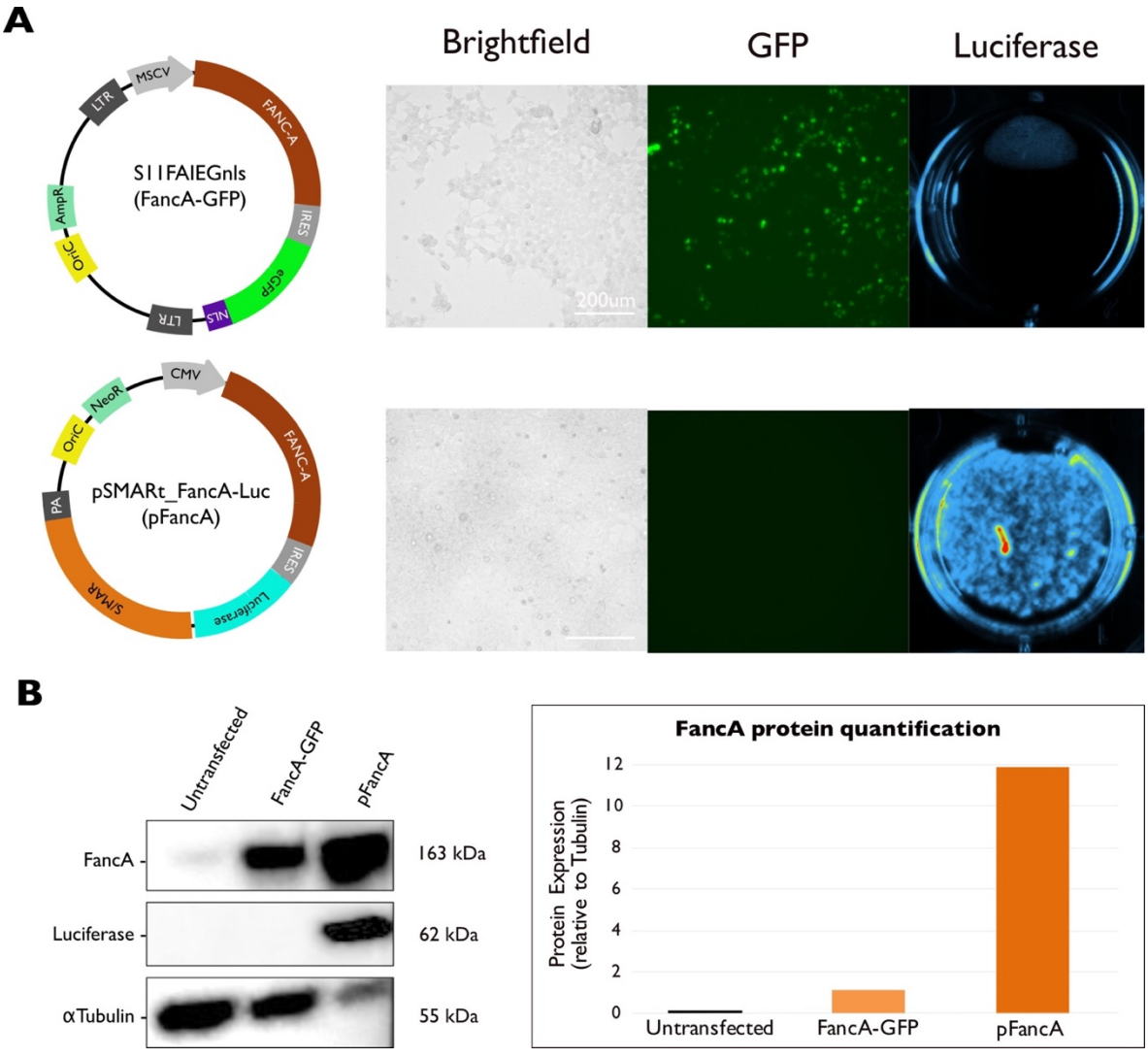
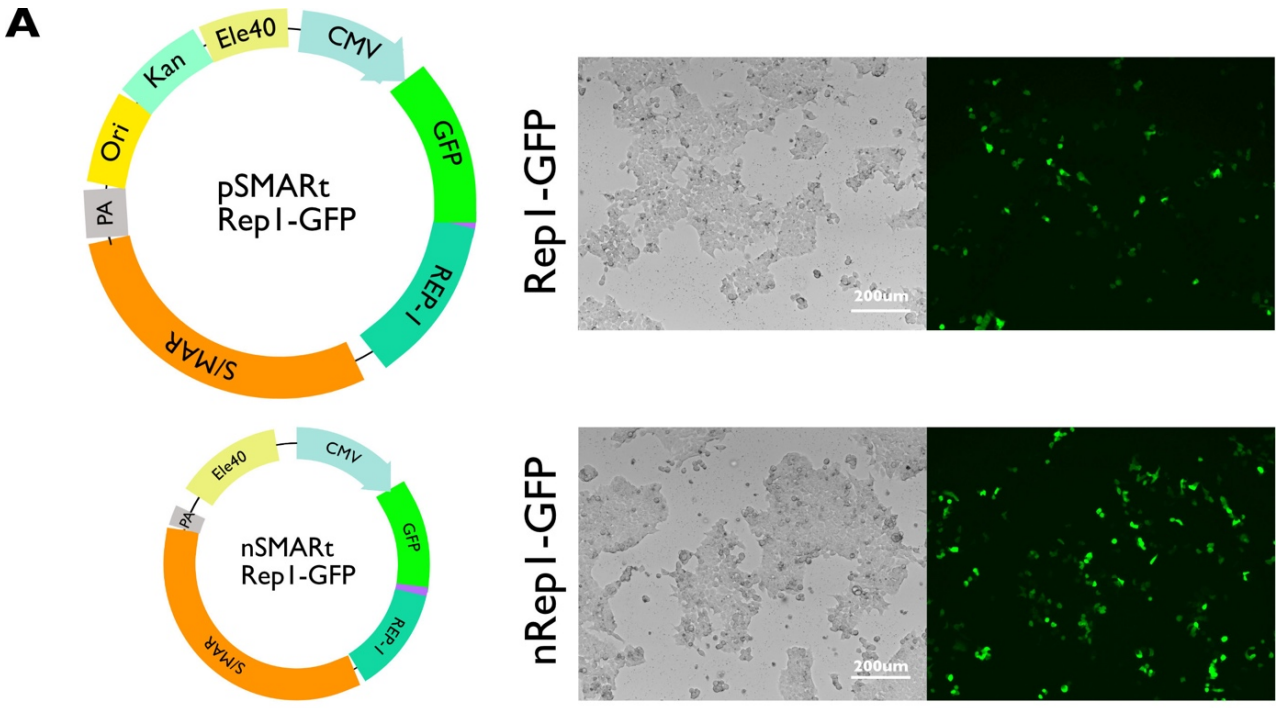


Figure 39: Testing the therapeutic FancA vectors in HEK293T cells.
The lentiviral S11FAIEGnls vector was provided by Dr.Milsom (Hi-STEM, DKFZ) and used as a reference vector. A pSMARt_FancA-Luc vector was generated by subcloning the cDNA from FancA into a pSMARt backbone. Both vectors were transfected into HEK293T cells, and untransfected cells were used as negative control. The reporter gene expression was checked either by fluorescent microscopy or addition of 10ug/ml of Luciferin **(A).** 30ug of Whole Cell Lysates were used to detect FancA expression in a Western Blot, which was later on quantified **(B).** Both FancA vectors were functional and expressed the FancA as well as their respective reporter genes.

5.1.12.2 REP-1 vectors

The *CHM/REP1* lentiviral vector was kindly provided by Dr Tanya Tolmachova [28]. The *Rep1* cDNA was subcloned into a pSMARt backbone, replacing the selective marker Puromycin and generating pSMARt_Rep1-GFP. A minimally sized nano vector (nSMARt_Rep1-GFP) was synthesised in collaboration with Nature Technology Corporation (NTX). To test the vectors' functionality, the same amount of molecules from both vectors were transfected into HEK293T cells using jetPEI. After 24h, the expression of the reporter gene was determined by fluorescent microscopy **(Figure 40a).** Also, 30 μg of total protein extracts were used in a Western Blot. Untransfected HEK293T cells were used as a negative control and αTubulin was used as a loading control. Rep1 expression was quantified by measuring the intensity of the protein band in the western blot and using the FusionSL analysis software **(Figure 40b).**

Both vectors were functional and expressed both the Rep1 protein and GFP. Although having equalled the number of molecules of pSMARt_Rep1-GFP and nSMARt_Rep1-GFP, the latest vector showed higher expression of the therapeutic gene. It is important to note that HEK203T cells are of human origin and they endogenously express Rep1. However, the overexpression of exogenous Rep1 can be readily observed by comparing the intensity of the band between the negative control and Rep1 overexpressing cells.



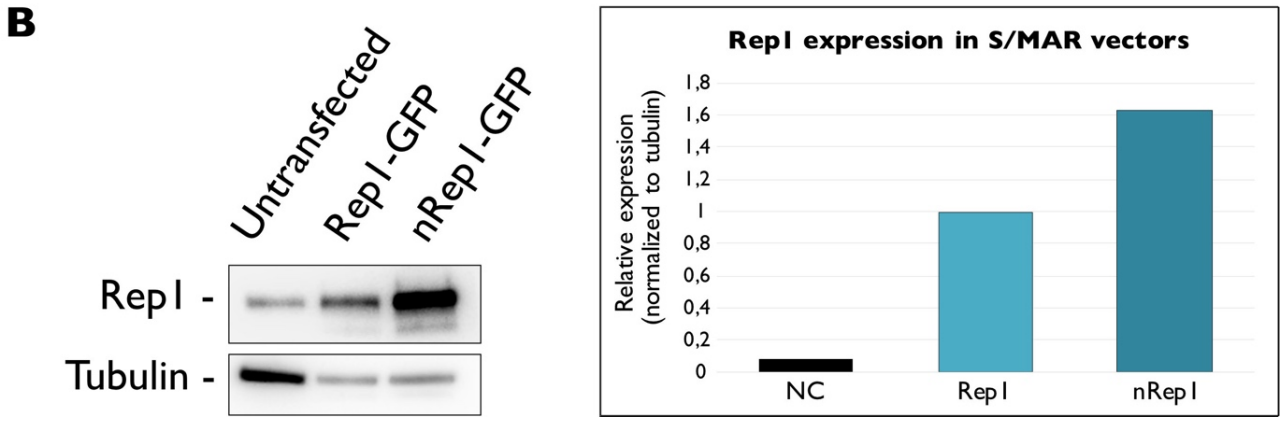


Figure 40: Testing the therapeutic Rep-1 vectors in HEK293T cells.
The lentiviral Rep1 vector was provided by Dr Tolmachova and used as a reference vector. A pSMARt_Rep1-GFP vector was generated by subcloning the cDNA from Rep1 into a pSMARt backbone. A minimally sized nSMARt_Rep1-GFP was also synthetised. Both vectors were transfected into HEK293T cells, and untransfected cells were used as negative control. The reporter gene expression was checked by fluorescent microscopy **(A).** 30ug of Whole Cell Lysates were used to detect Rep1 expression in a Western Blot, which was later on quantified **(B).** The functionality of both vectors was confirmed by expression of GFP and Rep1. The expression levels of nSMARt_Rep1-GFP were higher than those in pSMARt_Rep1-GFP.

5.1.13 Summary

An overview of the S/MAR DNA vector's evolution from the originally described pEPI vector is provided in this chapter. These modifications included resulted in a series of vectors that, by all means, outperformed the original vector. Also, relevant vectors for this study are described. Their expression is assessed by fluorescent microscopy or chemiluminescence as well as by Western Blot.

- The change and relocation of the selection marker gave raise to **pSMARt** series of vectors, which was later on refined by removing the majority of its bacterial backbone **(nSMARt).**
- Different mammalian promoters were evaluated and the CAG promoter proved to be the most suitable for both mESC and also MEFs work (which will eventually be reprogrammed into iPSCs).
- Vectors such as **nSMAR_spliced** and **pSMARter** proved to be the best performing vectors in expression and establishment (in HEK293T cells). However, pSMARter seems unsuitable for mESC work, probably due to the presence of CMV promoter.
- The vector series pSMARt, nSMARt (cloned early in this study) were used for further *in vitro* and *in vivo* studies, whereas the suitability of both nSMAR_spliced and pSMARter (cloned later) should be further evaluated.

5.2 Evaluating the potential of S/MAR vectors to survive differentiation and de-differentiation in ATRA-mediated neuronal differentiation model

To investigate whether S/MAR vectors would retain expression throughout differentiation and reprogramming, a model based on All-Trans Retinoic Acid (ATRA) mediated differentiation of Be2C neuroblastoma cells was used **(Figure 41).** Be2C cells were labelled with pSMARt vectors, and their expression during differentiation was assessed through different molecular analysis. Domenic Hartmann and Dr Jeannine Lacroix (DKFZ), kindly provided the neuroblastoma cells and the expertise on the topic.

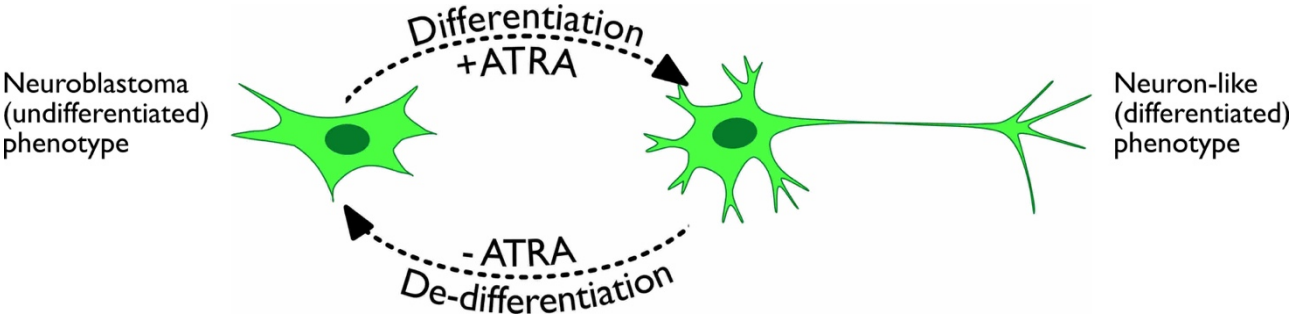


Figure 41: ATRA-mediated neuronal differentiation
In this model, Be2C neuroblastoma cells are forced to 'differentiate' into a neuron-like phenotype following the addition of ATRA. This simulates a differentiation process from a pluripotent to a differentiated cell. The reverse 'de-differentiation' process is achieved by withdrawing the ATRA and allowing the cells to recover their original neuroblastoma phenotype, which imitates the reprogramming process from a differentiated to a pluripotent state.

Although the terms 'differentiation' and 'de-differentiation' are used to refer to transitions between a cancer- and a neuron-like phenotype, such processes cannot be taken as real or biologically accurate. Due to the nature of the neuroblastoma cells, which carry an altered genotype, any differentiated progeny coming from the cells would also carry an altered genotype. Therefore, the results obtained using this model are only indicative and might serve as a tool to understand the behaviour of S/MAR vectors during a real differentiation from a stem cell to a fully differentiated cell.

5.2.1 Persistent expression of S/MAR vectors in neuroblastoma cells

Be2C neuroblastoma cells, a cancer cell line derived from a bone marrow biopsy of a 2-year-old boy, were labelled with pSMARt_GFP, pSMARt_Luciferase or a dual pSMARt_GFP-Luciferase. The cells were selected with 1µg/ml Puromycin for a month, and the selective pressure was removed afterwards. The cells were kept in culture and checked regularly for GFP and Luciferase expression as shown in **Figure 42.**

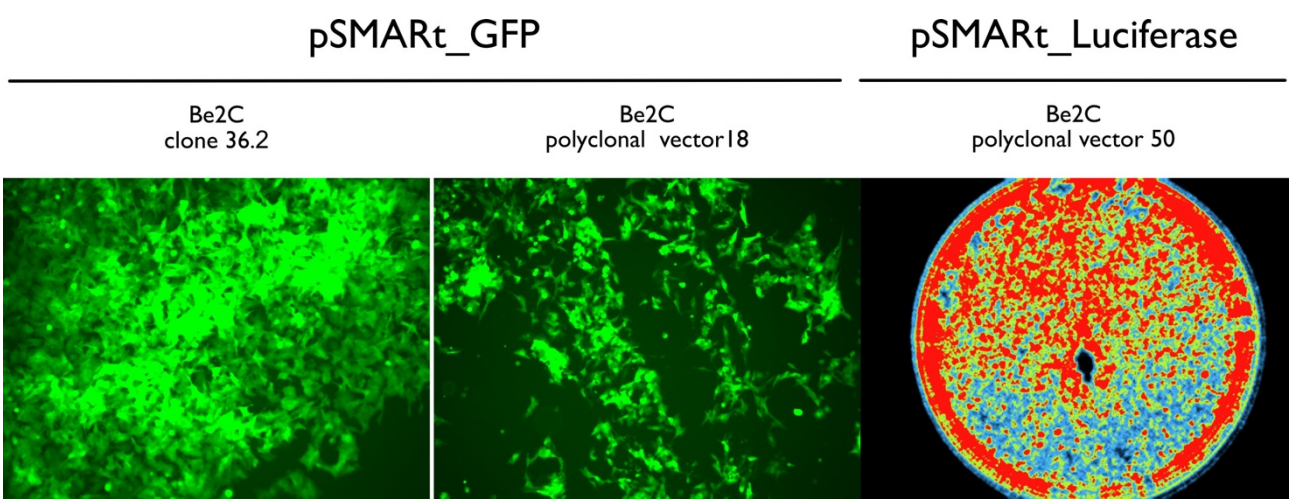


Figure 42: pSMARt vectors are persistently expressed in Be2C neuroblastoma cells
Be2C neuroblastoma cells were transfected with pSMARt vectors expressing either GFP (vector 18), Luciferase (vector 50) or a dual GFP-Luciferase cassette (vector 36). The cells were kept under selection with Puromycin for one month, and the selective pressure was then removed. The cells continued expressing the respective reporter genes in the absence of selection.

5.2.2 ATRA mediated neuronal differentiation of S/MAR labelled cells

5.2.2.1 S/MAR expression during 'differentiation' and 'de-differentiation'

We then wondered how pSMARt vectors would behave when labelled cells were grown in the presence of ATRA. Because fluorescence is more straightforward to monitor than chemiluminescence, pSMARt_GFP clones were chosen for differentiation experiments. As depicted in **Figure 43,** upon addition of 10µM of ATRA, pSMARt-GFP labelled Be2C cells started showing morphological changes, including elongation and neurite extensions in a dose-dependent manner as compared to the untreated cells (DMSO), as early as day 4. After 8-10 days, the cells clustered into 'ganglion-like' structures and stopped proliferating, as compared to the untreated control. At that point, the cells were still expressing GFP. At day 10, ATRA was withdrawn, and the cells recovered their cancer-like morphology as well as their proliferation, also in a dose-dependent manner.

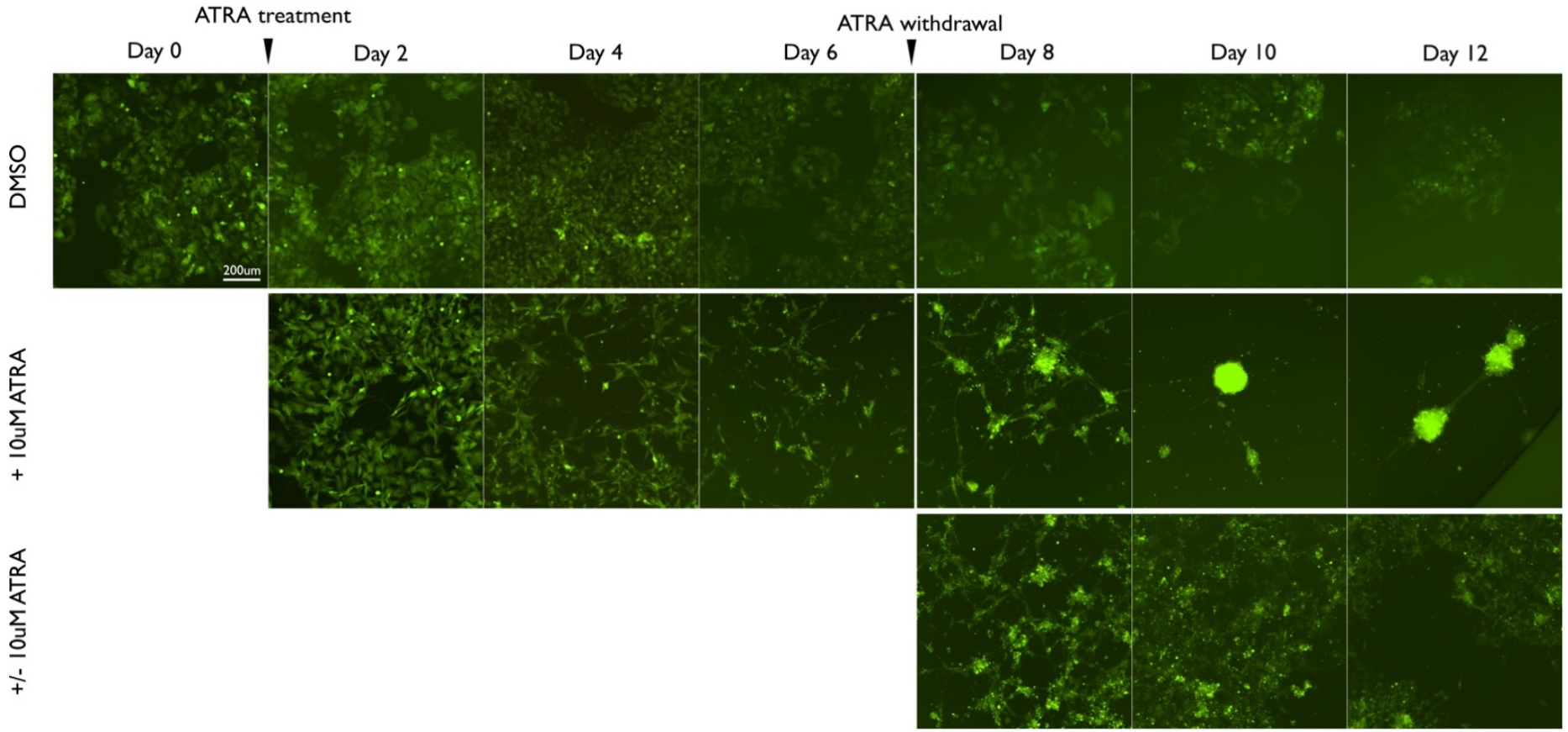


Figure 43: S/MAR expression is maintained during neuronal differentiation and de-differentiation.
Established Be2C neuroblastoma cells were labelled with pSMARt_GFP and treated with either 10uM ATRA or DMSO (untreated control). At day 4, ATRA-treated cells showed morphological changes such as elongation and neurite extensions as compared to the untreated control. At day 10, the cells stopped dividing and clustered into "ganglion-like" structures, which still expressed GFP. At that point, ATRA was withdrawn and the cells recovered their proliferation and original cancer phenotype, still expressing GFP. The cells expressed GFP throughout the experiment, which suggests that the pSMARt vector survived a "differentiation" and "de-differentiation"-like processes.

5.2.3 S/MAR vectors are kept at low copy number

After labelling Be2C cells with the respective pSMARt vectors, we determined how many copies of the vector were present per cell. For this, a copy number assay was performed in the GFP and Luciferase Be2C-expressing clones using an absolute quantification (qPCR) of the transgenes compared to the housekeeping gene (GAPDH). This allowed inferring the number of cells present in the sample as well as the number of copies of each vector. The generated pSMARt_GFP (Be2C 18c2 and 18c4) and pSMARt_Luc (Be2C 50c5 and 50c6) showed low vector copy number, ranging from 2 to 4 copies per cell (**Figure 44**).

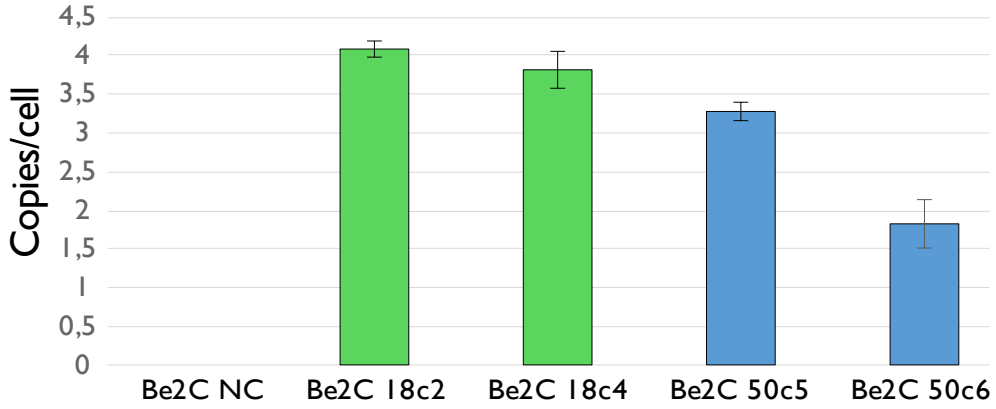


Figure 44: Determination of S/MAR vector copy number in labelled neuroblastoma cells.
Genomic DNA of Be2C clones labelled with pSMARt-GFP (18c2 and 18c4) or pSMARt-Luciferase (50c5 and 50c6), was purified and used for absolute quantification of GFP and Luciferase and compared to the expression of the housekeeping gene GAPDH. The analyzed clonal populations showed low copy numbers of the pSMARt vectors, ranging from 2 to 4 copies/cell.

5.2.4 S/MAR vectors remain episomal

To ascertain whether pSMARt vectors were kept episomal or integrated, a plasmid rescue experiment was performed. This method relies on the ability of circular DNA (but not linear DNA) to transform bacteria. For this, total DNA of stably labelled pSMARt_GFP Be2C cells was isolated, and the circular DNA fraction was enriched and used for electroporation of DH10b competent cells, which were then cultured and expanded. Finally, the plasmid was purified in a mini-prep and digested with *BglII*. The presence of intact S/MAR and GFP-2A-Puro units in the rescued vectors was confirmed by PCR amplification and sequencing of the amplicons **(Figure 45).** Also, total DNA of HEK293T cells labelled with the same pSMARt_GFP vector (lanes 1-5) was also used in this experiment, to gather evidence from different cell lines.

All clones contained an intact S/MAR and transcription unit (GFP-2A-Puro) as shown by PCR amplification. The restriction analysis revealed that clones 7, 8, 11-14 had an identical band pattern compared to the original vector, confirming then a successful 42,85% rescue efficiency of the vectors.

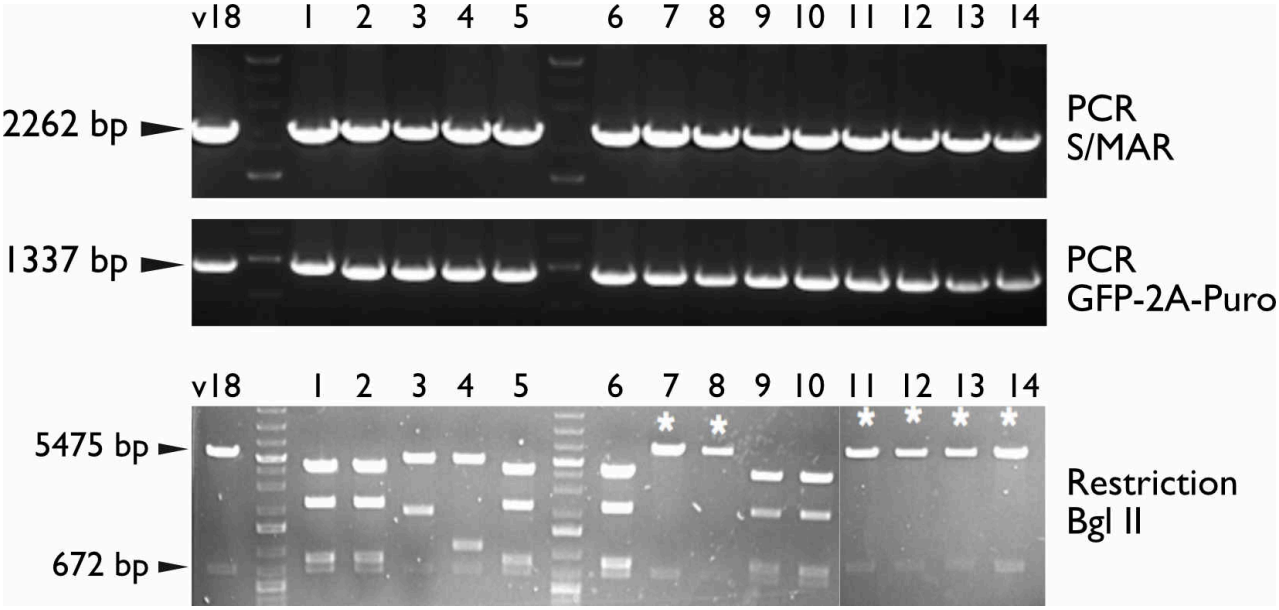


Figure 45: pSMARt_GFP can be rescued as an episome from stably transfected Be2C cells.
HEK293T (lane 1-5) or Be2C (lane 6-14) cells were labelled with pSMARt_GFP (vector 18). Total DNA from stable cells was extracted, and plasmid DNA was purified and used to transform DH10b competent cells. Transformed colonies were grown, and plasmid DNA was prepped and digested with *BglII* or used as a PCR template to amplify both the S/MAR region and the transcription unit GFP-2A-Puro. The S/MAR motif, as well as the GFP-2A-Puro unit, were present in all clones. Clones 7, 8, 11-14 showed an identical band pattern after restriction as compared to the original vector 18, which indicated a successful plasmid rescue and therefore, proved the episomal existence of pSMARt_GFP in stably transfected cells.

5.2.5 S/MAR vectors do not modify the cells' behaviour

5.2.5.1 S/MAR-labelled cells retain the expression of neuronal markers

Following up on previous findings of the minimal impact of S/MAR vectors in the cell transcriptome **(Figure 27),** we evaluated the expression of three neuronal markers (N-Cam, Gap43 and β3-Tubulin). Immunofluorescence stainings were performed in Parental or pSMARt_GFP labelled Be2C cells, before and after differentiation and in DMSO or ATRA-treated cells **(Figure 46).** Both parental, as well as GFP cells, showed expression of all three neuronal markers (in red) and its expression was independent of the treatment. Also, the GFP expression was kept during differentiation as observed by endogenous transgene expression (green).

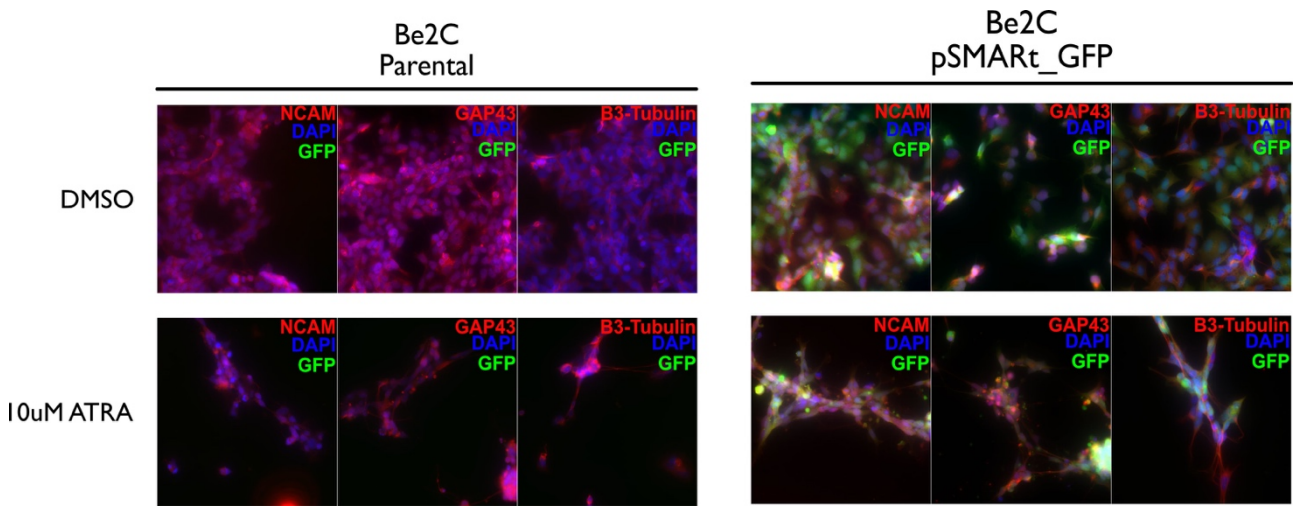


Figure 46: pSMARt-labelled Be2C cells express neuronal markers

Parental as well as pSMARt_GFP labelled Be2C were stained for three neuronal markers: Gap43, N-Cam and β3-Tubulin, before and during differentiation, under DMSO or ATRA treatments. Both parental and S/MAR labelled cells showed expression of the neuronal markers (red), suggesting that the S/MAR presence does not modify the cell's identity. Expression of GFP was kept during differentiation as observed by endogenous GFP expression (green). Nuclei were stained with DAPI (blue).

5.2.5.2 S/MAR-labelled cells engraft when injected into SCID mice

We then evaluated whether pSMARt modified Be2C cells would retain their tumorigenic and engrafting potential when heterotopically injected into SCID mice. In collaboration with Domenic Hartmann, either 10^6 or $5x10^6$ Be2C cells labelled with pSMARt_GFP or pSMARt_Luciferase were injected subcutaneously into SCID mice. Parental (unlabelled) Be2C cells were used as a control. As soon as day 4 post-injection, tumour masses started developing and were measured every three days *[tumour volume (cm^3) = width x length x depth]* until the tumours reached a critical size of 15mm in diameter, at that point the animals had to be sacrificed. The tumour growth is shown in **Figure 47.**

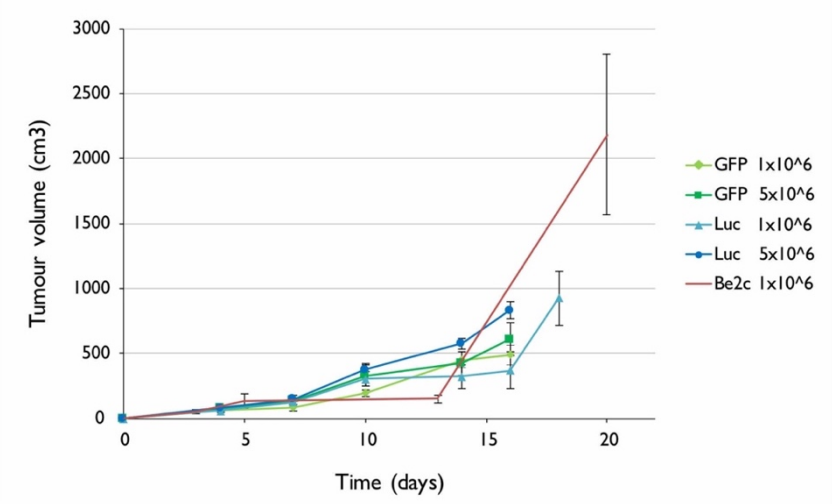


Figure 47: pS/MARt-labelled Be2C cells engraft when injected into SCID mice.

Either 10^6 or $5x10^6$ Be2C cells labelled with pS/MARt_GFP (v18) or pS/MARt_Luciferase (v50) were injected subcutaneously into SCID. The neuroblastoma cells engrafted and formed tumours as early as day 4 and kept growing until reaching a critical mass at day 16. The tumour size was measured every three days. Be2C parental cells were injected in a separate experiment.

Also, in collaboration with Sascha Dehler and Prof. Dr Ana Martin-Villalba from the Molecular Neurobiology department (DKFZ), 500.000 pSMARt_GFP or pSMARt_Luciferase Be2C cells were stereotactically into the striatum of SCID mice. After two weeks, injected mice showed impaired motor abilities and were sacrificed at that point. The brains were removed and directly imaged to check for luciferase expression **(Figure 48a)** or fixed, sectioned and imaged for GFP expression, without the need to stain **Figure 48b.**

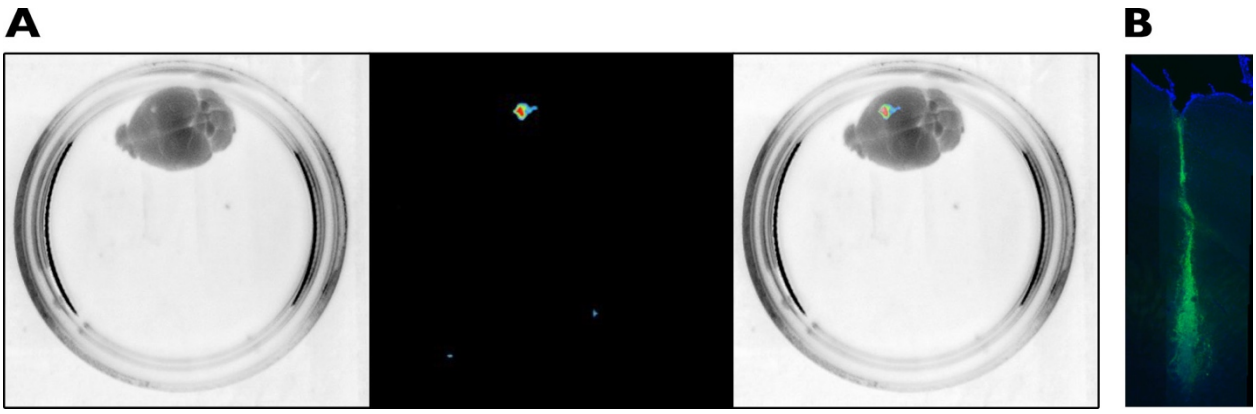


Figure 48: pSMARt-labelled Be2C cells form tumours when injected when injected intracranially
pSMARt_Luciferase and pSMARt_GFP-labelled cells (500.000 cells/mice) were injected stereotactically into the striatum of SCID mice. After two weeks, the mice showed impaired motor abilities and were sacrificed at that point. The brains were removed and directly imaged upon Luciferin administration **(A)** or fixed, sectioned and imaged directly for GFP expression using DAPI as counterstaining **(B).**

5.2.5.3 S/MAR-labelled tumours express the transgene

Images of the cultured neuroblastoma cells expressing the reporter genes were taken before injection **(Figure 49a).** After removal, the tumours were either snap-frozen for further molecular analysis or homogenised and kept in culture. Primary tumours cells kept expressing the reporter genes GFP and Luciferase after engrafting **(Figure 49b).** Also, protein extracts from snap-frozen tumour samples were used to determine GFP expression in a Western Blot **(Figure 50).** GFP could be detected in all pSMARt_GFP labelled tumours whereas the reporter gene expression was undetectable in the parental Be2C cells.

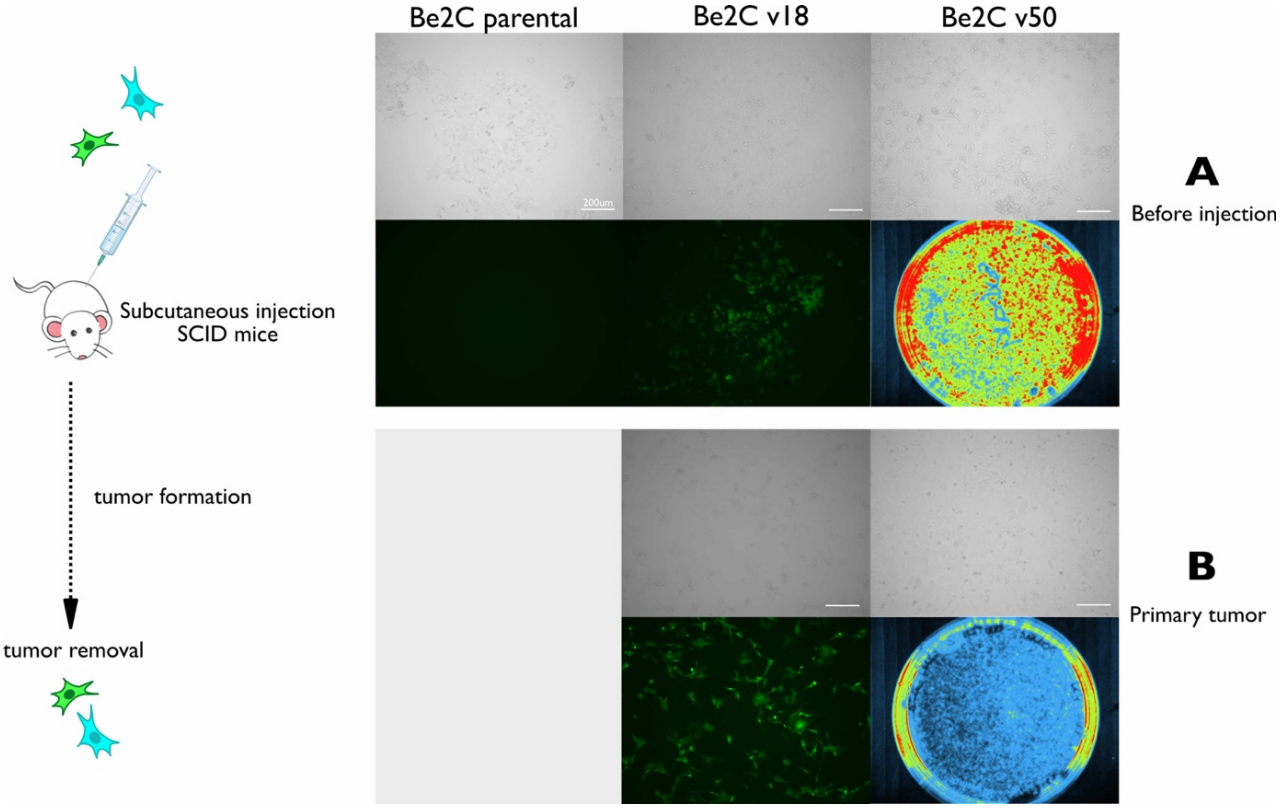


Figure 49: pS/MARt-labelled Be2C cells retain transgene expression after forming tumours.
Be2C cells labelled with pS/MARt_GFP (v18) or pSMARt_Luciferase (v50) were injected subcutaneously into SCID mice. Images of cultured cells expressing the reporter genes were taken before injection **(A).** After two weeks, the animals developed tumours and were sacrificed after reaching a critical tumour mass of 15mm. The tumours were then removed, homogenised and the primary tumour cells were kept in culture. After engrafting and removal, pSMARt-labelled cells remained expressing the reporter gene **(B).**

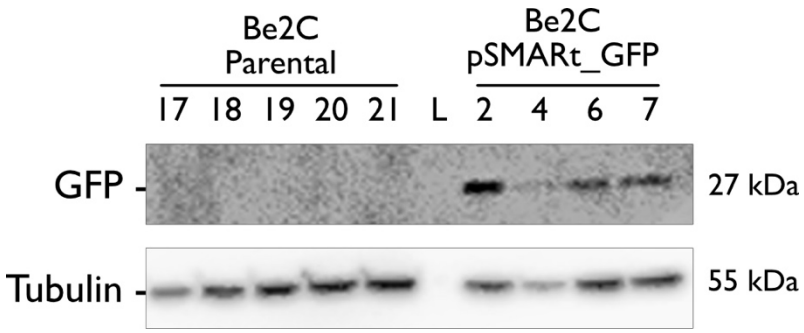


Figure 50: pSMARt-labelled Be2C xenografts retain expression of reporter genes.
Total protein from tumours developed upon injection of Be2C pSMARt_GFP cells was used for Western Blot analysis. The reporter gene GFP (27kDa) was detected in all pSMARt_GFP tumours analysed, whereas no GFP could be detected in parental Be2C tumours. Tubulin (55 kDa) was used as loading control. *Credit: Julia Hermann, DNA Vector Lab, DKFZ.*

5.2.6 Summary

To elucidate whether S/MAR vectors could retain the levels of transgene expression as well as to remain episomal throughout differentiation and de-differentiation, a neuronal differentiation model induced by All-Trans Retinoic Acid (ATRA) was used. For this, Be2C neuroblastoma cells were stably labelled with either pSMARt_GFP or pSMARt_Luciferase vectors, which robustly and persistently express the transgene. The DNA vectors remained episomal in neuroblastoma cells, as they could be rescued in their circular form, and neither altered the cells' behaviour *in vitro* nor *in vivo*. We evaluated the ability of the S/MAR DNA vectors to sustain expression during a differentiation-like process, in which the cells were forced to differentiate into neurons upon addition of ATRA. Neuronal traits, including elongation and neurite extensions, were apparent as early as day 4 and by day 10, the cells clustered into neurospheres and stopped proliferating. At that point, the cells still expressed robust levels of the transgene.

Also, we evaluated the DNA Vector's ability to be retained throughout the reverse process in which the neuron-like cells were allowed to regress into to their original neuroblastoma state by withdrawing the ATRA. They recovered their cancer-like morphology and maintained the expression of GFP. In addition, immunocytological analysis of parental and labelled cells showed no difference in the expression of neuronal markers, suggesting that the vectors did not modify the cells' molecular profile. Finally, we engrafted genetically labelled cells by subcutaneous or intracranial injections into SCID mice. These cells formed representative tumours, with robust expression of the reporter genes.

5.3 Immortalisation, labelling and reprogramming of murine fibroblasts into iPSC with S/MAR vectors – Do S/MAR vectors survive reprogramming?

In this chapter, the use of S/MAR vectors is further challenged by testing their suitability and performance in a real reprogramming process, from murine somatic cells (*e.g.*: fibroblasts) into induced pluripotent stem cells (miPSC). For this, mouse primary lung fibroblasts were first GFP-labelled and SV40LT-immortalised to get enough starting material for reprogramming. Then, the immortal GFP-fibroblasts were reprogrammed, first using the Lentivirus 4in1 and then testing the S/MAR reprogramming vectors POP and nPOP. An additional reprogramming efficiency comparison between Lentivirus and S/MAR DNA vectors in human dermal fibroblasts is shown in **Section 5.6.2.2.**

5.3.1 Immortalisation and labelling of mouse fibroblasts

Primary murine lung fibroblasts, kindly provided by Dr Joschka Willemsen (Binder Lab, DKFZ), were nucleofected using the bifunctional labelling-immortalising pSMARt_SV40LT vector. The cells were grown in the absence of selection, as successfully transfected cells would possess a selective 'immortal' advantage over the untransfected ones. The cells were cultured and remained in division for over 100 passages. Due to the presence of GFP in the immortalising S/MAR vector, the immortal cells also expressed GFP as compared to primary untransfected fibroblasts **Figure 51.**

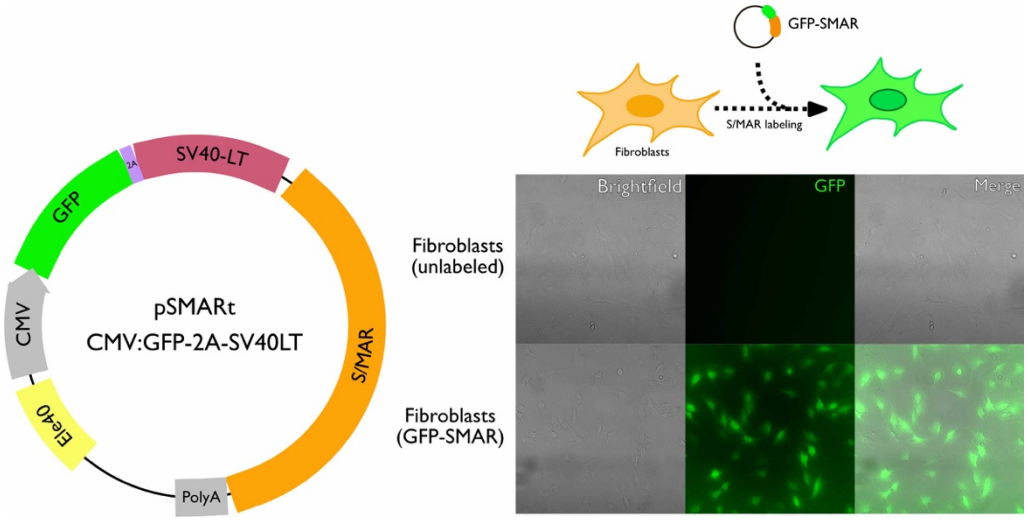


Figure 51: S/MAR vectors can label and immortalise murine fibroblasts.
A bifunctional pSMARt_SV40LT vector expressing GFP was nucleofected into murine primary lung fibroblasts. Successfully transfected cells not only expressed GFP but could also be kept in culture for more than 100 passages, confirming immortalisation.

5.3.2 Do S/MAR vectors survive reprogramming?

5.3.2.1 Reprogramming S/MAR labelled fibroblasts with Lentivirus

To answer whether S/MAR vectors were able to survive a real reprogramming process from somatic cells into iPSC, pSMARt_SV40LT-immortalised fibroblasts were transduced with 4in1 Lentiviral particles containing the OKSM reprogramming factors. Wild-type lung fibroblasts were used as negative control. Due to its higher transduction efficiency, we decided in favour of lentiviral reprogramming over S/MAR reprogramming (*e.g.*: POP, nPOP vectors). The reprogramming capabilities of these vectors are tested below. After transduction, the fibroblasts were provided with the right culturing conditions to allow them to undergo a mesenchymal-epithelial transition (MET) and reprogramming. After a couple of weeks, the original elongated-mesenchymal morphology typical from fibroblasts transitioned into a smaller and more compact epithelial shape, with the cells clustered in dome-like colonies (miPSC colonies). Those miPSC colonies retained the expression of GFP, also at the pluripotent stage, as compared to the unlabelled primary fibroblast control (**Figure 52**). Both iPSC expressed dTomato as proof that the reprogramming factors were being expressed. After reaching pluripotency, the expression of dTomato was switched off (not shown) in wildtype cells but not in immortalised cells.

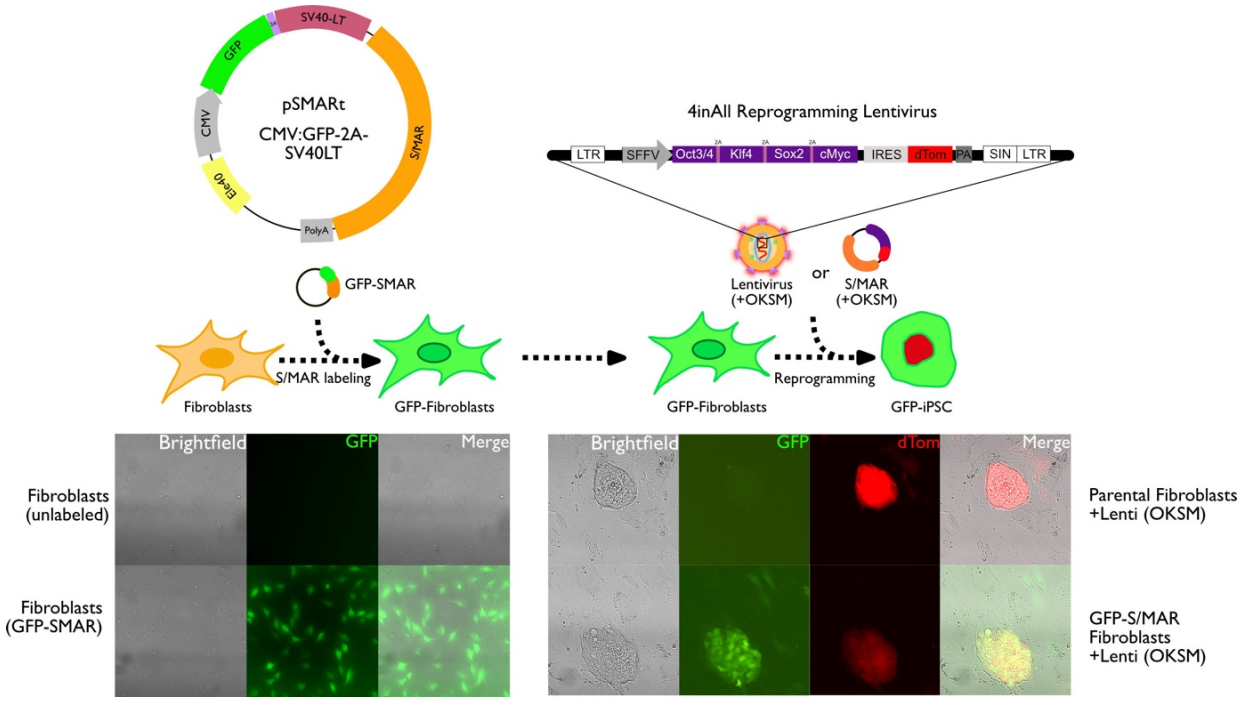


Figure 52: pSMARt vectors remain expressed through reprogramming.
Wildtype (negative control) and pSMARt_SV40LT-immortalised fibroblasts were transduced using the 4in1 Lentivirus coding for the Yamanaka Factors OKSM. The fibroblasts were kept under reprogramming conditions until dome-like iPSC colonies were observed after a couple of weeks. The iPSC colonies were imaged and checked for reporter gene expression. Both labelled, and unlabelled iPSC colonies expressed the reprogramming factors, as indicated by the expression of dTomato. pSMARt_SV40LT fibroblasts kept expressing GFP, suggesting that S/MAR vectors survived the reprogramming process.

5.3.2.2 Reprogramming mouse embryonic fibroblasts (MEFs) with the S/MAR reprogramming vectors POP and nPOP

Because S/MAR-labelled cells could sustain the expression of GFP throughout reprogramming after Lentiviral delivery of the reprogramming factors, we then tested a safer approach by using the S/MAR vectors as reprogramming tools.

For that, the aforementioned S/MAR reprogramming vectors POP and nPOP were nucleofected into low passage commercial murine embryonic fibroblasts (MEFs), transducing the Lentivirus 4in1 as a positive control. However, the transfection efficiency of the S/MAR reprogramming vectors POP and nPOP was low, as we could barely observe dTomato-positive cells 24hpt. Therefore, no miPSC colonies formed, due to insufficient expression of reprogramming factors (data not shown). To increase the transfection efficiency and the delivery of reprogramming factors, repeated transfections of POP and nPOP were performed. Nucleofected MEFs were then kept under puromycin selection to remove untransfected cells, and dTomato-positive cells were placed under reprogramming conditions until miPSC colonies could be observed after four weeks. We observed that the morphology of nPOP-reprogrammed miPSC was more similar to that obtained using the Lentiviral 4in1 vector, whereas POP miPSC had a less defined colony morphology **(Figure 53).**

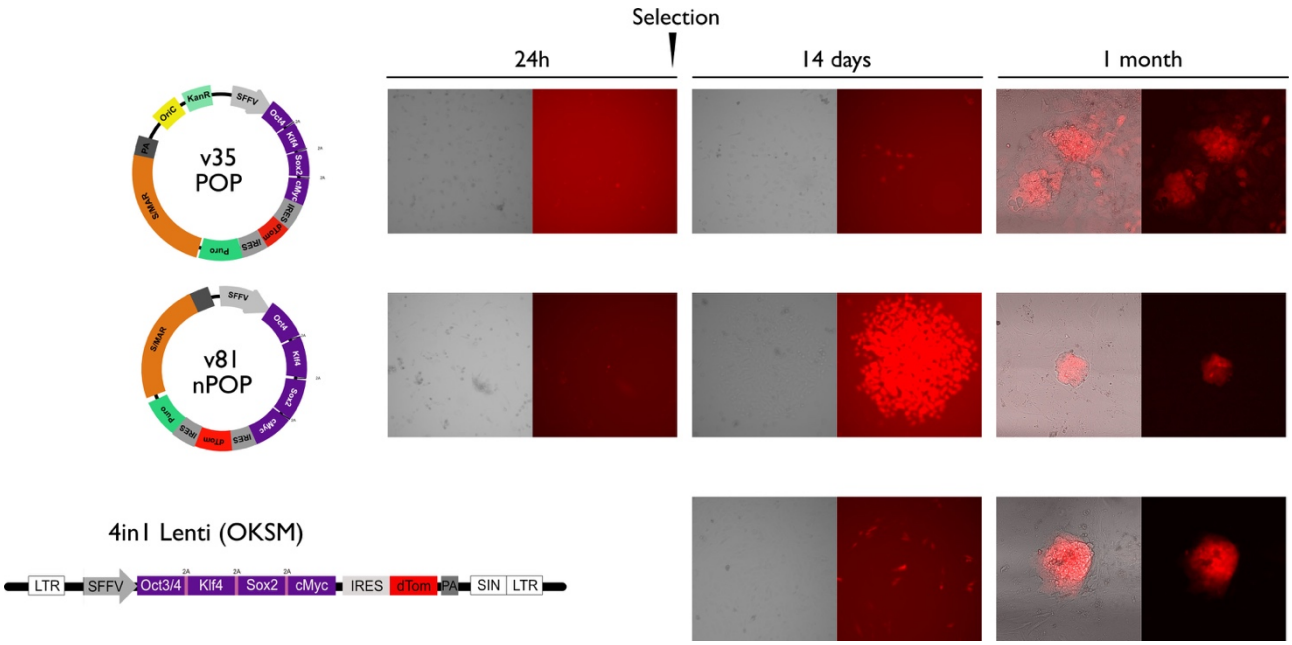


Figure 53: The S/MAR reprogramming vector nPOP can reprogram murine fibroblasts into iPSC.
The ability of the S/MAR reprogramming vectors to turn MEFs into miPSC was assessed by nucleofecting the S/MAR-based reprogramming vectors POP and nPOP. Transduction with 4in1 lentivirus was used as positive control. Two repeated transfections were performed to increase the transfection efficiency. The cells were selected with puromycin and allowed to reprogram for a month. Although the transfection efficiencies were lower than viral transduction, miPSC colonies emerged after four weeks.

5.3.3 Summary

In this chapter, the ability of S/MAR vectors to survive reprogramming from somatic cells (*e.g.*: fibroblasts) to induced pluripotent stem cells (miPSC) was challenged. To ensure enough starting material, murine lung fibroblasts were first immortalised using a bifunctional labelling and immortalizing pSMARt_SV40LT vector. Then, immortal GFP-fibroblasts were reprogrammed first by using 4in1 Lentiviral particles and later, by testing the reprogramming ability of the S/MAR reprogramming vectors themselves. pSMARt-immortalised fibroblasts were able to maintain the reporter gene expression throughout the reprogramming process, suggesting that the vector was not lost nor silenced. Also, mouse embryonic fibroblasts could be reprogrammed using the nPOP vector, although the reprogramming was not as efficient as with Lentiviral particles. Although possible, the low reprogramming efficiency of nPOP was mostly due to the low transfection efficiency, possibly due to MEFs being refractory to transfection.

5.4 Labelling and *in vitro* differentiation of murine Stem Cells with S/MAR vectors – Do S/MAR vectors survive differentiation?

The use and suitability of S/MAR-based vectors for labelling dividing and differentiating cells has been tested in the aforementioned ATRA-mediated differentiation and de-differentiation model using neuroblastoma cells, suggesting that S/MAR-based vectors were capable of surviving differentiation. In this chapter, the use of S/MAR vectors in a real differentiation process from murine embryonic stem cells to somatic cells is validated. For this, mESC were labelled with either pSMARt or nSMARt vectors, and stable GFP-expressing mESC lines were generated. Then, basic molecular analysis were performed to determine the copy number and the integration or episomal status of the vector. The pluripotency and differentiation abilities of either 1) mESC labelled with pSMARt or 2) pSMARt_SV40LT reprogrammed miPSC were also tested; as well as the ability of the vector to survive differentiation. For this, immunofluorescence stainings for murine pluripotency markers were performed. Additionally, pSMARt-mESC cells were forced to differentiate either randomly into representatives of the three germ layers or directly into hematopoietic progenitors by using specific cytokines. The expression of GFP and presence of the S/MAR vector, were evaluated through differentiation.

5.4.1 Labelling of mESC with pSMARt and nSMARt vectors.

The suitability of pSMARt or nSMARt to transfect and label mESC has been shown in **Section 5.1.7.1** and **5.1.7.2.** We then generated stable GFP-mESC lines using these vectors to evaluate their ability to remain active through differentiation. For this, 10 μg of either pSMARt or nSMARt were delivered using the Amaxa II nucleofector and the mESC kit (Lonza). The cells were kept under Puromycin selection for a couple of weeks to remove untransfected cells as well as to favour the retention of the S/MAR vectors. After removal of the selective pressure, the cells kept expressing GFP for months after and until today, as shown in **Figure 54.**

Clones from both cell lines were picked and expanded and one clone each was chosen for further analysis. The pSMARt clone v71c22 homogeneously expressed GFP, whereas clone v85c17 (nSMARt) was somewhat heterogeneous due to selection and picking of a mixed mESC colony, which was generated from cells expressing various levels of GFP.

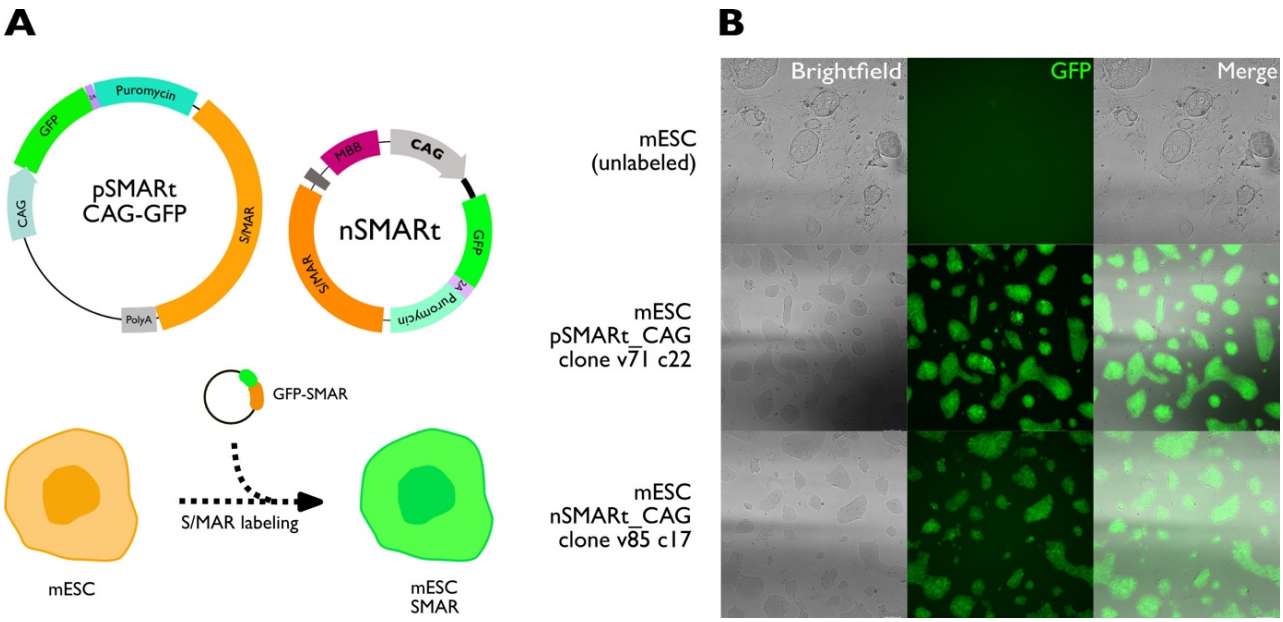


Figure 54: pSMARt and nSMARt can persistently modify mESC
A representative cartoon showing the mESC labelling process using pSMARt and nSMARt **(A).** For this, 500.000 mESC were nucleofected with 10ug of plasmid DNA and kept under selection with 500ng/ul Puromycin for two weeks. Afterwards, the selection was removed, and the cells kept expressing GFP for months after, until today. Then, mESC clones from both cell lines were picked and expanded and their GFP expression is shown **(B).**

5.4.2 S/MAR Vectors are kept at low copy number

The selected mESC clones labelled with pSMARt (v71c22) and nSMARt (v85c17) were then analysed to determine the number of vector copies per cell. For this, total DNA from a confluent 6cm-plate was extracted using a DNeasy Blood and tissue kit (QIAGEN) or Phenol/Chloroform. Standard curves for both genomic and plasmid DNA were prepared to be able to quantify the number of cells and vector in each the sample. The vector was quantified by amplification of Puromycin (Primers Puro3-4), and the cells were quantified by amplification of the housekeeping gene GAPDH (primers mGAPDH3-4). The number of copies per cell was calculated as explained in **Section 4.2.9.1** and untransfected mESC were used as negative control. We observed that S/MAR vectors were kept at low copy number (between 1-2 copies/cell).

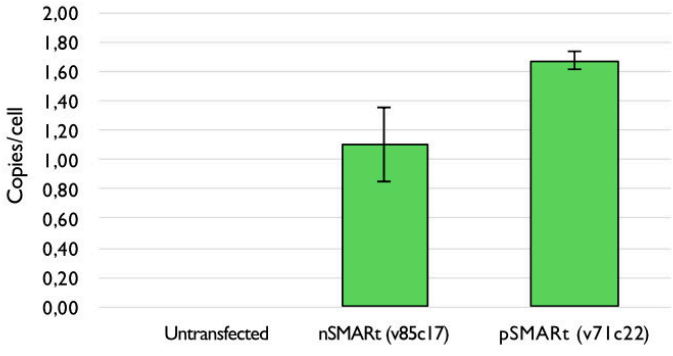


Figure 55: pSMARt and nSMARt vectors are kept at low copy numbers
Total DNA from mESC clones (pSMARt v71c22 and nSMARt v85c17) was purified and used for absolute quantification of Puromycin as compared to a constitutively active gene GAPDH. The results showed more copies/cell in the homogeneous v71c22 clone, whereas the copy number is underestimated in the heterogeneous clone v85c17.

5.4.3 S/MAR vectors remain episomal

Next, we performed a **plasmid rescue experiment** and a **southern blot** to exclude genomic integrations and prove the vectors' episomal maintenance. **Integration analysis** of pSMARt labelled mESC clones and polyclonal populations is currently being performed in collaboration with Prof. Dr Christof von Kalle, Dr Manfred Schmidt and Dr Irene Gil-Farina, from NCT, but the results could not be included in this thesis due to time constraints.

5.4.3.1 Plasmid rescue of pSMARt episomes from mESC-labelled cells

Total DNA from pSMARt labelled mESC was extracted using phenol-chloroform and 10µg of genomic DNA were then digested with the vector non-cutter restriction enzyme (*EcoRV*). The undigested pSMARt was purified and transformed into DH10ß competent cells. Then, bacterial colonies were picked, miniprepped and digested using the restriction enzyme *BglII*. The restriction pattern of rescued clones was compared to the original pSMARt vector. Also, the integrity of the S/MAR motif as well as the expression cassette containing GFP-2A-Puro was confirmed by PCR amplification, as shown in **Figure 56.** A total of 9/10 rescued clones showed identical restriction pattern as compared to the original pSMARt (v71) vector as well as successful PCR amplification of both vector regions, suggesting that the vector integrity is not disturbed upon establishment in mESC.

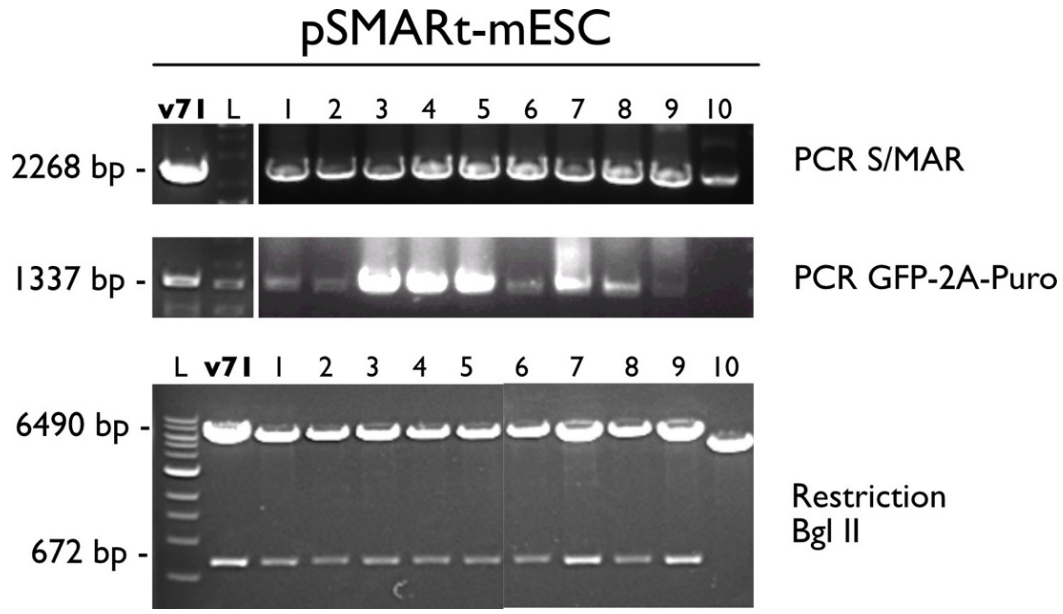


Figure 56: Episomal forms are detected in stably transfected mESC
Total DNA from pSMARt labelled mESC was extracted using phenol-chloroform and digested with the restriction enzyme *EcoRV*, which did not digest pSMARt. Then, circular vectors were purified and transformed into competent cells. The grown clones were miniprepped and digested with *BglII*. Their restriction pattern was compared to that from the original pSMARt vector (v71). The integrity of the vector was confirmed by PCR amplification of both the S/MAR motif (2268bp) and the expression cassette (1337bp). The successful rescue of pSMARt was finally confirmed by sequencing the rescued clones.

5.4.3.2 Southern Blot of S/MAR vectors from mESC-labelled cells

40 µg of total DNA extracted with Phenol/Chloroform was digested with a vector unique-cutter enzyme (*BamHI*, common in all S/MAR vectors), to linearise the plasmids. Additionally, 2-5ng of parental plasmids were also digested as positive controls. Then, digested DNA was ethanol-precipitated, and 100ng was loaded on a 0,8% agarose gel, which was run overnight at 15V. The DNA was transferred to a Hybond-XL nylon membrane and hybridised with GFP probes. Linearised plasmids (**Figure 57,** left blot) showed sharp bands at their expected sizes: pSMARt (7162bp), nSMARt (5915bp), nSpliced (6546 bp) and pSMARter (6085bp). The amount of nSMARt loaded was underestimated, and the band did not appear as intense as the other vectors. However, only nSpliced-labelled cells showed a faint band (**Figure 57,** right blot), suggesting that the S/MAR was present in the sample. The upper left part of the blot showed some unspecific background, with an intense spot over Untransfected and pSMARt-labelled cells.

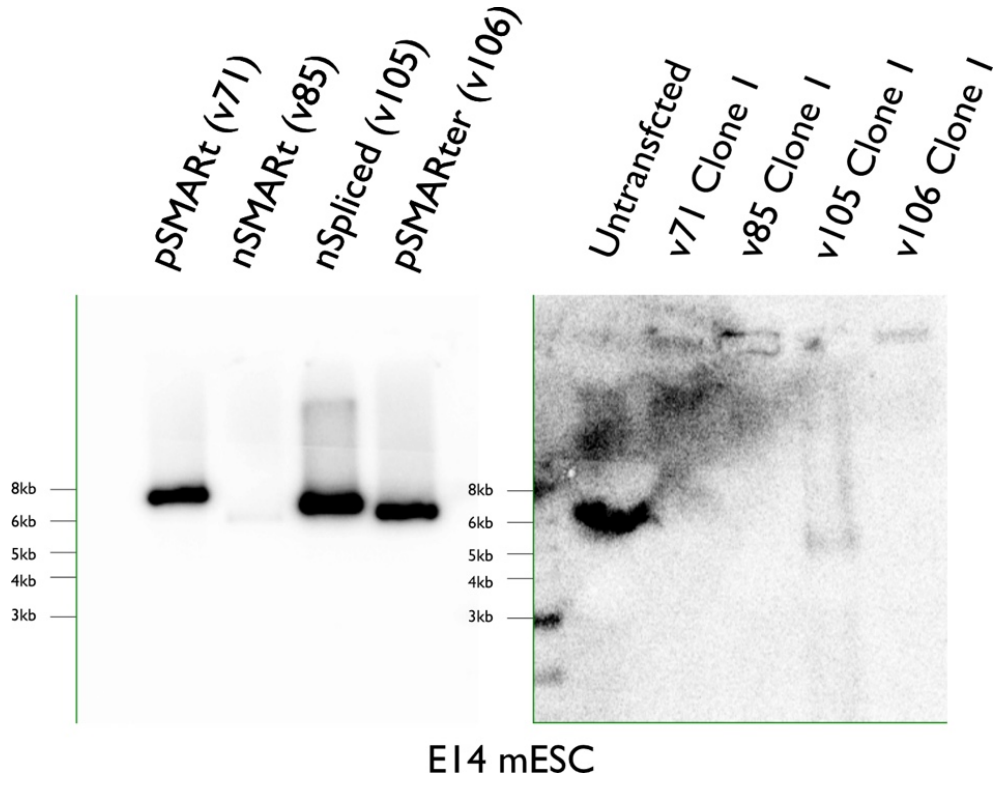


Figure 57: Integrated forms cannot be detected in S/MAR-labelled mESC clones.
Total DNA from S/MAR-labelled cells (40ug) was digested with a unique vector cutter (*BamHI*). Parental plasmids were used as positive controls. Then, 100ng of digestion product, including linearised episomes, were loaded on a 0,8% agarose gel and run overnight at 15V. The DNA was transferred into a membrane and hybridised with GFP probes. Positive control plasmids (left blot) showed sharp bands at their expected sizes, although the amount of nSMARt was somehow underestimated. Despite the background of the blot, only nSpliced could be detected from mESC cells (right blot).

5.4.4 S/MAR vectors do not alter pluripotency in murine stem cells

Next, the functionality and pluripotency of generated pSMARt-labelled mESC and miPSC were assessed. For this, the expression of pluripotency markers was analysed via immunofluorescent staining. Then, S/MAR labelled cells were subjected to *in vitro* differentiation, either randomly or directed towards the hematopoietic lineage with the help of cytokines. The expression of GFP was monitored and quantified before, during and after differentiation.

5.4.4.1 Assessment of pluripotency in S/MAR-pluripotent cells

The pSMARt_SV40LT-immortalised lung fibroblasts reprogrammed into miPSC using the Lentivirus 4in1, as well as the pSMARt-labelled mESC, were subjected to Alkaline Phosphatase (AP) staining. This hydrolytic enzyme, which is responsible for dephosphorylating molecules under alkaline conditions, is highly expressed in undifferentiated cells with self-renewal potential. After fixation and staining with Alkaline Phosphatase staining kit II (Stemgent), undifferentiated cells appeared stained in purple while differentiated cells appeared colourless. Also, murine pluripotent cells are characterised by expression of a set of pluripotency markers, amongst them: the surface marker SSEA-1 as well as the endogenous transcription factors Oct4 and Nanog. The expression of these pluripotency markers was also assessed by performing immunofluorescence stainings as shown in **Figure 58**. Unlabelled mESC, as well as unlabelled miPSC, were used as negative controls. All pSMARt-labelled cells expressed GFP, as shown by the fluorescent images obtained before staining. Control and pSMARt-labelled mESC as well as control miPSC showed expression of all pluripotency markers tested. However, pSMARt_SV40LT immortalised and reprogrammed iPSC did not stain positive for the pluripotency marker Nanog and the cells kept expressing the exogenous reprogramming factors, as observed by dTomato expression.

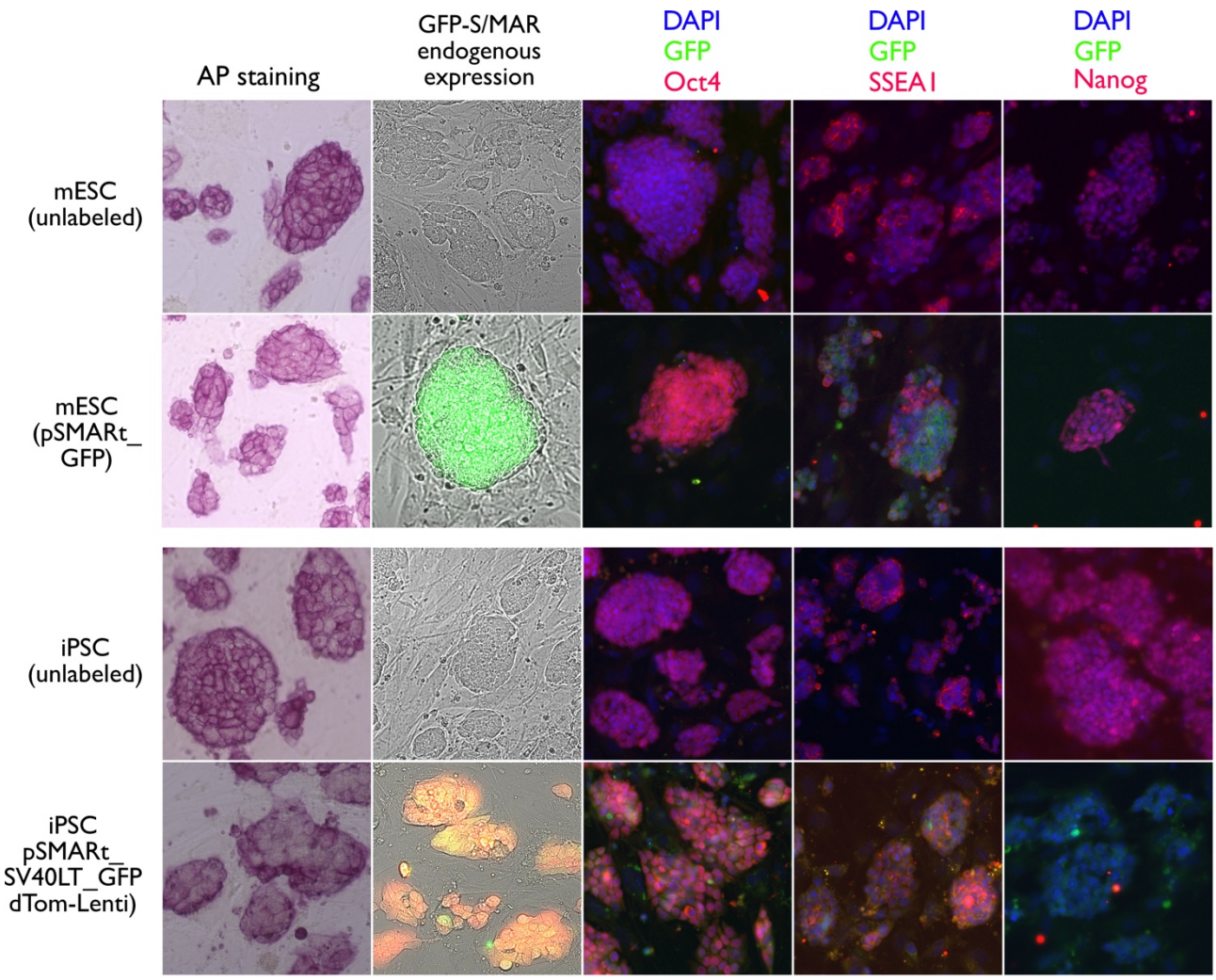


Figure 58: S/MAR-modified stem cells express pluripotent markers

A panel of pluripotent markers was assessed in unlabelled and pSMARt-labelled mESC and miPSC. This included Alkaline Phosphatase staining and immunofluorescence for the surface marker SSEA-1 and the transcription factors Oct4 and Nanog. After fixation and staining, both unlabelled controls and S/MAR-labelled mESC showed expression of all pluripotent markers tested. However, immortalised iPSC did not show expression of Nanog and retained the expression of endogenous reprogramming factors, as indicated by the reporter gene dTomato, suggesting partial reprogramming.

5.4.4.2 Random differentiation of S/MAR-mESC via embryonic bodies

After successful confirmation of pluripotency in both mESC and miPSC, the cells were subjected to a more stringent test consisting on *in vitro* random differentiation into representatives of the three germ layers (ectoderm, mesoderm and endoderm) via the transient formation of embryonic bodies (EBs). For this, the cells were withdrawn from all differentiation inhibitors (2i) and LIF and allowed to collapse into hanging drops, which were later on placed into adherent culture and allowed to attach and spread while differentiating **(Figure 59a).**

During the experiment, the EBs were monitored and imaged regularly to check for GFP expression and therefore, for the presence and functioning of the S/MAR vectors during the differentiation process.

Data collected from two independent experiments revealed that both mESC (unlabelled) and pSMARt-labelled mESC formed compact EBs, which differentiated into different cell types. More importantly, pSMARt-mESC kept the expression of GFP throughout differentiation **(Figure 59b).** Control iPSC reprogrammed with 4in1 Lentivirus, formed EBs and differentiated into distinct structures. However, iPSC generated from pSMARt_SV40LT-immortalised lung fibroblasts, initially formed EBs, which kept the GFP expression **(Figure 59c)**, but failed to differentiate. Immortal cells kept expressing the exogenous reprogramming factors (dTom) and could not differentiate properly.

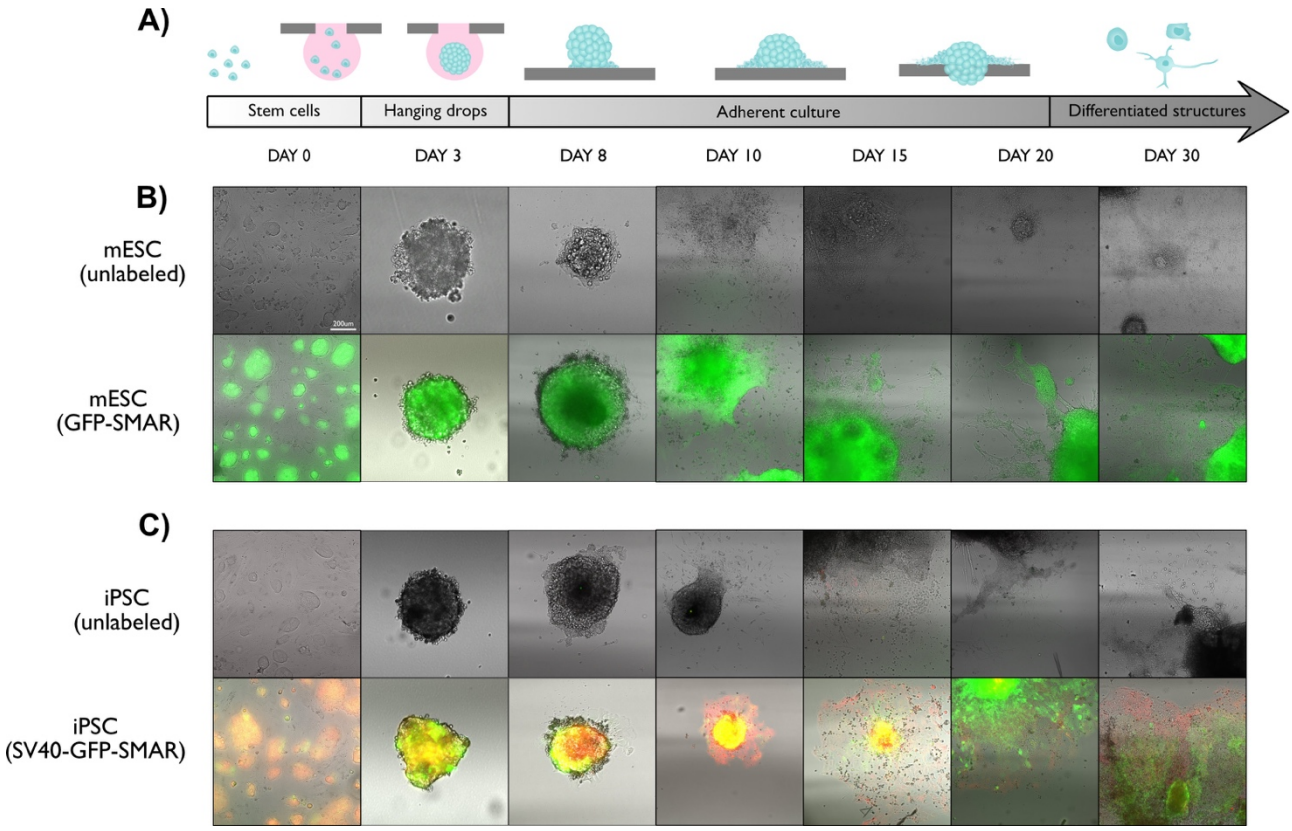


Figure 59: S/MAR vectors are expressed in EBs and through mESC differentiation

To generate Embryonic Bodies (EBs), single stem cell suspensions were allowed to collapse at the bottom of a hanging drop, devoid of differentiation inhibitors (2i) and LIF. Then, the EBs are transferred into gelatinised plates and allowed to attach and spread, while the proliferating cells differentiated into representatives of the three germ layers **(A).** Both unlabelled (negative control) and pSMARt-labelled mESC formed EBs that differentiated into defined structures. pSMARt cells kept the expression of GFP throughout the differentiation process **(B).** Unlabelled miPSC formed EBs, which differentiated and formed distinct structures. However, immortalised miPSC formed EBs but failed to differentiate. The exogenous reprogramming factors remained expressed, as suggested by the expression of dTomato **(C).**

After differentiated structures could be observed microscopically, such as neurons or beating cardiomyocytes, the cells were fixed and stained for ectodermal (ß3-Tubulin), mesodermal (α-SMA) and endodermal (FoxA2) markers **Figure 60.**

Both control and pSMARt-mESC differentiated into representatives of the three germ layers. Also, endogenous GFP expression, although faded due to the fixation of cells, could also be observed. Unlabelled miPSC differentiated into ectoderm and mesoderm, although endodermal structures could not be found. In contrast, immortalised miPSC could not differentiate as they remained expressing reprogramming factors. Therefore, they were excluded from further *in vivo* experiments.

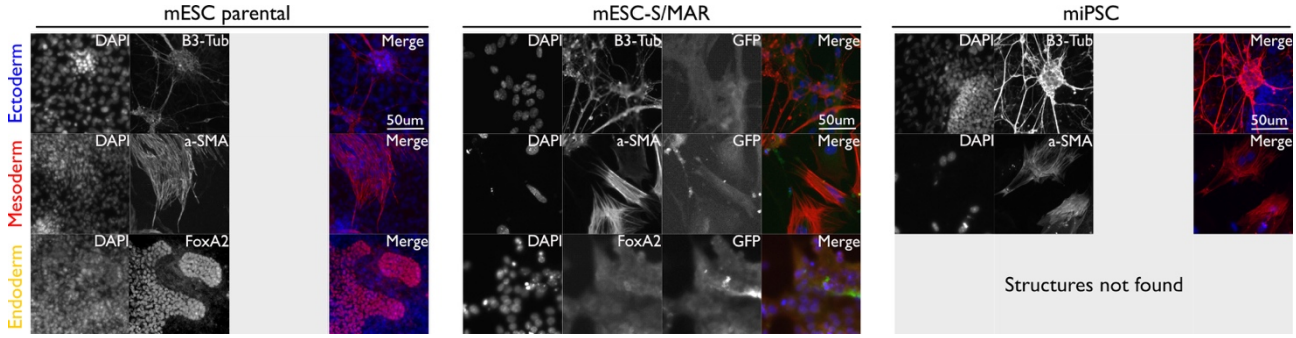


Figure 60: S/MAR-labelled stem cells can differentiate into derivatives from the three germ layers
After formation of EBs, differentiating structures, such as neurons or beating cardiomyocytes, could be observed microscopically. Then the cells were fixed, permeabilised and stained for markers from the three embryonic layers: ectoderm (ß3-Tubulin), mesoderm (α-SMA) and endoderm (FoxA2). Both control and pSMARt-labelled mESC differentiated into all three germ layers. Also, pSMARt-labelled cells showed expression of endogenous GFP. Unlabelled miPSC differentiated into ectoderm and mesoderm structures although no endodermal structures were found. Immortalised miPSC failed to differentiate, as they kept expressing the reprogramming factors.

5.4.4.3 Directed hematopoietic differentiation of S/MAR-mESC

After confirming that GFP was expressed through random differentiation and that mESC could differentiate into all embryonic layers, we then quantified the GFP expression of different S/MAR-labelled cell lines in a directed differentiation setting. For this, an *in vitro* hematopoietic differentiation experiment was performed in collaboration with Marleen Büchler (Dr Milsom Lab, Hi-STEM, DKFZ).

Different lines of stably transfected E14 mESC with pSMARt_CAG (pCAG), nSMARt_CAG (nCAG), nSMAR_Spliced or pSMARter, were forced to collapse into EBs under hypoxic conditions (5%O2), as this enhances the efficacy. Unlabelled E14 cells were used as negative control. All the samples were prepared in triplicates. After 60h, cytokines such as BMP-4 (mesodermal development), Activin A; and growth factors such as VEGF (endothelial lineage) and FGF2 were added into the media to push the cells down the hematopoietic differentiation path. After another 60h, the EBs were collected, disaggregated and stained with a panel of endothelial (c-kit, VE-Cadherin) and hematopoietic markers (CD41) markers for FACS analysis (**Figure 61).**

At the mESC stage (Day 0), the cells were negative for the hematopoietic marker CD41 and showed predominant expression of the endothelial marker c-kit but not CD144/VE-cadherin. After 6 days, the

cell population shifted and acquired expression of the hematopoietic marker CD41 and lost endothelial characteristics, as observed by the loss of c-kit and VE-Cadherin expression. However, expression of c-kit and VE-cadherin should still be observed at the HSC stage [302].

Also, the GFP expression was quantified at the mESCs (day 0) and HSC (day 6) stage. A slight decrease in GFP expression was observed in pCAG and pSMARter-labelled cells, whereas the GFP expression from nCAG and nSpliced remained constant. Interestingly, the decrease in GFP expression corresponded to cells labelled with vectors which contained bacterial sequences and was not observed in minimally sized bacterial backbones or nano vectors. Within the nanovectors, the CAG promoter provided more constant levels of GFP as compared to CMV shielded with Element 40.

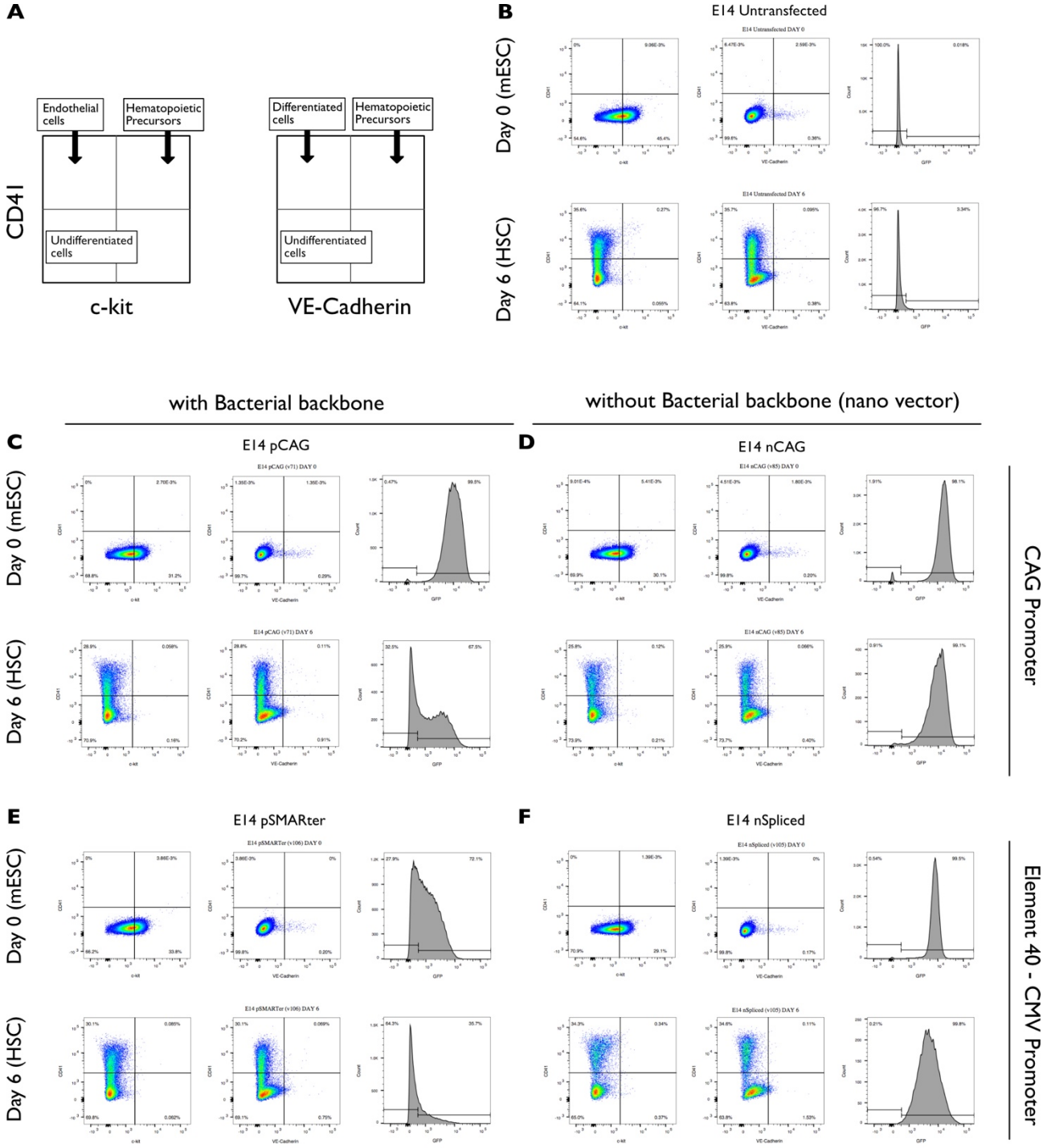


Figure 61: S/MAR-labelled mESC retain expression of GFP during hematopoietic differentiation

Schematic representation of expected populations according to hematopoietic (CD41) and endothelial markers (c-kit and VE-Cadherin) **(A).** E14 mESC were transfected and stably labelled with pCAG (v71), nCAG(v85), nSpliced (v105) and pSMARter (v106). **(C-F).** Untransfected E14 cells were used as negative control **(B).** Before differentiation (day 0), mESC were stained for endothelial (c-kit and VE-Cadherin) and hematopoietic (CD41) markers; and the GFP expression was quantified in the viable gated population. Then, mESC cells were forced to collapse into EBs under hypoxic conditions. Unlabelled E14 cells were used as negative control. All the samples were prepared in triplicates. After 60h, cytokines and growth factors were added into the media to push the cells down the hematopoietic differentiation path. After 120h (day 6), the EBs were disaggregated, counted and stained with the same hematopoietic panel as day 0. The same gating scheme used on mESC was applied to HSC and GFP was quantified in the HSC viable population. pCAG and pSMARter showed a slight decrease in GFP expression, whereas the expression from nano vectors nCAG and nSpliced remained constant.

5.4.5 Summary

This chapter shows that pSMARt vectors can modify murine stem cells without compromising the cells' capabilities. The S/MAR vectors provided high and robust levels of transgene expression during extended periods of time, despite the low number of vector copies per cell. The molecular analysis suggested that the vectors were kept episomal, as they could be rescued from stably labelled mESC and integrations were never detected in a southern blot.

Moreover, pSMARt did not interfere with the stem cells' pluripotent features, as shown by the expression of all pluripotency markers tested as compared to control mESC. The S/MAR vectors were further challenged in a random differentiation experiment, in which both labelled, and unlabelled mESC cells, as well as miPSC, were forced to collapse into embryonic bodies and to differentiate into representatives of the three germ layers. The expression of GFP, which reflected the presence and functionality of the vector, was kept throughout the process. However, immortalised miPSC failed to differentiate and kept expressing the reprogramming factors and therefore, they were excluded from further *in vivo* experiments. The same vector behaviour pattern was observed during directed hematopoietic differentiation using specific cytokines. Here, a range of mESC lines labelled with different S/MAR vectors, including pSMARt, nSMARt, nSpliced and pSMARter were tested and quantified for GFP expression before (mESC) and after (HSC) differentiation. A slight decrease in GFP expression was observed in pCAG and pSMARter, whereas nCAG and nSpliced were able to sustain the levels of GFP throughout differentiation.

5.5 A novel tool to generate transgenic animals – Do S/MAR vectors survive *in vivo* differentiation?

This chapter introduces the use of S/MAR vectors for *in vivo* applications, such as the generation of chimeric and transgenic mice. Most importantly, it shows in a real *in vivo* developmental process that S/MAR DNA vectors can survive and keep their expression throughout a differentiation from an embryo into a whole organism.

There are two methods to generate transgenic animals: either with an **'indirect' or cell-mediated approach** using genetically modified stem cells, which are then implanted into blastocysts; or with a **'direct' DNA-mediated approach,** in which the DNA is directly delivered into the zygote's pronucleus. Based on our expertise in *in vitro* modification of cells with DNA vectors, we first generated chimeric mice using stably pSMARt or nSMARt-labelled mESC that were implanted into embryos. Then, we attempted to generate (iso)transgenic mice by directly injecting pSMARt vectors into the pronucleus of 1-cell stage zygotes. The presence of the vector in the transgenic tissues was evaluated through GFP expression using a variety of techniques. Finally, the suitability of S/MAR vectors for the inheritable modification of organisms was challenged by analysing the germ-line transmission of S/MAR vectors into the offspring of transgenic animals.

5.5.1 Cell-mediated approach – pSMARt labelled mESC cells can form chimaeras

A more stringent pluripotency test is to assess the stem cells' ability to form chimaeras when injected into embryos at the *morulae* or blastocyst stage. In order to investigate whether S/MAR labelled cells were able to do so and to ascertain if the vectors maintained their expression through an *in vivo* differentiation process, two rounds of chimeric mice were generated in collaboration with Franciscus van der Hoeven and Ulrich Kloz, from the Transgenics Service (DKFZ).

First, high-passage pSMARt-labelled mESC from BL6 origin were injected into blastocysts of CD1 albino mice. Although we could observe coat-chimerism in the born pups with black patches resulting from the contribution of our modified mESC, the percentage of chimerism was unsatisfyingly low (data not shown). In a second attempt to generate better chimaeras, a more suitable, low-passage stem cell line (E14 129Ola) was used. E14 cells were transfected either with pSMARt_CAG (pCAG) or nSMARt_CAG (nCAG) and kept under puromycin selection for two weeks. The clones were then picked and expanded, which resulted in the aforementioned clones v71 c22 and v85 c17. These E14

129Ola clones were injected into *morulae* of BL6 embryos and transferred into foster mothers. 49 chimeric pups were born, in which the presence and contribution of E14 cells could be observed by an agouti/chinchilla coat chimerism over the black background **(Figure 62).**

All pups showed varying degrees of chimerism: more than half of the animals had over 50% of coat chimerism, in some cases reaching a 100% chinchilla coat colour, which translates into almost the whole animal being originated from the genetically modified S/MAR-mESC.

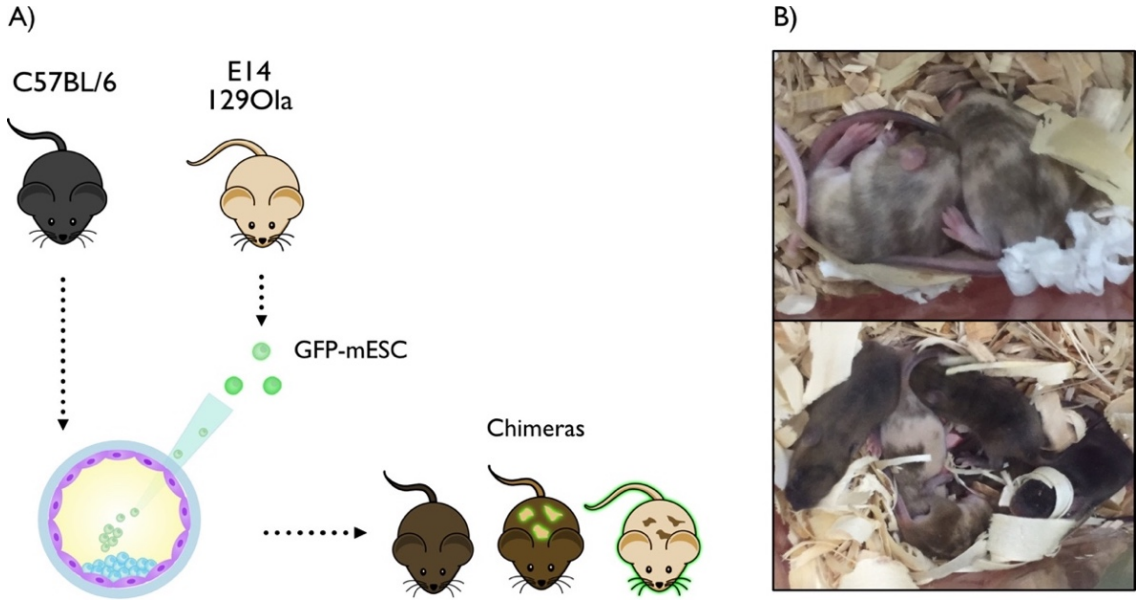


Figure 62: pSMARt and nSMARt-labelled mESC can form chimaeras
E14 129Ola mESC were transfected and established *in vitro* with pSMARt. Expanded clones were injected into black (BL6) embryos **(A),** which resulted in the formation chimaeras, as observed by the agouti/chinchilla coat colour **(B).** The percentage of chimerism was dramatically improved as compared to using high-passage mESC, in some cases obtaining 100% chinchilla mice.

5.5.2 DNA-mediated approach – Microinjection of S/MAR vectors into 1-cell stage zygotes

All previous work performed in this thesis was accomplished by transfecting, selecting, and modifying cells *in vitro*. These modified cells carrying additional genetic information (*e.g.*: GFP) could, later on, be transplanted into living organisms (*e.g.*: to form chimaeras), which would also express the transgenic traits. However, the selection of positive cells via the addition of antibiotics was somehow artificial. So far, the cells have been kept under selective pressure for a defined period to force them to retain the exogenous DNA vector; in order words, either the cell kept the vector at all costs, or it died. The need of keeping exogenous DNA to survive under pressure conditions could be at the expense of integrating, breaking or rearranging such DNA. To get a better insight on the vector's behaviour in a 'selective-free'

in vivo environment, we sought to investigate the establishment and expression of pSMARt vectors during embryonic development, from a 1-cell zygote to a fully developed living mouse.

Also in collaboration with the Transgenic Service, between 130 and 780 molecules of pSMARt_CAG (pCAG) vector were injected into 1-cell stage zygotes of BL6 mice. 250 microinjected embryos were then transferred into 10 foster mothers and brought to term. 34 pups were born, and 3 died shortly after birth, which corresponded to approximately 13% survival rate. To check for toxicity of the transgene (GFP) and the to assess the microinjection performance, 12 embryos were kept aside in culture and monitored daily to check for correct development and GFP expression until hatching occurred **(Figure 63).** One of the embryos got arrested at the 1-cell stage and did not complete the first cellular division. Other three embryos failed to compact and also stopped dividing. The remaining eight embryos underwent compaction and developed into blastocysts. Embryo 9 hatched and attached to gelatinised and feeder-coated plates, and expressed GFP at that point. Unfortunately, the embryo died after a couple of days in adherent culture.

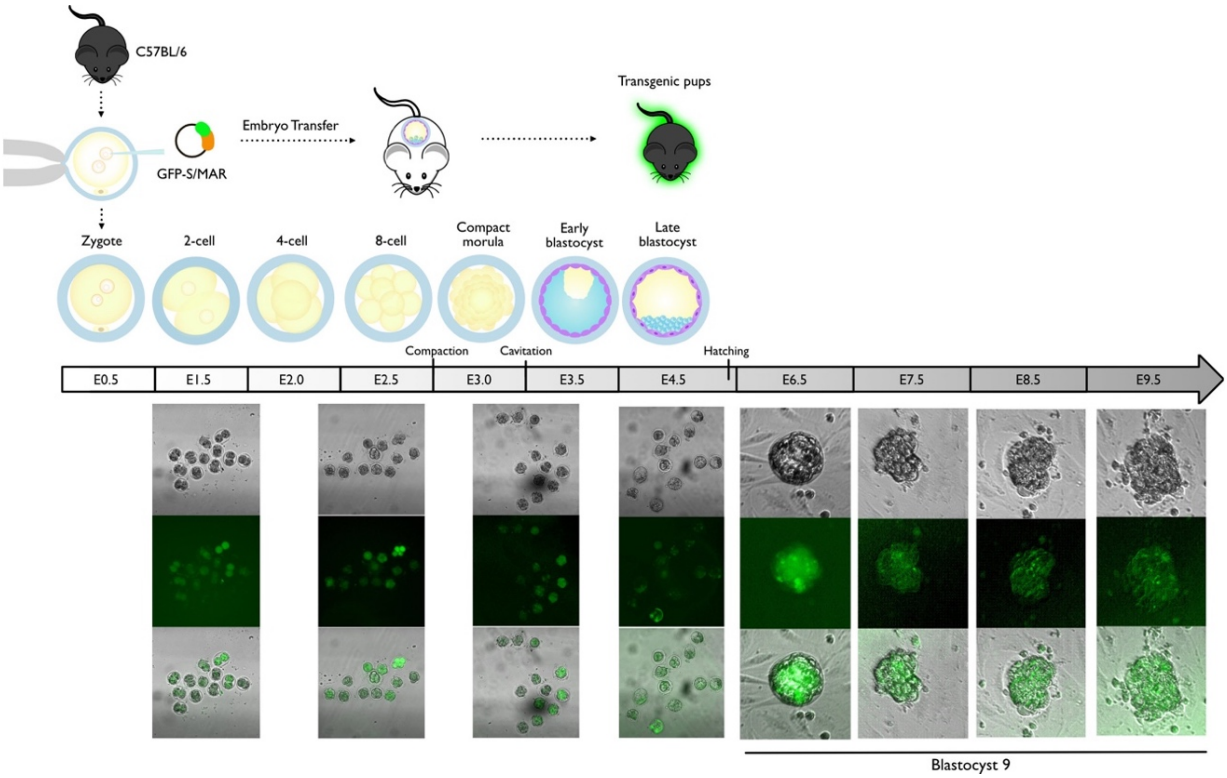


Figure 63: Expression of pSMARt vectors during the first stages of development
The establishment and expression of S/MAR vectors in a 'selection-free' environment, were analysed during the first stages of embryonic development. For this, pSMARt was microinjected into 250 C57BL/6 embryos at a 1-cell stage, which were then transferred into foster mothers. Twelve embryos were kept aside and monitored daily for transgene toxicity, GFP expression and correct embryonic development. Eight out of twelve embryos underwent compaction into blastocysts and one hatched from the zona pellucida and attached to a feeder-coated plate. The embryos expressed GFP throughout the first stages of embryonic development, suggesting that the vector does not damage the embryos and can be expressed from the first cell divisions.

5.5.3 S/MAR vectors are expressed transgenic tissues

5.5.3.1 GFP expression in chimeric mice

Although the S/MAR-labelled cells contributed to form chimaeras, as observed by the apparent coat-chimerism, the expression of GFP and therefore the presence and functionality of the vector remained unknown. To tackle this question, ear biopsies from 49 born chimaeras were analysed via fluorescent microscopy and by PCR. No difference was observed in the intensity of GFP between pSMARt and nSMARt-generated chimaeras **(Figure 64a).** The overall mean fluorescence intensity was significantly higher in the chimeric ears as compared to biopsies from BL6 negative controls **(Figure 64b-c),** although the fluorescence from individual ear biopsies was in some cases difficult to distinguish from background autofluorescence (represented by a red threshold line around 6 relative light units (RLU)) **(Figure 64d).** Since microscopic images were not always conclusive, PCR was used to amplify a 515 bp fragment of coGFP (*Primers GFP1-2*). The highly conserved non-coding region of the mammalian gene Sox21 (237bp) was used as an internal control (*Primers IC1-2*). GFP was detected in 26/49 biopsies **(Figure 64e),** although the positive samples did not always correspond to the GFP detected in the images and *vice versa*.

Since PCR and fluorescent results were not always coinciding, ten chimaeras were selected for further analysis (c6, c19, c32, c34, c35, c39, c44, c47, c48 and c49). The selection criteria applied was: GFP amplification via PCR, detection of fluorescence in the microscopic images and above 50% coat chimerism. Chimeras c6, c39, c44, c47 and c48 were bred with C57BL/6J mice. Chimeras c19, c32, c34 and c49 and later on c39 and c44, were sacrificed and the GFP expression was analysed in the chimeric organs as shown in **Figure 66.**

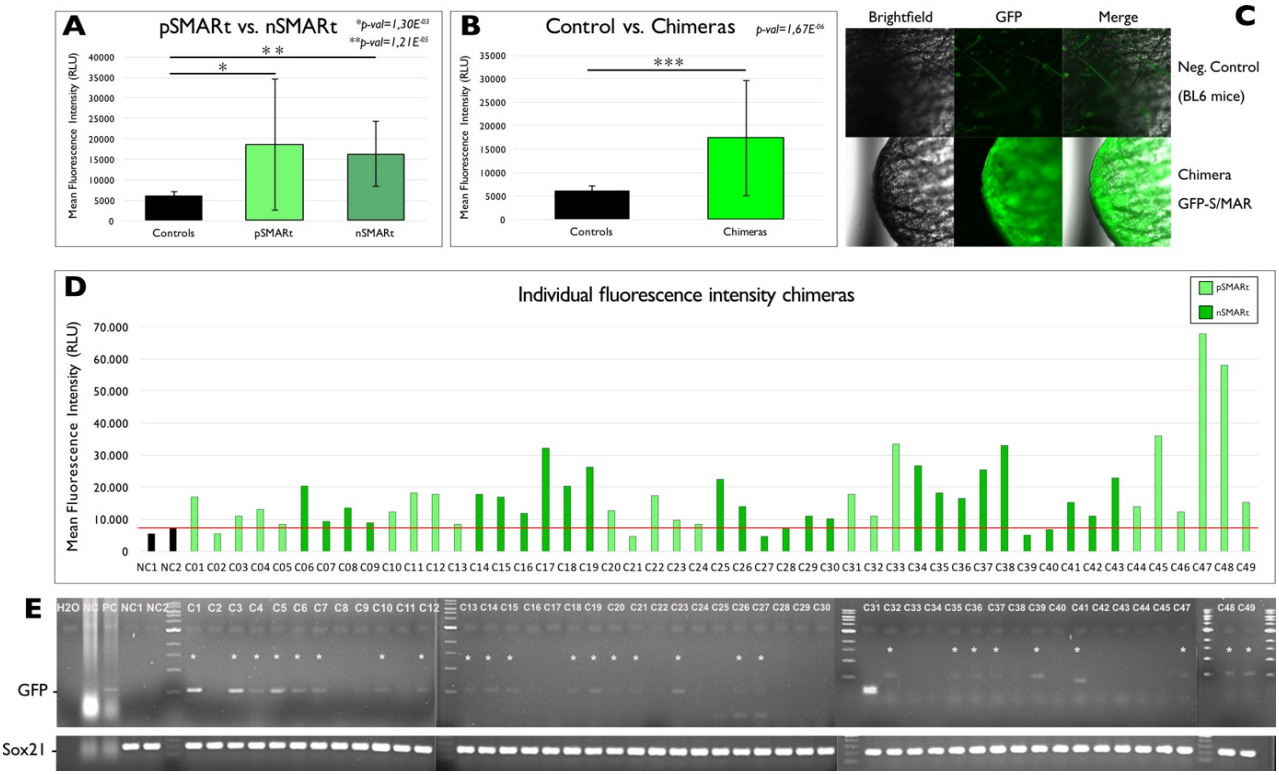


Figure 64: S/MAR vectors are expressed in chimeric ear biopsies

The GFP expression from pSMARt or nSMARt vectors was analysed in ear biopsies of 49 born chimaeras, via fluorescent microscopy and PCR. The overall fluorescence of pSMARt and nSMARt biopsies was significantly higher than biopsies of black mice with the same genetic background (BL6), although there was no difference in the overall intensity between pSMARt and nSMARt chimaeras **(A).** Similarly, when grouped, the overall fluorescent intensity of all chimeric biopsies was significantly higher than the autofluorescence from the negative controls **(B-C).** The fluorescence of individual ear biopsies was also analysed and compared to the background fluorescence, represented by a threshold line at 6 RLU **(D).** Also, a 515bp fragment of coGFP was amplified in 26/49 samples. A 237bp corresponding to the conserved mammalian Sox21 gene was used as internal control **(E).**
**p-val=1,30E-03. **p-val=1,21E-05. ***p-val=1,67E-06.*

5.5.3.2 GFP expression in pronuclear injected mice

Similarly, tail biopsies from 31 pronuclear injected pups were taken and analysed both microscopically as well as by PCR. The tails were compared to biopsies from C57BL/6 mice. Overall, transgenic tails showed a significantly higher fluorescence as compared to the background autofluorescence from the negative controls, which corresponded to 6 relative light units (RLU). When analysed individually, the specimens showed varying degrees of GFP. Also, GFP was analysed by PCR, which did not always correspond to the GFP detected by fluorescence microscopy **(Figure 65).**

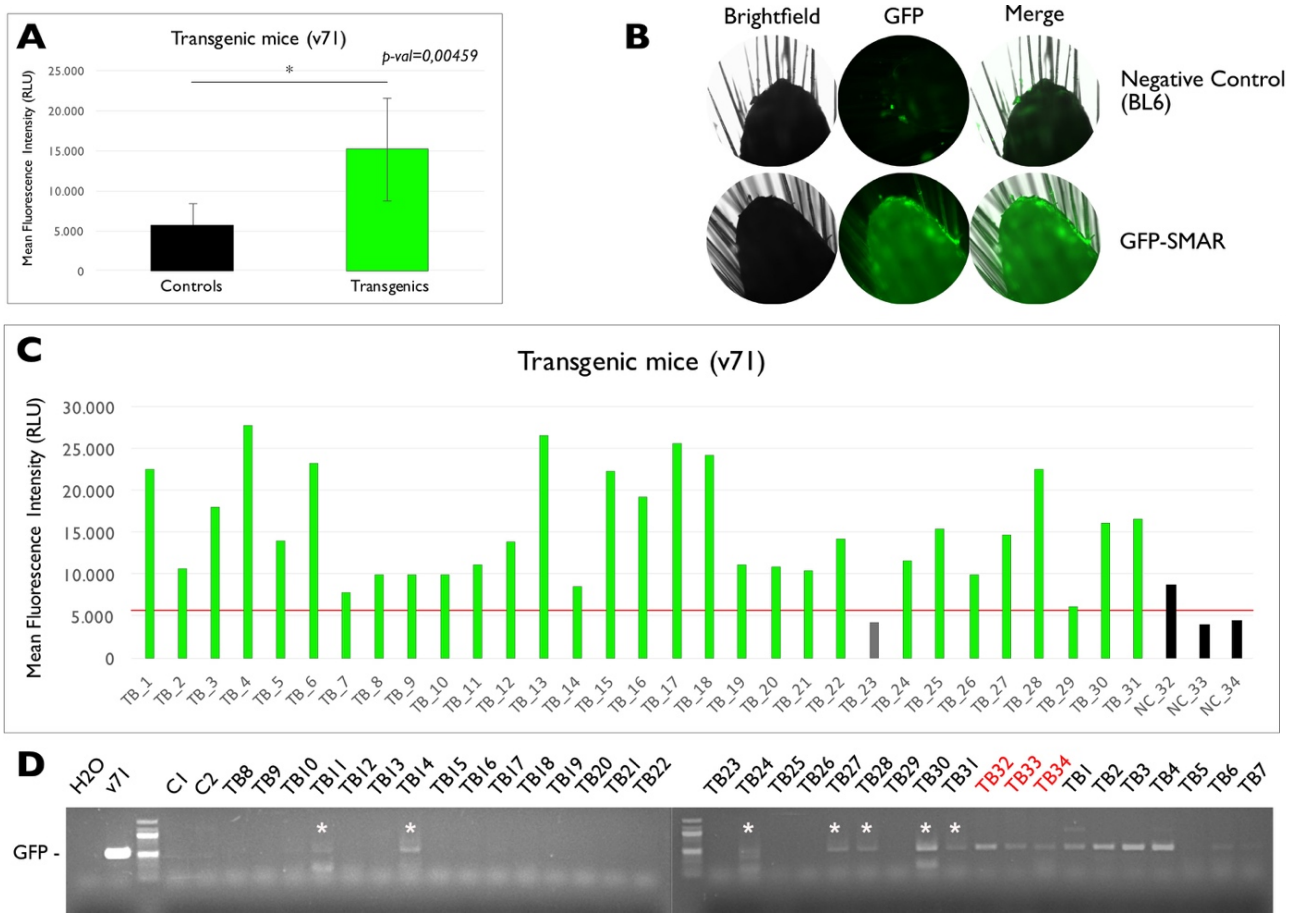


Figure 65: S/MAR vectors are expressed in biopsies of pronuclear injected mice.
The GFP expression from the pSMARt vector was analysed in tail biopsies from 31 born pups and negative controls, via fluorescent microscopy and PCR (Primers GFP 1-2). The overall fluorescence of pSMARt biopsies was significantly higher than biopsies of black mice with the same genetic background (BL6) **(A-B).** The fluorescence of individual tail biopsies was also analysed and compared to the background fluorescence, represented by a threshold line at 6 RLU **(C).** Also, a 515bp fragment of coGFP was amplified in some samples **(D)** as compared to the negative controls (C1 and C2). **p-val=0,00459.*

Since fluorescence and GFP amplicons were not always coinciding, 13 transgenic animals that showed GFP amplification as well as fluorescent expression were selected (T3, T4, T5, T6, T7, T11, T12, T14, T18, T21, T28, T30 and T31). Transgenics T11, T14, T18, T21, T28 and T30 were used for breeding and assessing the germline transmission of the vectors. Transgenics T4, T5, T6, T7, T8, T12 and later on T11, T18 and T21, were sacrificed and their tissues were analysed for the presence of S/MAR vector **Figure 66.**

The selected chimaeras and transgenics, together with C57BL/6J mice (negative control) and a constitutively active integrated UBC::GFP mice (positive control), were anaesthetised and terminally bled via heart puncture. The blood was collected in EDTA-tubes and analysed via Flow Cytometry. The animals were then dissected, and internal organs from different embryonic derivatives such as liver, kidney, heart, skeletal muscle, skin and testes (**Table 48**) were taken and cut into four pieces. One piece

was fixed in 4% PFA and used for immunohistochemistry, another piece was used for fluorescent imaging, and two pieces were snap frozen for further molecular analysis.

Together with the blood, spleens and femurs (bone marrow) were collected and placed in IMDM media for additional FACS analysis of hematopoietic organs **(Figure 67 - Figure 69).**

Table 48: Embryonic origin of organs and tissues

Organ	Embryonic origin
Liver	Endoderm
Skeletal Muscle	Mesoderm
Kidney	Mesoderm
Heart	Mesoderm
Skin	Ectoderm/ mesoderm
Spleen	Mesoderm
Bone Marrow/ blood	Mesoderm
Testes	Mesoderm

The tissues were imaged using a motorised fluorescent stereomicroscope (Leica M205 FA. Exposure 1 sec, Amp gain 1,9x, digital exposure 4). GFP was detected in all chimaeras analysed and was highly expressed in the muscle, skin and liver and to a lesser extent in the heart. No GFP was detected in the kidney. One transgenic (T11) showed high GFP expression in muscle and skin and less in the heart, but no expression in kidney or liver. The constitutively GFP-expressing mice used as positive control showed GFP expression across all organs whereas the C57BL/6 negative control showed no detectable fluorescence. Organs and tissues from a selection of analysed animals are shown in **Figure 66.**

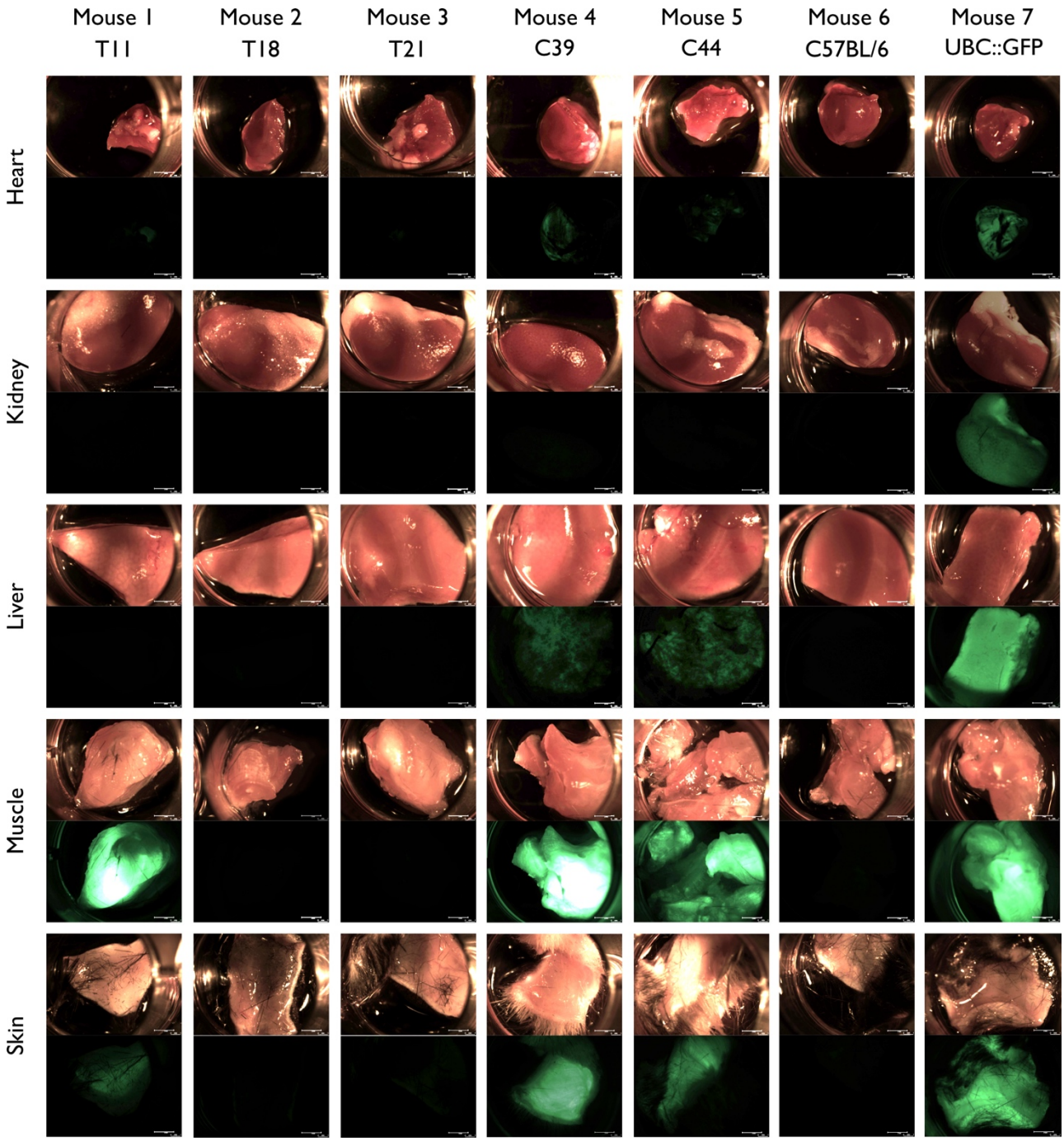


Figure 66: S/MAR vectors are expressed in transgenic organs

Organs derived from the three germ layers from chimeric and pronuclear injected mice were imaged using a motorised fluorescent stereomicroscope (Leica M205 FA. Exposure 1 sec, Amp gain 1,9x, digital exposure 4). GFP was detected in all chimaeras analysed and was highly expressed in muscle, skin and liver and to a lesser extent in kidney and liver. Only one of the analysed mice derived from pronuclear injection of pSMARt (T11) showed high GFP expression in the muscle and skin and to a lesser extent in the heart. As expected, the constitutively GFP-expressing mice used as positive control showed GFP expression across all organs whereas the C57BL/6 negative control showed no detectable fluorescence. Scale bars = 2mm.

Also, hematopoietic tissues such as the spleen, bone marrow and peripheral blood, were processed, stained for the common leukocyte surface antigen CD45, and their GFP expression was quantified via Flow Cytometry.

Blood samples were stained with a different panel of surface markers (CD45.1, CD11b, B220, CD4 and CD8a) to subgroup the GFP+ blood cells into T cells, B cells and myeloid cells **(Figure 67).** However, the EDTA-tubes were filled with too much blood, which resulted in insufficient EDTA to prevent coagulation and affected the overall quality of blood samples. No GFP was detected in the blood of Mouse 6 (Negative control) **(A),** whereas the majority of the blood from Mouse 7 (positive control) was GFP positive **(B).** None of the analysed mice generated by pronuclear injection showed GFP positive blood **(C-E).** The analysed chimaeras **(F-G)** showed the varying percentage of GFP blood cells (46,3 – 63,2%), with different blood cell type contribution. GFP population gated on viable cells and blood subpopulations gated on GFP positive cells.

A similar pattern was observed in the GFP expression from bone marrow. None of the analysed mice generated by pronuclear injection showed GFP -positive bone marrow cells (**Figure 68**C-E), while the analysed chimaeras **(F-G)** showed the varying percentage (31 - 46,7%) of GFP-positive cells.

A slight difference was detected in splenocytes **(Figure 69).** Almost 3% of GFP-positive splenocytes were detected in Mouse 1 (pronuclear injected mice T11) **(C)** but not in the others **(D-E).** The analysed chimaeras **(F-G)** showed the varying percentage of GFP-positive spleen cells (26,9 – 55,8%).

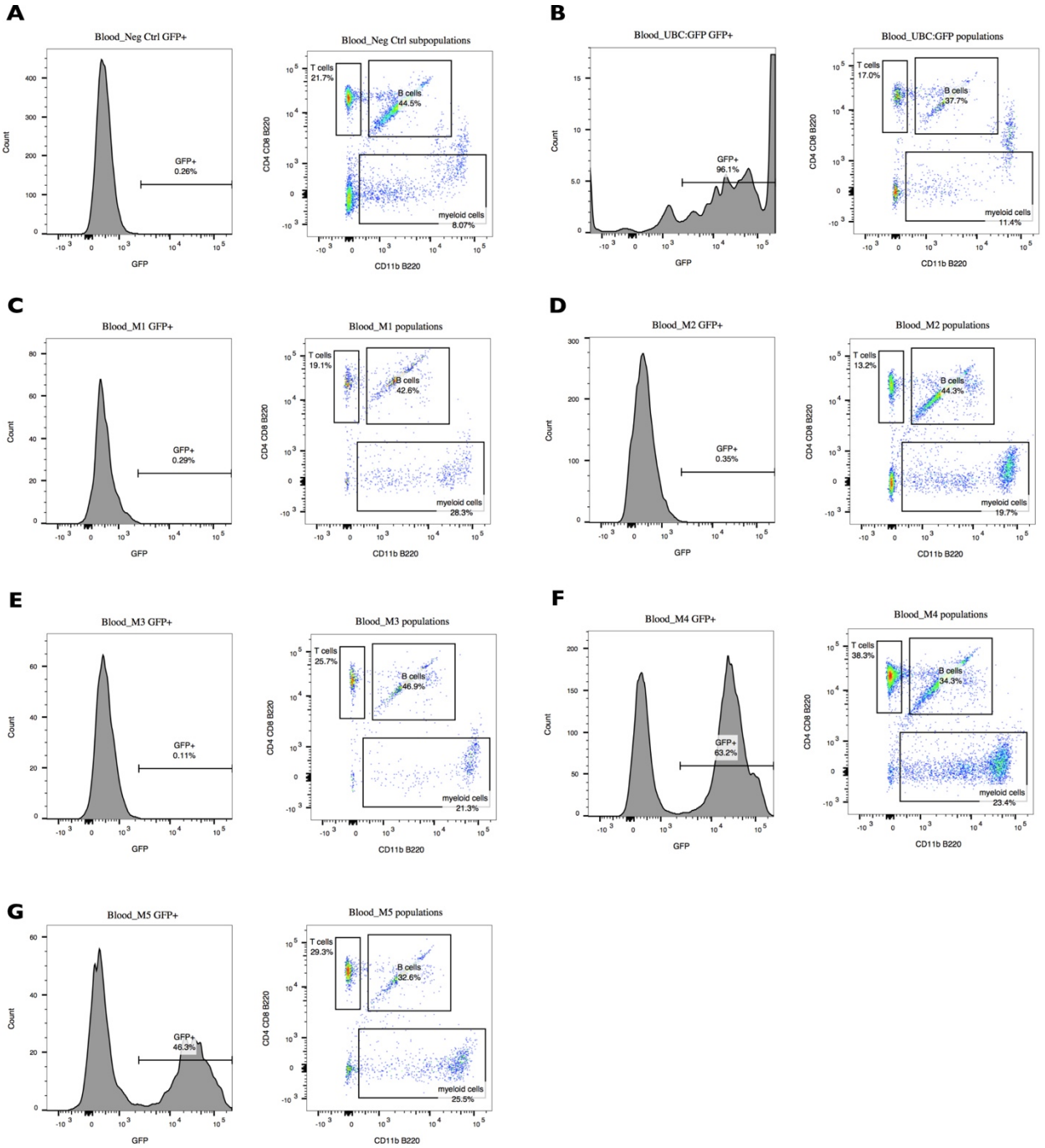


Figure 67: GFP analysis of Blood from chimeric and transgenic animals via Flow Cytometry.
Blood samples from chimaeras and pronuclear injected mice were obtained by terminal heart puncture and analysed for GFP expression. Also, a panel of surface markers (CD45.1, CD11b, B220, CD4 and CD8a) was used to subgroup GFP positive cells into T cells, B cells and myeloid cells. No GFP was detected in the blood of Mouse 6 (Negative control) **(A),** whereas the majority of the blood from Mouse 7 (positive control) was GFP positive **(B).** None of the analysed mice generated by pronuclear injection showed GFP positive blood **(C-E).** The analysed chimaeras **(F-G)** showed the varying percentage of GFP blood cells (46,3 – 63,2%), with different blood cell type contribution. GFP population gated on viable cells and blood subpopulations gated on GFP positive cells. *Note: the EDTA-tubes were filled with too much blood, which resulted in insufficient EDTA to prevent coagulation and affected the overall quality of blood samples.*

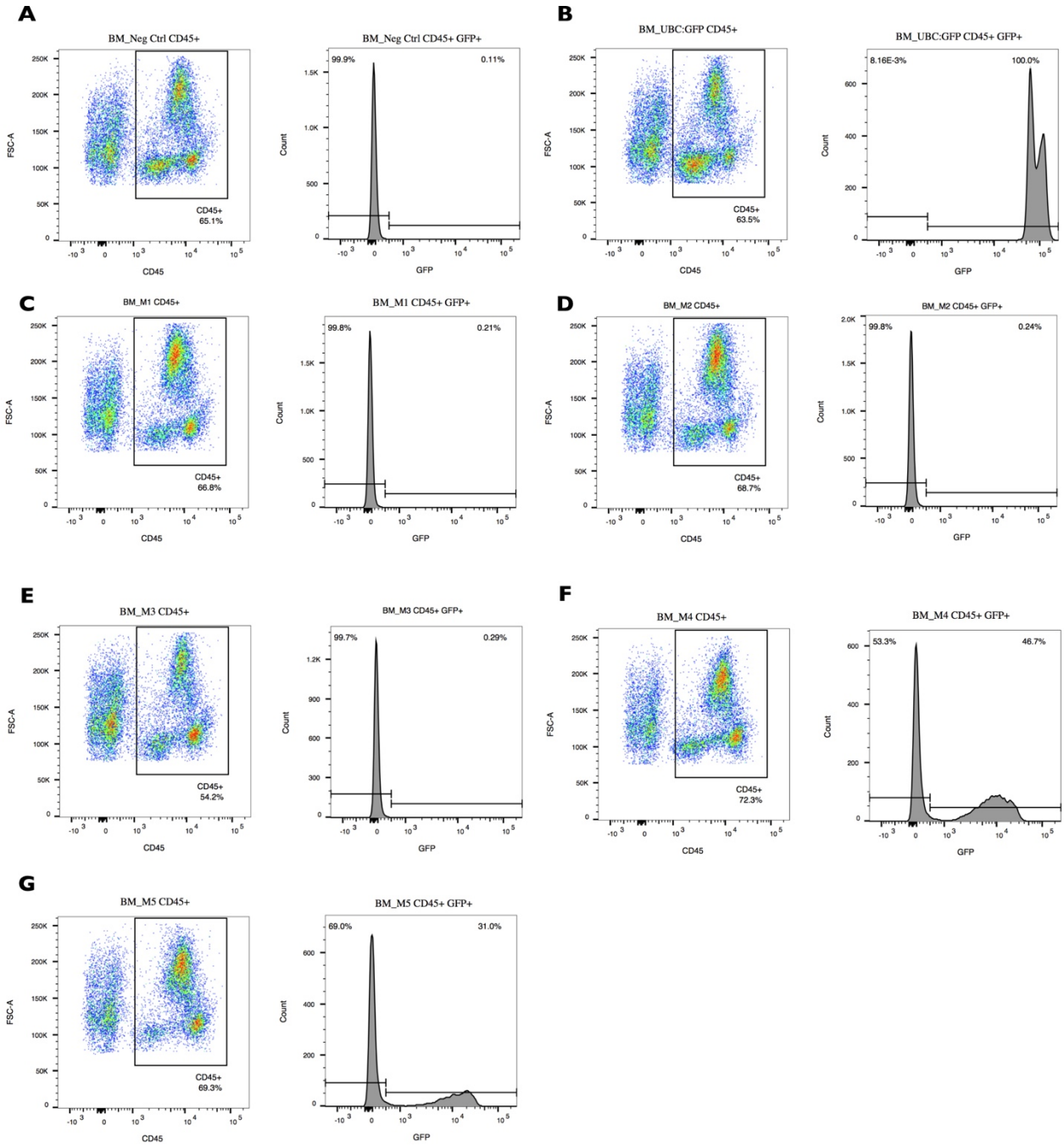


Figure 68: GFP analysis of Bone Marrow from chimeric and transgenic animals via Flow Cytometry.
The bone marrow (BM) was flushed from the femur and stained against the common leukocyte surface antigen CD45, common in all hematopoietic cells. No GFP was detected in BM from Mouse 6 (Negative control) **(A),** whereas the majority from Mouse 7 (positive control) was GFP positive **(B).** None of the analysed mice generated by pronuclear injection showed GFP positive BM **(C-E).** The analysed chimaeras **(F-G)** showed the varying percentage of GFP cells (31 - 46,7%). CD45 population gated on viable cells and GFP population gated on CD45 positive cells.

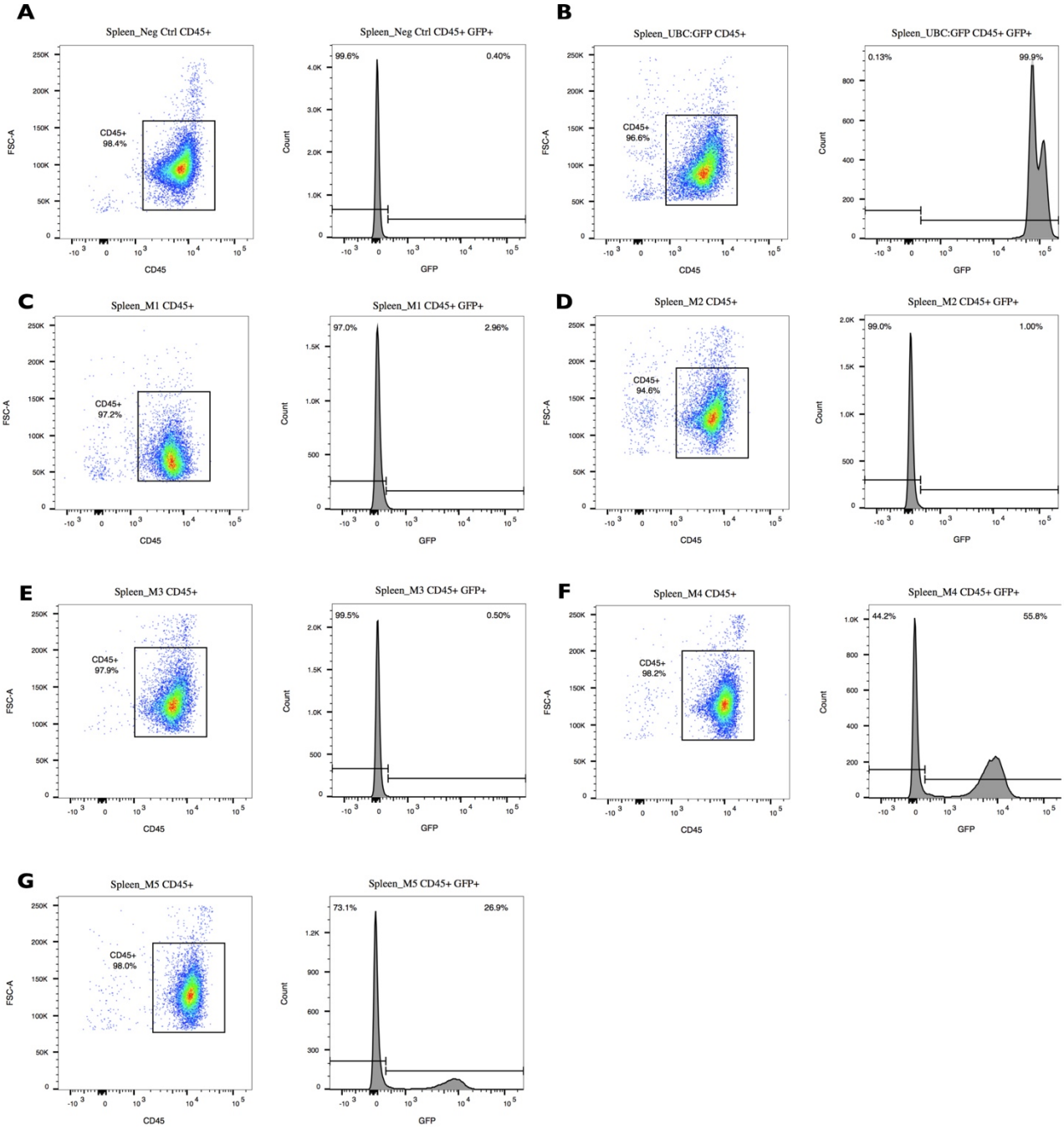


Figure 69: GFP analysis of Spleen from chimeric and transgenic animals via Flow Cytometry.
The spleens were homogenised and the recovered splenocytes were stained with CD45. No GFP was detected in the spleen from Mouse 6 (Negative control) **(A),** whereas the majority from Mouse 7 (positive control) was GFP positive **(B).** Almost 3% of GFP-positive splenocytes were detected in Mouse 1 **(C)** but not in the other pronuclear injected mice **(D-E).** The analysed chimaeras **(F-G)** showed the varying percentage of GFP positive spleen cells (26,9 – 55,8%). CD45 population gated on viable cells and GFP population gated on CD45 positive cells.

5.5.4 Germ-line transmission of S/MAR vectors

5.5.4.1 Stem Cell transmission of S/MAR-labelled mESC in chimaeras

After confirming that S/MAR-mESC were able to form chimeras and to retain the vector expression throughout cell division and differentiation both *in vitro* and *in vivo*, we wondered whether the pSMARt or nSMARt labelled mESC would contribute to germ-line transmission (stem cell transmission) and most importantly, whether the vector would be able to survive meiosis and be passed onto the offspring (vector transmission).

To investigate this, six chimaeras (♀c6 and ♂c35, c39, c44, c47 and c48), which had above 80% of coat chimerism and showed GFP amplification in a screening PCR, were selected for breeding with C57BL/6J mice. In the positive scenario where S/MAR-labelled mESC (129 Ola) would contribute to the germ-line (ectoderm), either a mixed black and agouti or a 100% agouti coat-colour offspring would be expected. In the negative scenario, only black offspring would be produced.

Surprisingly, two of the best chimaeras (c39 and c44, both males) showed stem cell germ-line transmission as the totality of their offspring was agouti **(Figure 70).** Both pSMARt (c44) and nSMARt (c39) labelled mESC were able to contribute to the germ-line.

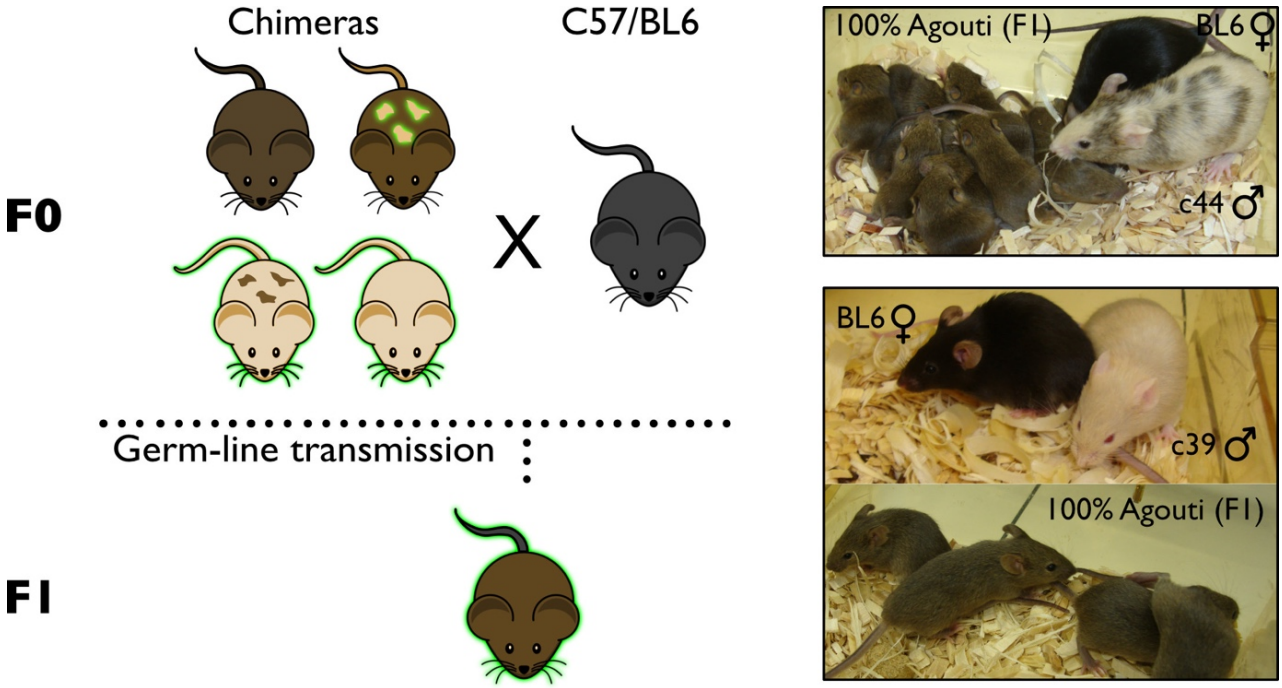


Figure 70: S/MAR-modified stem cells contribute to the germ-line
To investigate whether S/MAR-labelled mESC were able to contribute to the germ-line, the best six generated chimaeras (generation F0) were backcrossed to C57BL/6J mice. In a positive scenario of germ-line transmission, all offspring (F1 generation) would have an agouti coat colour. In a negative scenario, all the offspring would be black. Two out of six bred chimaeras c39 (nSMARt) and c44 (pSMARt), both males, resulted in 100% agouti offspring.

5.5.4.2 Vector transmission in chimaeras

After confirming that the transplanted 129Ola mESC were able to contribute to the germ-line and could be passed onto the offspring, we wondered if that would be the case for the S/MAR vectors and if vector germ-line transmission was possible. However, little is known about episomal vector behaviour during meiosis and gametogenesis. To tackle this, a GFP screening was performed in blood (protein level) and tail biopsies (DNA level) from agouti pups (F1 generation) from c39 (nSMARt) and c44 (pSMARt), which had a 50% 129Ola mESC contribution. Therefore, assuming vector transmission, GFP would be expected in all F1 pups. First, a GFP expression screening was performed in blood, which was obtained from a puncture in the facial vein, collected in EDTA tubes and analysed via Flow Cytometry to detect expression of GFP (**Figure 71**).

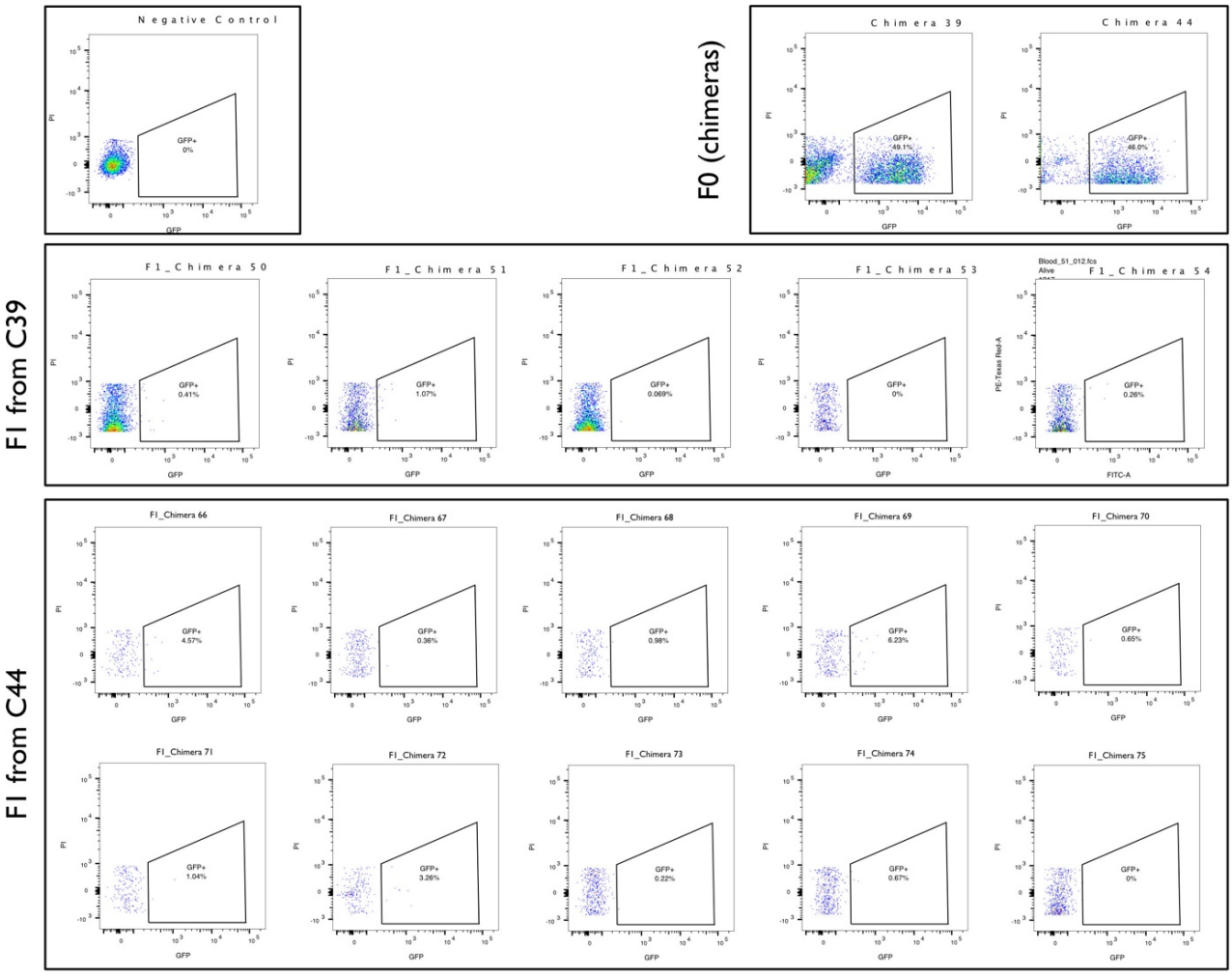


Figure 71: pSMARt and nSMARt are not expressed in the chimeric F1 generation
Six generated chimaeras (F0) were bred with C57BL/6 mice. Two male chimaeras: c39 (nSMARt) and c44 (pSMARt), showed stem cell transmission and produced agouti offspring. To investigate whether the F1 generation expressed either nSMARt or pSMARt, blood from the facial vein was collected from all pups and analysed using flow cytometry. As expected, black pups without 129Ola mESC contribution did not show expression of GFP (data not shown). Surprisingly, GFP was only detected between 1-6% of blood cells in agouti pups, which have a 50% contribution of exogenous mESC.

As expected, the blood from the F0 generation (c39 and c44) was almost 50% chimeric, with 46 – 49,1% of GFP-positive cells, as compared to the negative control (C57BL/6 mice). However, only one pup from c39 showed four pups from c44 showed above 1% of GFP positive cells, ranging from 1,04 to 6,23%.

We then wondered if the lack of GFP expression was a consequence the DNA being lost during meiosis or being present but silenced. For that, a screening PCR was performed on two different litters from c39 and c44, which stem cell transmission (agouti coat). Although GFP could be amplified in the F0 generation, no amplification was found in the F1 offspring, regardless of the litter; suggesting that the vector was not present either episomally or integrated (**Figure 72**).

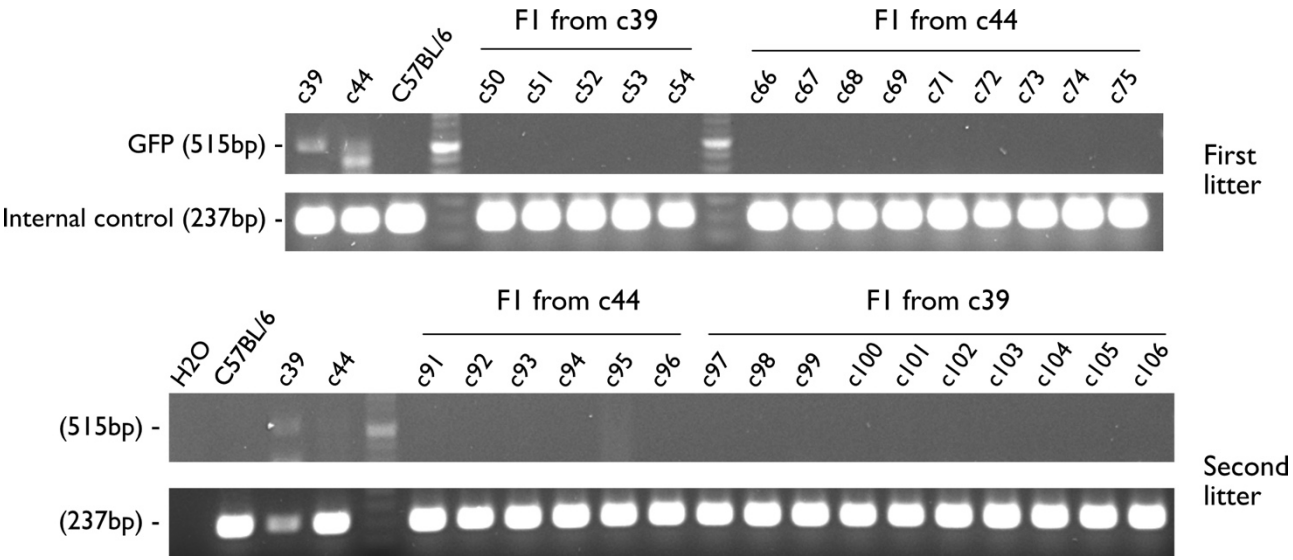


Figure 72: pSMARt and nSMARt are not present in F1 chimeric tissues
Six generated chimaeras (F0) were bred with C57BL/6 mice. Two male chimaeras: c39 (nSMARt) and c44 (pSMARt), showed stem cell transmission and produced agouti offspring. However, the F1 generation did not show GFP expression when peripheral blood was examined. To investigate whether the offspring lacked GFP expression due to loss or silencing of the DNA, a genotyping PCR was performed. Although GFP could be amplified in the two F0 males, no GFP amplification was detected neither in the C57BL/6 (negative control) nor two litters from the F1 generation, suggesting that the transgene was not transmitted to the offspring, neither episomally or integrated. The internal control corresponds to a conserved fragment of the mammalian Sox21 gene.

This data suggested that the vector was either lost or silenced during gametogenesis and that the outcomes are independent of which vector type was used.

5.5.4.3 Vector transmission in pronuclear injected mice

Differently, from the chimeric animals, pronuclear injected mice had no phenotypic trait that could suggest the presence or absence of S/MAR vectors. Therefore, a GFP screening was performed in all 31 generated F0 mice, either via PCR amplification of GFP (**Figure 65**) or flow cytometry.

Surprisingly, three F0 animals (T11, T18 and T21) expressed GFP, ranging from 4 to 10% of positive cells in peripheral blood **(Figure 73),** as compared to C57BL/6 mice (negative control).

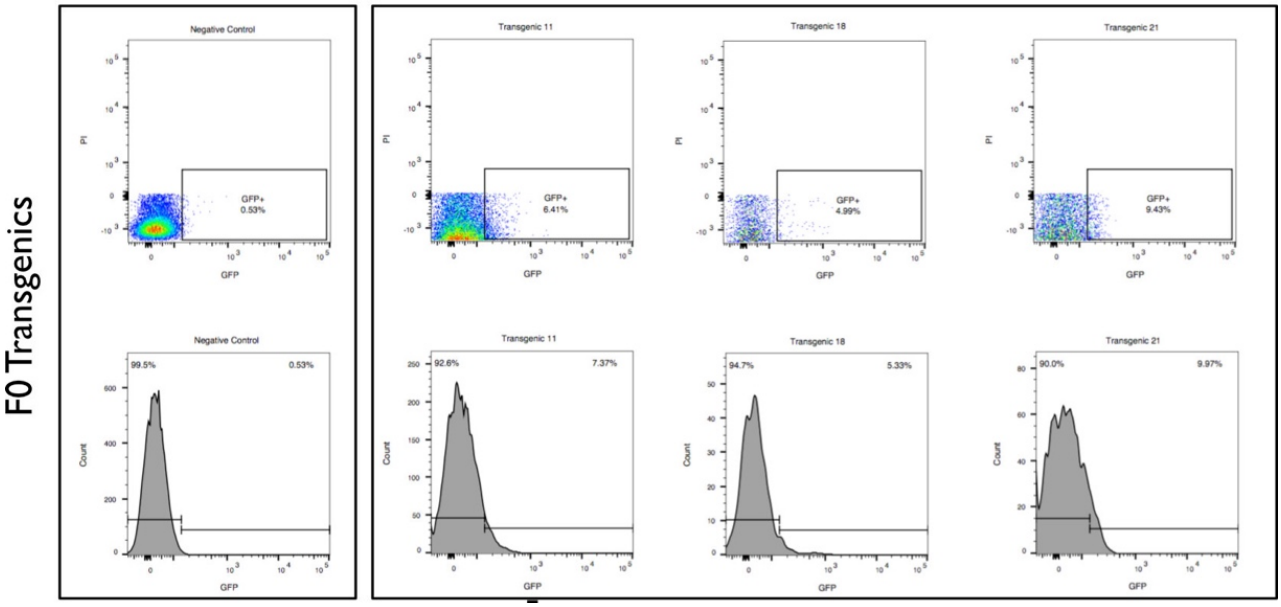


Figure 73: pSMARt is expressed in F0 pronuclear injected mice
All 31 born mice (F0) generated by pronuclear injection of pSMARt, were analysed for GFP expression in peripheral blood from a facial vein puncture. From these, three males (T11, T18 and T21) showed varying degrees of GFP expression ranging from 4-10% as compared to the negative C57BL/6 mouse.

Some animals (T11, T14, T18, T21 T28, T30 and T31) were selected for breeding with C57BL/6J mice to check for germ-line transmission of the pSMARt vector, as well as to analyse its presence or absence in the germ-line. When the litters were born, blood from the facial vein was used to determine the expression of the vector as well as PCR to confirm its presence.

Mouse T11, which showed GFP expressing cells in blood, generated two litters that were subjected to a genotyping PCR to check for the presence of pSMARt **(Figure 74).** No GFP could be detected in biopsies from the first litter. The second litter contained two pups which showed amplification of GFP.

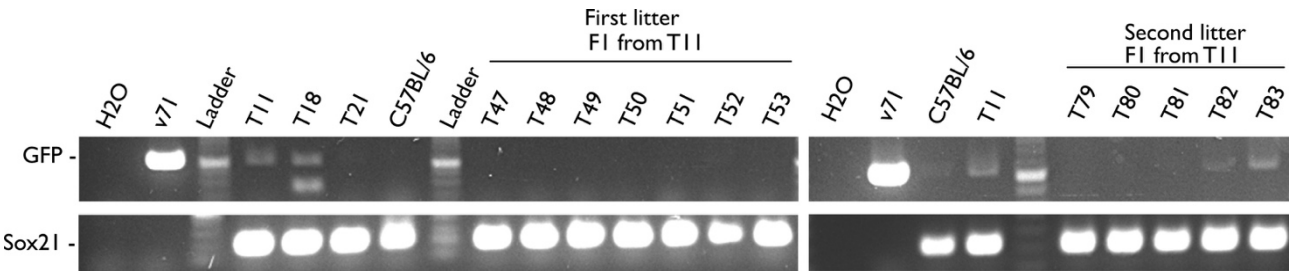


Figure 74: pSMARt can be detected in the F1 generation of pronuclear injected mice
Mouse T11 was bred with a C57BL/6 female and produced two litters. No GFP could be amplified in any of the pups from the first litter, although GFP expression could be detected via Flow cytometry. Surprisingly, two pups from the second litter showed amplification of GFP via PCR.

5.5.4.4 Study of S/MAR behaviour during meiosis

The previous analysis of GFP in the F1 generation pointed at meiosis as a significant obstacle for episomal germ-line transmission. To investigate the behaviour of S/MAR vectors during this particular cell division process, the presence and expression the S/MAR vector were evaluated before and after gametogenesis.

For this, the testes and sperm from chimaeras c39 and c44 as well as from the pronuclear injected mice T11, T18 and T21 were collected and analysed for both expression (GFP fluorescence) and presence (GFP amplification) of the S/MAR vectors. Black C57BL/6J mice as well as a constitutively expressing mouse with integrated GFP driven by the UBC promoter (UBC::GFP mice, Jackson Lab) were used as negative and positive controls respectively.

The testes were imaged with a Leica M205 FA microscope (Exposure 1 sec, Amp gain 1,9x, digital exposure 6. Scale bars = 2mm) as shown in **(Figure 75a).** GFP expression could be detected in both chimaeras (c39 and c44) as well as in one of the pronuclear injected mice (T11). To exclude that the fluorescence detected was emitted by the external testicular membrane or *Tunica Albuginea*, the *seminiferous tubules* were homogenised to isolate the germinal cells, comprised by *spermatogonias*, *spermatocytes* and *spermatids*. The GFP fluorescence from these cells was analysed by flow cytometry **(Figure 75b)**, which revealed the same expression pattern than fluorescent images from testes. Finally, the sperm was collected from the *epididymis* and imaged using a Nikon Ti microscope (scale bars = 100um). GFP could only be detected in sperm from the positive control UBC::GFP **(Figure 75c).**

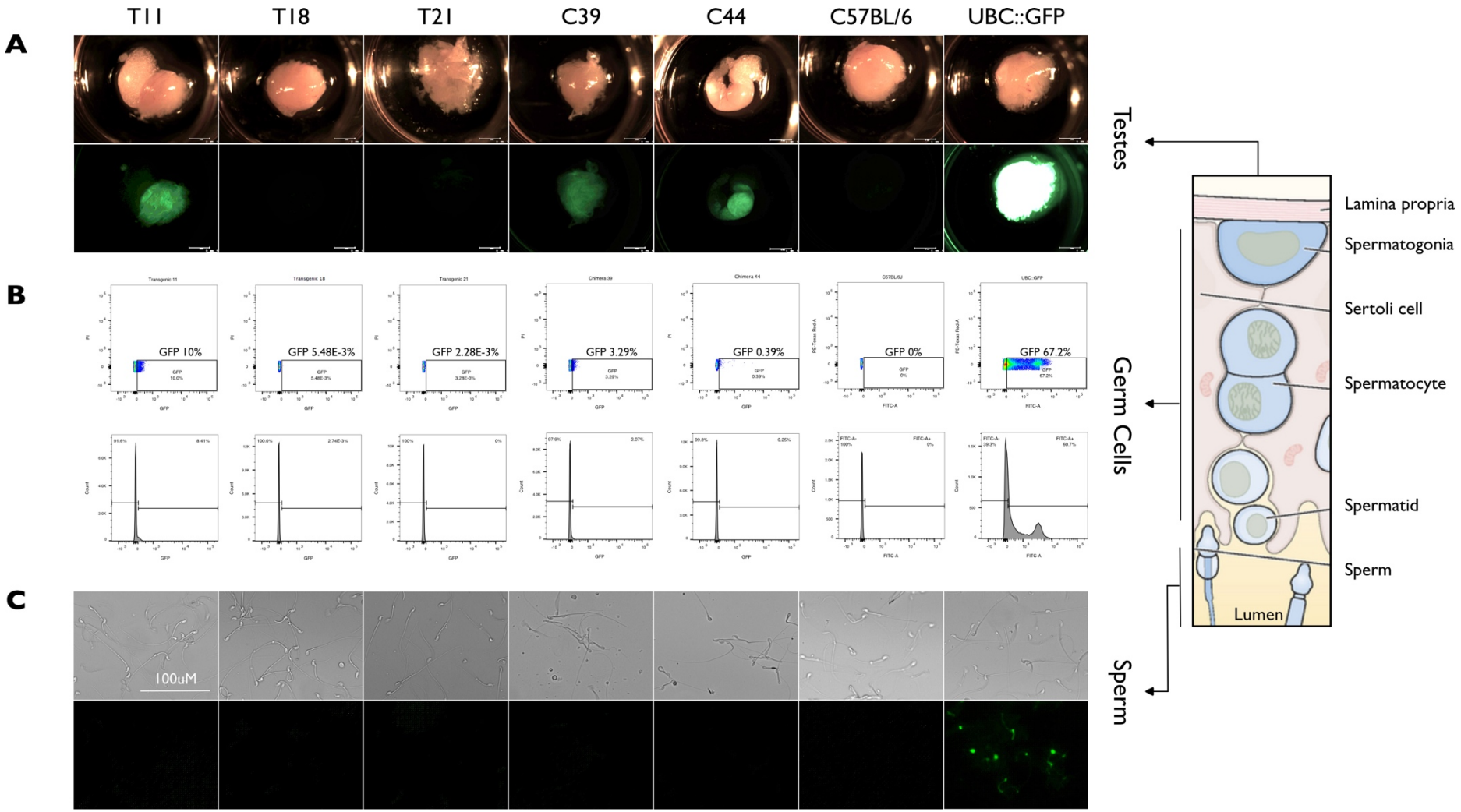


Figure 75: GFP expression along the male reproductive tract

The reproductive organs from chimeras (c39, c44), pronuclear injected mice (T11, T18, T21), a negative control (C57BL/6J) and a positive control (UBC::GFP) were imaged using a a motorized fluorescent stereomicroscope (Leica M205 FA. Exposure 1 sec, Amp gain 1,9x, digital exposure 6. Scale bars = 2mm.). GFP was detected in both chimeras as well as in one pronuclear injected mice **(A).** In addition, the seminiferous tubules were digested to isolate the germinal cells. The GFP expression was quantified via flow cytometry and followed the same expression pattern than the testes **(B)**. The sperm was collected and imaged using a Nikon Ti microscope (Scale bars = 100uM) **(C).** GFP could only be detected in the sperm from mice constitutively expressing integrated GFP (positive control). Schematic adapted from *Falcone, T., & Hurd, W. (2007). Clinical Reproductive Medicine and Surgery. Elsevier Health Sciences.*

We then addressed whether the lack of GFP fluorescence in the sperm was due to silencing or loss of the episome during spermatogenesis. For this, DNA from sperm and testes was amplified via PCR **(Figure 76).** GFP could be detected in both sperm and testes from mice c39, c44 and T11, as well as in the technical and biological positive controls, suggesting that S/MAR vectors are present at both initial and end stage of spermatogenesis but are only expressed before meiosis occurs.

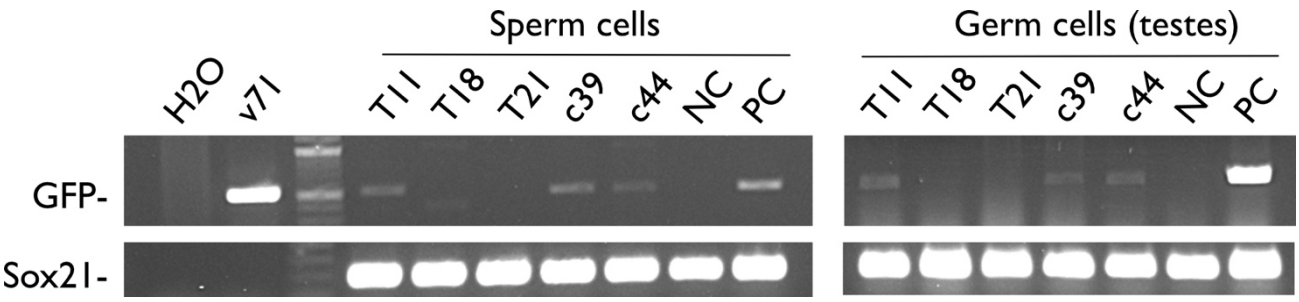


Figure 76: S/MAR vectors can be detected in both testes and sperm
To investigate the presence or absence of GFP (and therefore, episomal DNA) in the reproductive tissues, both sperm and testes from chimeric and pronuclear injected mice were subjected to PCR amplification of the transgene. GFP could be detected in sperm and testes from the chimaeras c39 and c44 as well as from the pronuclear injected mice T11. The positive control mice (UBC::GFP) as well as the pSMARt_CAG plasmid (v71) showed successful GFP amplification whereas the biological (C57BL/6J mice) and technical (water) negative controls did not.

We then investigated at which stage of spermatogenesis did the loss of expression occur **(Figure 77).** For this, the germinal epithelia from the *seminiferous tubules* were subjected to immunohistochemistry and stained against GFP, with haematoxylin counterstaining (upper row). Sections stained only with secondary antibody were used as technical controls (middle row), and Haematoxylin-Eosin (HE) staining revealed the tissue architecture (lower row).

Pronuclear injected mice T11 as well as chimaeras c39 and c44 showed expression of GFP in the most peripheral layer of cells, which corresponds to *spermatogonial cells* (see the cartoon in **Figure 75** for histological reference). However, no GFP could be detected in more advanced meiotic cells, such as *spermatocytes*, *spermatids* or *sperm cells*. The negative control C57BL/6 mice showed no GFP expression at all, whereas the constitutively active UBC::GFP mice showed GFP expression along the germinal epithelia.

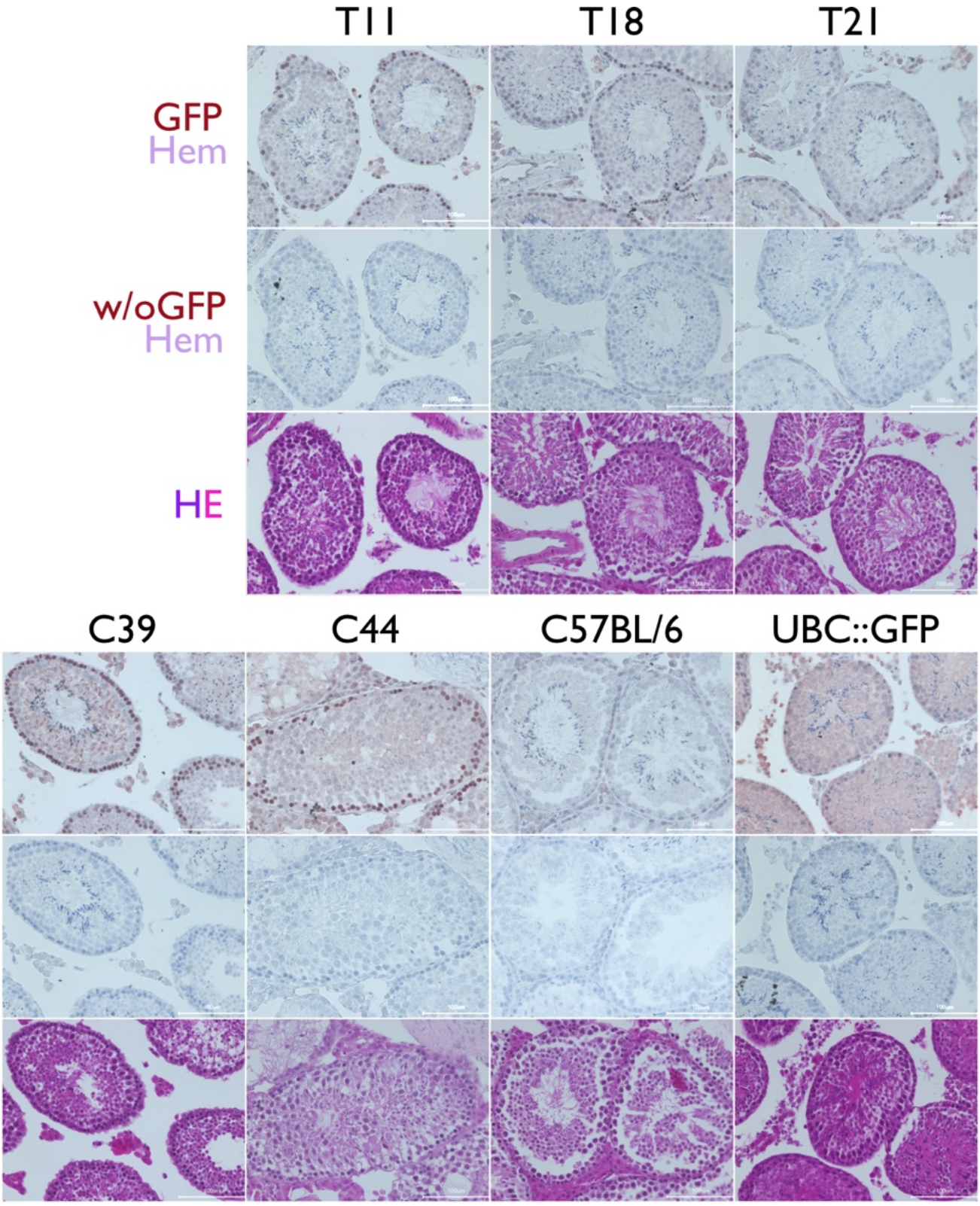


Figure 77: S/MAR expression is lost during meiosis

Immunohistochemical staining for GFP with haematoxylin counterstaining (upper row), unstained control (middle row) and HE (bottom row) of testicular sections from chimeric and pronuclear injected mice. T11, as well as C39 and C44, showed GFP expression in the outer cell layer, which corresponds to spermatogonial cells, but no GFP expression could be detected in spermatocytes or sperm. *Scale bars=100μm*

5.5.5 Summary

The generation of *bona fide* chimaeras can be achieved by using 'freshly' generated and low passage mESC, in some cases reaching almost complete coat chimerism. Both pSMARt (pCAG) and nSMARt (nCAG)-labelled mESC contributed to form chimaeras, which expressed significantly higher levels of fluorescence as compared to background autofluorescence from C57BL/6 negative control mice. However, no significant difference in overall GFP expression was observed between pSMARt and nSMARt chimaeras. The presence of S/MAR vectors was also detected by PCR amplification of GFP, which was observed in some of the chimaeras. Similarly, transgenic mice generated by pronuclear injection of pSMARt into the zygote also showed a significant degree of GFP expression as compared to negative controls. Transgenics generated both '*ex vivo*' and '*in vitro*' with either S/MAR-labelled mESC or pronuclear injected zygotes, contributed to generate organs derived from the three germ-layers, confirming that the vectors could survive differentiation to fully committed and differentiated cell types of a living organism. Finally, the capabilities of S/MAR vectors were further challenged by investigating their germ-line transmission potential. At least two male chimaeras produced entirely agouti F1 litters, which proved that S/MAR-labelled mESC could contribute to the germ-line (stem cell transmission). The unknown behaviour of episomal DNA vectors during meiosis was challenged by checking the GFP expression in those agouti F1 litters from chimaeras and pronuclear injected mice (vector transmission). Although the S/MAR vectors could be detected before (testes) and after meiosis (sperm), their expression was lost at the spermatid stage.

5.6 Genetic modification and reprogramming of human fibroblasts to hiPSC

In this chapter, the reprogramming abilities of the S/MAR vector system are compared to the state-of-art for episomal reprogramming, the EBNA-1/OriP system, using the vectors described in **Section 5.1.11.** For this, the reprogramming efficiency as well as the pluripotent quality of generated EBNA and S/MAR-hiPSC clones were compared. In a separate section, the reprogramming abilities of the 4in1 Lentiviral-derived vectors POP and nPOP, described in **Section 5.1.10,** were also tested in a more suitable and easier to transfect cell line of human dermal fibroblasts.

5.6.1 Comparison between S/MAR vectors and the EBNA-1/OriP system

As explained earlier in this thesis, the transfection in murine fibroblasts (MEFs) was not efficient enough to deliver high levels of reprogramming factors to allow efficient reprogramming. To overcome the DNA delivery barrier, we tested different S/MAR reprogramming vectors in human dermal fibroblasts (HDFs), which can be routinely and efficiently transfected.

5.6.1.1 S/MAR vectors can reprogram hiPSC

In collaboration with Prof. Mckay (Manchester Metropolitan University, UK), we compared the reprogramming potential and efficiency of S/MAR vectors to the well established EBNA-1/OriP vector system. Prof. Mckay kindly provided the expertise and the EBNA-based vectors [203] from which S/MAR versions containing the human Reprogramming factors *Oct4*, *Sox2*, *Klf4*, *L-Myc* and *Lin28* (OKSML), as well as shRNA against p53 were subcloned **(Figure 37).**

Both EBNA-1 or S/MAR vectors were nucleofected into CLN7 474 human dermal fibroblasts (HDFs) from patients suffering from Batten disease, which were commonly used in Prof. Mckay's lab. Transfected cells were kept under reprogramming conditions for three weeks. At day 21, HDFs that received the reprogramming factors delivered in EBNA vectors showed signs of Mesenchymal-Epithelial Transition (MET). The cells that were transfected using S/MAR vectors had a slight delay in reprogramming but showed the same morphological changes at day 23 **(Figure 78a).** Thereafter, the emerging hiPSC colonies were picked and expanded, and their morphology was monitored over time. After a couple of passages, the pluripotency of hiPSC was assessed using Immunofluorescence stainings for human pluripotency markers, such as Tra160, SSEA3, SSEA4 and Nanog **(Figure 78b).** Both EBNA-1 and S/MAR reprogrammed colonies showed expression of all pluripotency markers as

well as the positive control (hESC). A slight expression of the embryonic surface marker SSEA-3 was detected in feeder layers, as they also are of embryonic origin. All work performed in human embryonic stem cells (hESC) was performed at Prof. Mckay's lab in Manchester, UK.

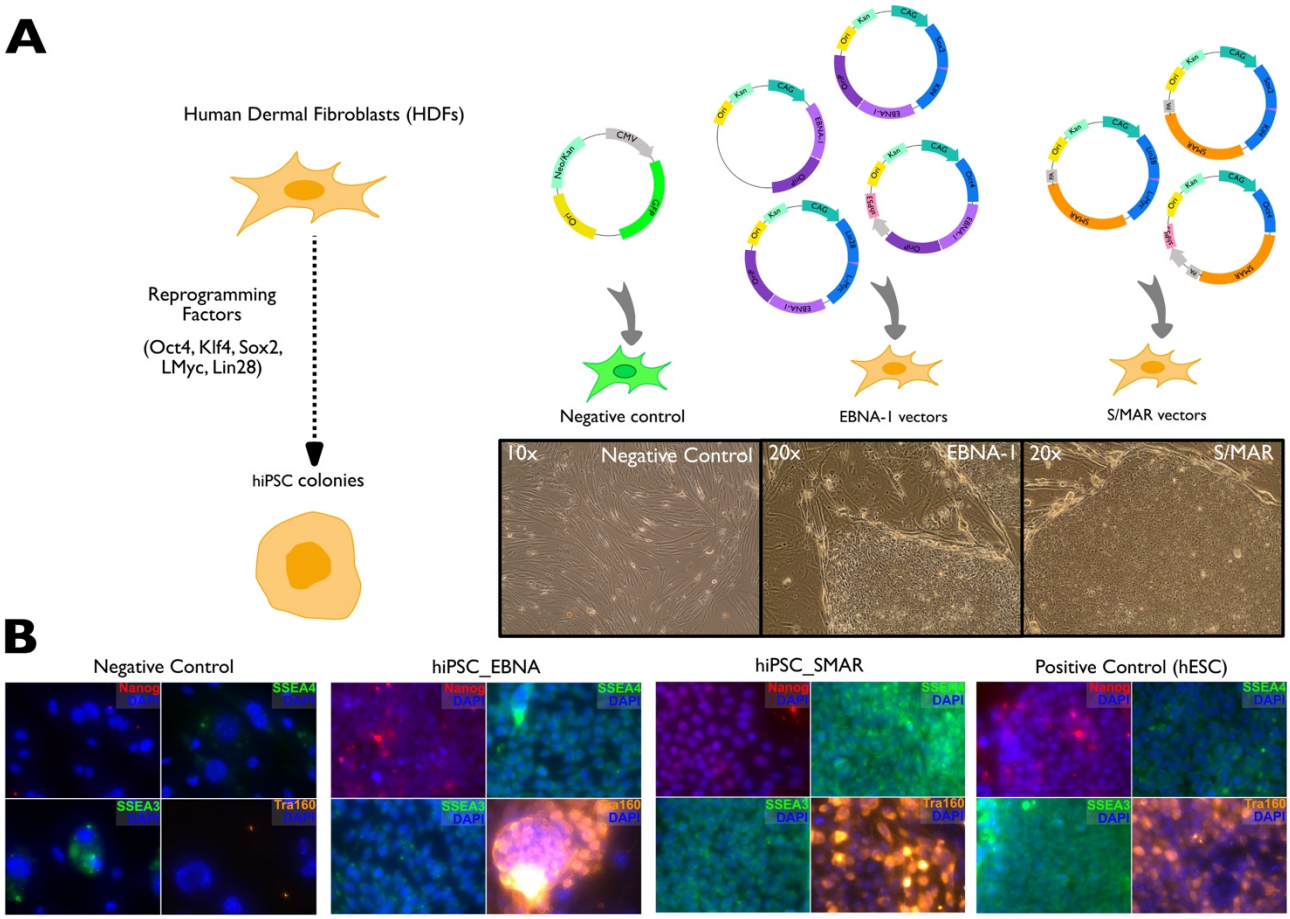


Figure 78: EBNA-1 and S/MAR episomal reprogramming and pluripotency assessment in hiPSC.
CLN7 474 human dermal fibroblasts (HDFs) were electroporated with EBNA-1 or S/MAR episomal vectors and kept under reprogramming conditions for over a month. At day 21, signs of MET were obvious in EBNA-1 transfected cells, whereas S/MAR containing fibroblasts showed a slight delay in reprogramming. However, after a month of reprogramming, there were no apparent microscopical differences between EBNA-1 and S/MAR reprogrammed fibroblasts **(A).** The expression of pluripotency markers (Tra160, SSEA-3, SSEA-4 and Nanog) was assessed in the generated hiPSC **(B).** Both EBNA-1 and S/MAR reprogrammed colonies showed expression of all pluripotency markers as well as the positive control (hESC). A slight expression of the embryonic surface marker SSEA-3 was detected in feeder layers, as they are from embryonic origin. All work performed in human embryonic stem cells (hESC) was performed at Prof. Mckay's lab in Manchester, UK.

5.6.1.2 Comparison between EBNA and S/MAR reprogramming efficiencies

The reprogramming efficiencies between EBNA-1 and S/MAR vectors were then compared. Due to limited access to Batten disease fibroblasts, early passage wildtype NHDFs (Promocell) were used for further experiments. The cells were nucleofected with the respective EBNA or S/MAR vectors (together with pSMARt as a GFP-labelling vector) and kept under reprogramming conditions for a month. Untransfected NHDFs were used as negative control. As observed with patient fibroblasts, EBNA colonies started showing mesenchymal-epithelial transition (MET) and a morphology resembling

hiPSC at around day 20, whereas S/MAR colonies were delayed by approximately 10 days. At day 30, the cells were fixed and stained with the Alkaline Phosphatase (AP) Staining Kit II (Stemgent).

Out of 60.000 cells that were allowed to reprogram, EBNA-nucleofected fibroblasts generated 55 AP-positive colonies (0.092%), whereas fibroblasts nucleofected with S/MAR vectors resulted in 3 AP-positive colonies (0.005%) and 2 colonies that showed some degree of MET but were AP-negative **(Figure 79a).** Both reprogramming and stainings were performed in quadruplicates of 15.000 cells each. The results are expressed as a total number of AP-positive colonies/total number of cells. Also, GFP-expressing colonies were found in both S/MAR and EBNA-1-reprogrammed hiPSC as a result of co-transfecting the labelling vector pSMARt **(Figure 79b),** suggesting that pSMARt vectors were also capable of surviving human reprogramming.

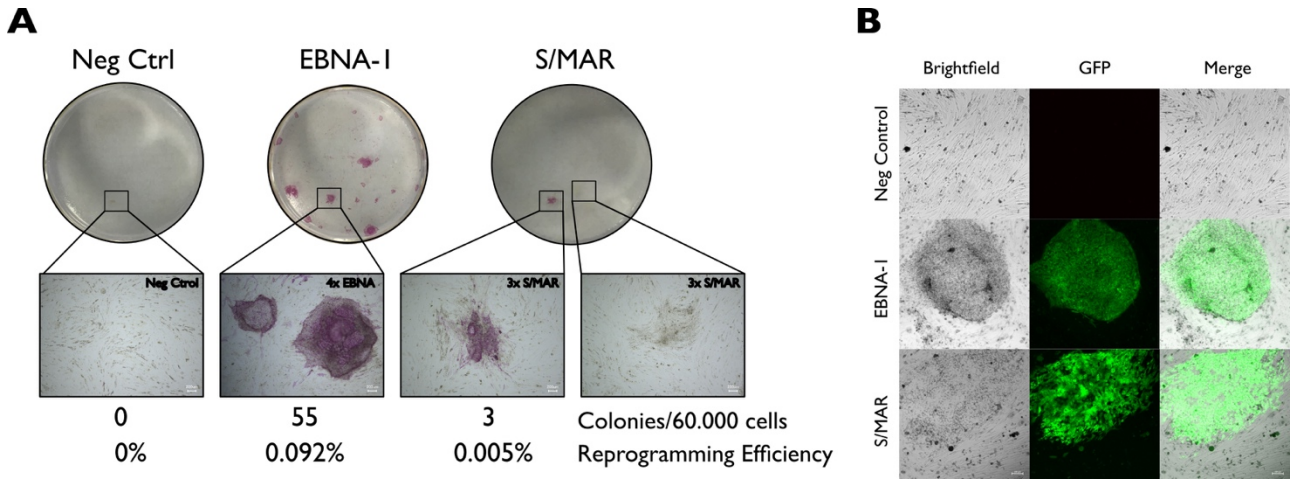


Figure 79: S/MAR vectors reprogram hiPSC less efficiently than EBNA-1 vectors
Wildtype NHDFs were nucleofected with either EBNA-1 or S/MAR vectors, in combination with the labelling vector pSMARt. Untransfected fibroblasts were used as negative control. After 30 days under reprogramming conditions, the cells were fixed and stained for Alkaline Phosphatase (AP). The reprogramming and stainings were performed in quadruplicates of 15.000 cells each (60.000 cells in total). A total of 55 AP-positive colonies emerged from EBNA-nucleofected cells (0.092% reprogramming efficiency), whereas 3 S/MAR colonies stained positive for AP (0.005%) **(A).** Also, NHDF cells were co-transfected with the GFP-labeling vector pSMARt **(B),** and GFP-expressing hiPSC were also obtained (pictures from two independent experiments).

5.6.1.3 S/MAR-labelled hiPSC retain expression through *in vitro* differentiation

pSMARt-labelled or unlabelled hiPSC generated with EBNA-1 vectors were forced to form Embryonic Bodies (EBs) and allowed to randomly differentiate into representatives of the three germ layers. For this, hiPSC colonies were passaged and plated into ultra-low attachment plates (Corning), in the presence of hiPSC media + 10 μM Rock inhibitor. The growing EBs were transferred to a gelatinised black plate and allowed to expand and differentiate in the absence of differentiation inhibitors.

pSMARt-labelled hiPSC remained expressing GFP through EB formation and differentiation, as observed in **Figure 80.**

After two weeks, differentiated structures could be observed microscopically, and the cells were then fixed, permeabilised and stained for ectodermal (ß3-Tubulin), mesodermal (α-SMA) and endodermal (FoxA2) markers **(Figure 81).** Both pSMARt-labelled and unlabelled cells were able to differentiate into all germ layers, although endodermal structures were difficult to obtain. pSMARt-labelled hiPSC retained expression of GFP **(Figure 81b).**

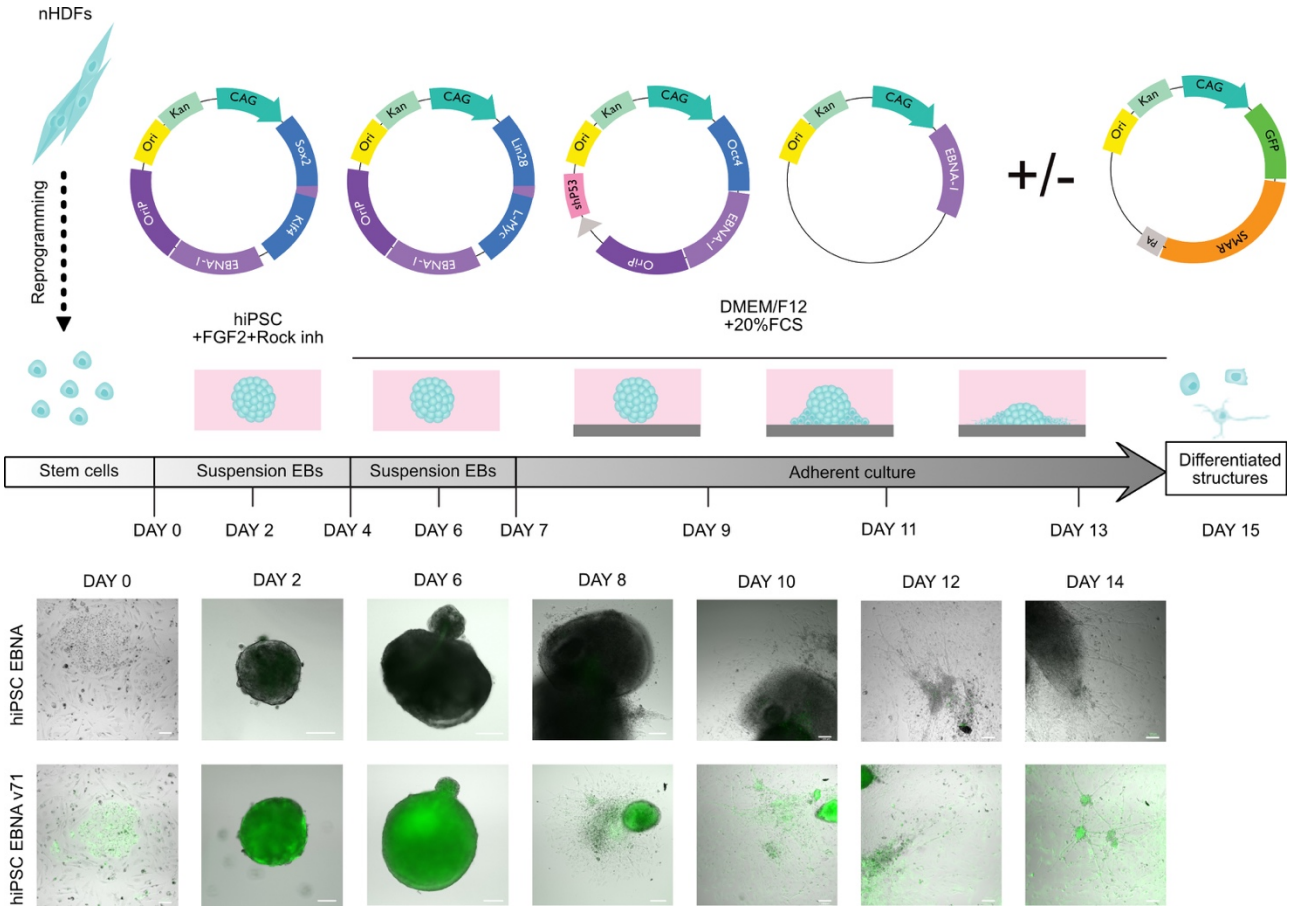


Figure 80: S/MAR vectors are expressed in EBs and through hiPSC differentiation
To generate human Embryonic Bodies (EBs), stem cell colonies were scratched, and hiPSC clumps were allowed to collapse in ultra-low attachment plates in the presence of hiPSC media with 10μM Rock inhibitor. After 7 days, the EBs were transferred into gelatinised plates in the presence of DMEM 20% FCS and allowed to attach and spread, while the proliferating cells differentiated into representatives of the three germ layers. Both unlabelled (negative control) and pSMARt-labelled hiPSC formed EBs that differentiated into defined structures. pSMARt cells kept the expression of GFP throughout the differentiation process. Scale bars=100μm.

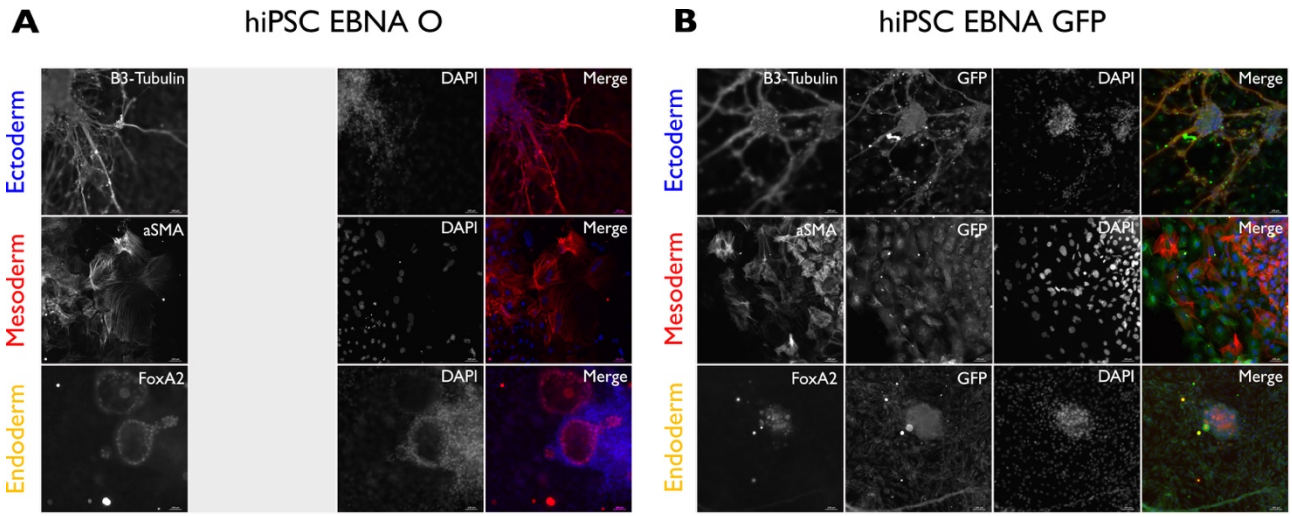


Figure 81: S/MAR-labelled hiPSC can differentiate into derivatives from the three germ layers and express GFP
After formation of EBs, differentiating structures, such as neurons or tubular structures, could be observed microscopically. The cells were then fixed, permeabilised and stained for markers from the three embryonic layers: ectoderm (ß3-Tubulin), mesoderm (α-SMA) and endoderm (FoxA2). Both unlabelled and pSMARt-labelled hiPSC differentiated into all three germ layers. Also, pSMARt-labelled cells remained expressing endogenous GFP. Scale bars=100μm

5.6.2 Assessment of nPOP as a hiPSC reprogramming tool

Because human fibroblasts could be routinely and easily transfected, overcoming the problem of DNA delivery in murine fibroblasts, we sought to investigate the suitability and efficiency of reprogramming of the previously described **(Section 5.1.10)** S/MAR pentacistronic vectors coding for Oct4, Klf4, Sox2, cMyc and the reporter gene dTomato (OKSMdTom) and compare them to the originally described 4in1 Lentivirus [285, 297]. The S/MAR vector (POP) contained the aforementioned pentacistronic cassette (OKSMdTom), whereas its minimally sized version (nPOP), was devoid of bacterial sequences. An additional aim of this experiment was to investigate the effect of bacterial sequences in transgene expression and reprogramming. We hypothesised that plasmids with reduced or absent bacterial sequences would be 1) smaller and therefore, easier to deliver to cells; and 2) less prone to silencing, which would yield to higher and longer expression of reprogramming factors.

5.6.2.1 S/MAR reprogramming vectors can reprogram hiPSCs

Also in collaboration with Prof. Mckay, we nucleofected low passage CLN7 474 HDFs with either POP or nPOP. Also, the cells were co-transfected with a pSMARt-GFP-shP53 labelling vector with shRNA against p53, since cells with reduced p53 levels have been reported to reprogram into hiPSC with better efficiencies than normal cells [301]. Fibroblasts transfected with pSMARt-GFP-shP53 were used as negative control, to discard the effect of either S/MAR or knockdown p53 in reprogramming. The transfected cells were kept under reprogramming conditions for a month. As observed in S/MAR

reprogrammed hiPSC **(Section 5.6.1.1),** POP and nPOP nucleofected cells also showed a delay in MET and reprogramming. We could only reprogram hiPSC with nPOP, while POP failed to generate fully reprogrammed colonies, as observed by the remaining expression of exogenous factors (dTomato) and the aberrant colony morphology. Also, we observed two nPOP hiPSC colonies which also expressed GFP, indicating a successful co-transfection, establishment and survival of pSMARt during reprogramming. Moreover, GFP-negative hiPSC colonies (without shP53) were also obtained, suggesting that shp53 might be desirable but not dispensable for human reprogramming.

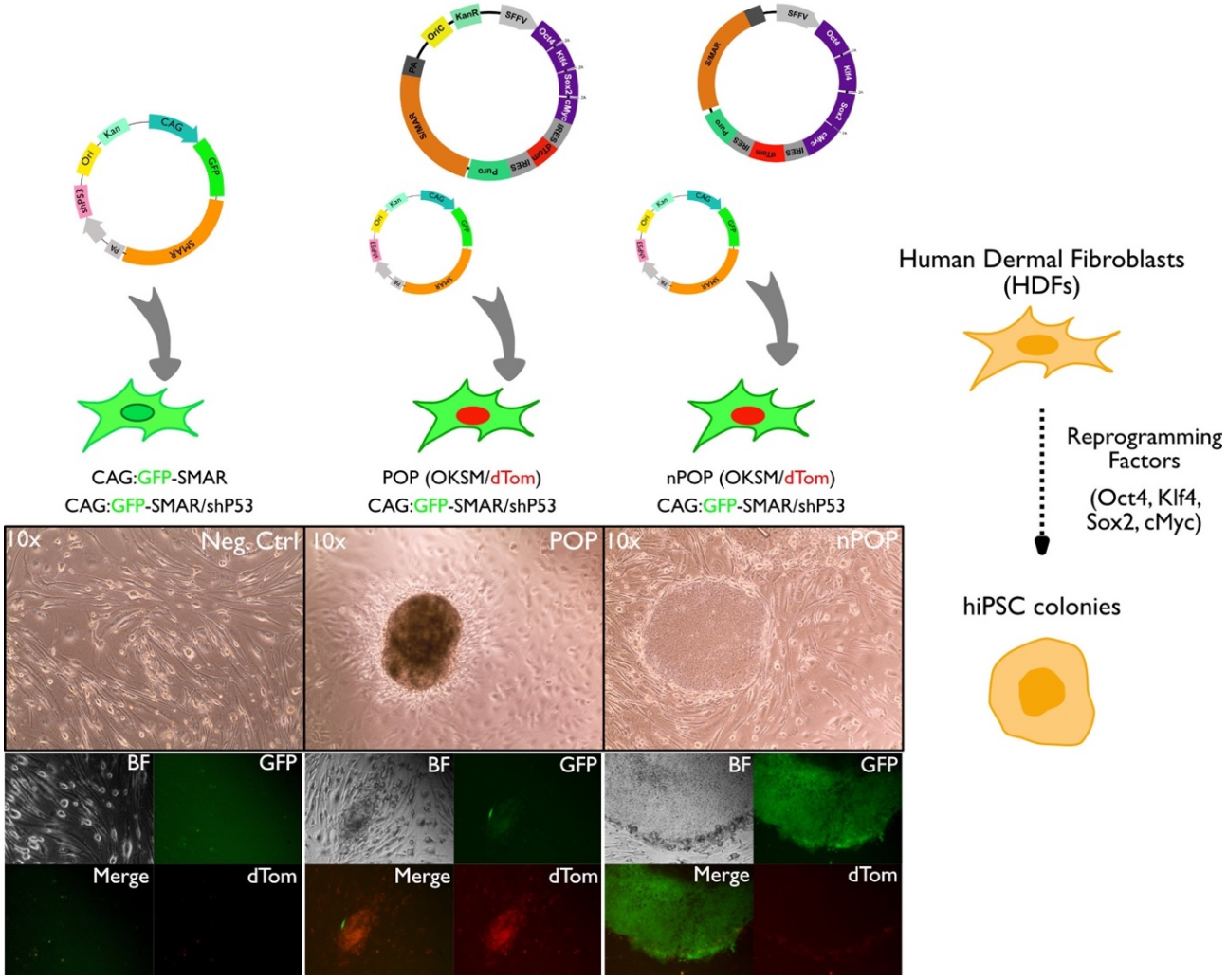


Figure 82: Assessment of the reprogramming capabilities of POP and nPOP in HDFs
Patient CLN474 Human Dermal Fibroblasts (HDFs) were co-nucleofected with a labelling pSMARt-GFP-shp53 vector and either an S/MAR vector containing the human reprogramming factors Oct4, Klf4, Sox2, c-Myc with a bacterial sequence containing OriC and antibiotic resistance (POP) or a minimally sized version (nPOP), devoid of bacterial sequences. After a month of reprogramming, POP failed to reprogram hiPSC fully and generated only partially reprogrammed colonies, which displayed an aberrant morphology and remained expressing the exogenous factors (dTom); whereas nPOP generated cell colonies resembling the morphology of hiPSC. Also, some of the nPOP-generated colonies expressed GFP (pSMARt-GFP-shP53), demonstrating that S/MAR vectors can survive and establish during the reprogramming process.

5.6.2.2 Comparison between reprogramming efficiencies of Lentivirus and S/MAR vectors

After successfully reprogramming patient HDFs with nPOP, we wondered how efficient the pentacistronic S/MAR vectors would be as compared to the state-of-art 4in1 Lentivirus, from which they were subcloned. Due to limited access to patient fibroblasts, we performed this comparison in low passage wildtype NHDFs (Promocell). For this, the cells were either nucleofected with the S/MAR vectors POP and nPOP or transduced with 4in1 Lentiviral particles. Scrambled Lentiviral particles based on the pWPI vector were used as a negative control. At day 30, the cells were fixed and stained with the Alkaline Phosphatase (AP) Staining Kit II (Stemgent).

Out of 60.000 cells that were allowed to reprogram, 4in1 Lentivirally transduced cells showed more colonies (n>400) than nPOP (n=25) and than any other episomal reprogramming system. The reprogramming and staining were performed in quadruplicates of 15.000 cells each. The results are expressed as a total number of AP-positive colonies/total number of cells. Although the efficiency of reprogramming in Lentivirally transduced cells was higher than nPOP (0.667% vs 0.042%), the morphology and quality of the AP-positive colonies did not resemble those from human stem cells (for example, EBNA-generated hiPSC or hESC), but rather the aberrant morphology obtained with POP vectors. No hiPSC colonies were observed in POP transfected cells **Figure 83.**

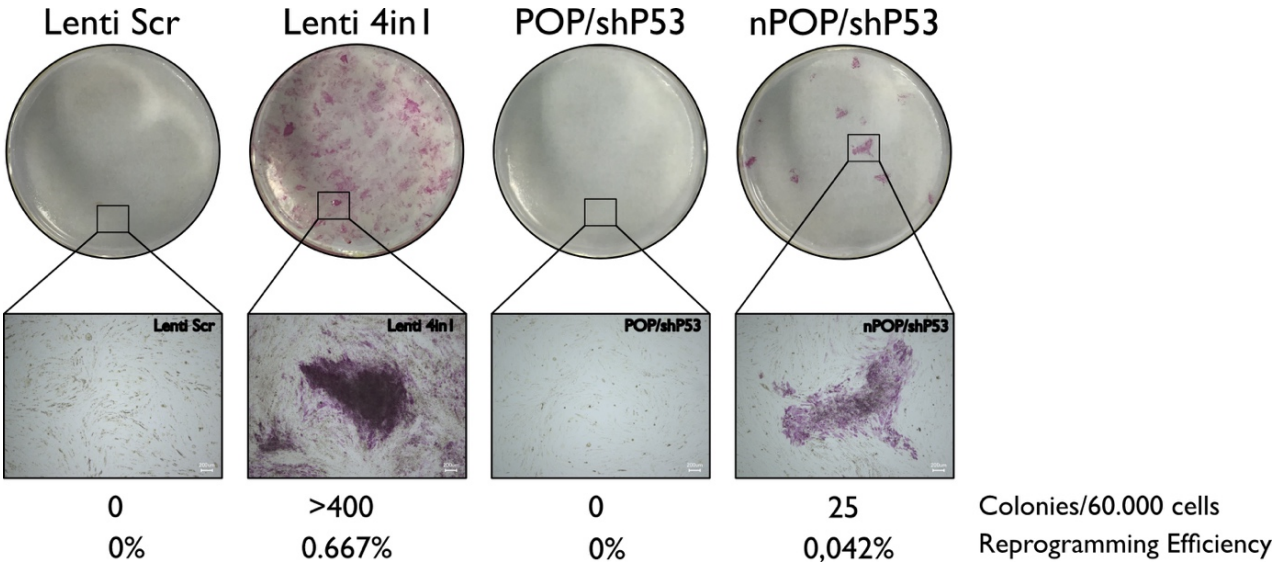


Figure 83: Reprogramming efficiencies of Lentivirus 4in1 and S/MAR-transfected HDFs
Wildtype NHDFs were either nucleofected with either POP or nPOP and co-transfected with pSMARt-GFP-shp53 or transduced with 4in1 Lentiviral particles. Scramble particles were used as negative control. After 30 days of reprogramming, the cells were fixed and stained for Alkaline Phosphatase (AP). The reprogramming and staining were performed in quadruplicates od 15.000 cells each (60.000 cells). More than 400 AP-positive colonies emerged from 4in1 transduced cells (0.667% reprogramming efficiency), whereas 25 nPOP colonies stained positive for AP (0.042%). No colonies emerged from POP-transfected cells.

5.6.3 Summary

In this chapter, three systems were tested and compared for the reprogramming of human fibroblasts: our S/MAR episomal vectors were compared to the state-of-art of episomal reprogramming based on EBNA-1/OriP coding for OKSML; as well as compared to the Lentiviral 4in1 construct.

When compared to other episomal systems, such as EBNA-1/OriP, S/MAR vectors showed a slight delay in reprogramming and a slightly lower reprogramming efficiency. However, their pluripotency and quality were comparable to those hiPSC colonies obtained with EBNA-1 vectors. Also, GFP-expressing hiPSC colonies were obtained when reprogramming fibroblasts were co-transfected with a pSMARt_GFP vector, suggesting that pSMARt vectors were established and kept their functionality during human reprogramming.

When the pentacistronic vectors POP and its bacterial backbone-reduced counterpart nPOP were compared to the described 4in1 Lentivirus, we observed a much better reprogramming efficiency in lentivirally transduced cells (0,667%) than in S/MAR transfected cells. However, those lentiviral-reprogrammed hiPSC showed aberrant morphology and failed to grow after passaging. No POP-hiPSC colonies were obtained, whereas the minimally sized nPOP, devoid of bacterial sequences, managed to reprogram hiPSC colonies with a 0,042% of efficiency.

6. DISCUSSION

6.1 Overview

The use of episomal DNA vectors, entirely comprised of human elements, to persistently modify stem cells has never been shown before. Therefore, we aimed to develop episomal DNA vectors based on S/MAR motifs as a tool to persistently genetically modify Stem Cells, which is typically achieved by using integrating viral vectors or virally-derived episomal vectors. Episomal vectors offer several advantages over integrative systems, as they avoid non-specific integrations into the host genome and the risk of cellular transformation; as well as over virally-derived episomal vectors, which still retain some oncogenic potential and can be detected as 'pathogenic' by the innate immune system, leading to silencing and short expression of the transgene. In particular, episomal vectors based on S/MAR motifs are attractive alternatives to virally-derived vectors as they can deliver sustained levels of transgene expression without compromising the viability of the host cell. However, the ability of S/MAR vectors to sustain high levels of transgene expression through different cellular states (differentiated vs pluripotent) without getting integrated or silenced was unknown to us.

6.2 Summary of results

The suitability of the originally described **pEPI** minicircle was first assessed **in mESC**. However, its performance was poor, and the transgene expression was heterogeneous and weak. Therefore, the **vector design was refined and improved** in all its functional aspects. First, the selection marker was replaced and relocated inside the transcription unit (pSMARt). Second, anti-repressive elements, such as UCOE or Element 40, were added to enhance expression and improve the vector establishment (pSMARt.1). Third, bacterial sequences were removed to decrease immunotoxicity and enhance the transgene expression (nSMARt). Fourth, the original -interferon S/MAR was replaced by a smaller and more effective S/MAR (pSMARter). Finally, the splicing of the S/MAR motif from the mRNA transcript resulted in better expression and establishment (nSpliced). Then, the activity of different promoters was evaluated, which resulted in the CAG promoter being the most suitable promoter for pluripotent cells' work. The vector series pSMARt, nSMARt were used for further *in vitro* and *in vivo* studies, whereas the suitability of both nSMAR_spliced and pSMARter should be further evaluated.

The reprogramming and differentiation survival of pSMARt vectors was first assessed in an all-trans retinoic acid **(ATRA)-mediated neuronal differentiation model,** in which neuroblastoma cells were

forced to differentiate into neurons upon addition of ATRA. When ATRA was withdrawn, the neurons de-differentiated and recovered their neuroblastoma morphology. The pSMARt expression remained active throughout **'differentiation' and 'de-differentiation'**, suggesting that the vector was neither lost nor silenced during the processes. Molecular analysis revealed that pSMARt was kept at **low copy numbers** (2-4 copies/cell) and could be rescued in its **episomal form**. More importantly, the vector did **not modify the cells' behaviour** as observed by expression of the neuronal markers tested, and by the retention of engraftment and tumorigenic potential when the cells were injected both subcutaneously and intracranially into SCID.

Then, both pSMARt and nSMARt vectors were used to **modify mESC persistently** and to **reprogram** somatic cells into **miPSC.** Molecular analysis of the vectors demonstrated that were kept at **low copy numbers** (1 - 2 copies/cell) and could be detected and rescued in their **episomal forms.** The presence of **integrated forms could not be detected** via Southern Blot but cannot be excluded; thus requiring further investigation and integration analysis.

The impact of S/MAR vectors into the host stem cell was evaluated by assessing their pluripotent potential and differentiation capabilities. S/MAR-modified stem cells stained positive for all the **pluripotency markers** tested and were able to **differentiate** into representatives of the three-germ layers when **randomly differentiated *in vitro***, while still being able to express the reporter gene GFP. Directed **hematopoietic differentiation** revealed that S/MAR vectors could deliver stable and high levels of the transgene, even at the HSC stage.

The pluripotent capabilities, as well as the differentiation potential of modified stem cells, were further challenged during ***in vivo* differentiation**. Modified mESC were injected into early-stage embryos and resulted in the formation of **chimeric mice**. The presence and expression of S/MAR vectors were observed in organs derived from all embryonic layers. Also, pSMARt vectors were microinjected into the pronucleus of 1-cell stage zygote and established in the absence of selection, resulting in **transgenic mice** that also expressed the vector across different tissues.

The behaviour of S/MAR vectors during meiosis was addressed by investigating their **germ-line transmission** in animals generated with both S/MAR-modified mESC or vectors microinjected into the pronucleus. Although modified stem cells could be transmitted into the offspring of chimeric animals, as agouti pups were obtained, the presence of S/MAR vectors in the F1 generation was rarely detected, suggesting that the vector was lost during meiosis.

The survival of S/MAR vectors during **reprogramming** was also addressed. For that, S/MAR-GFP vectors were used to label both **human and murine fibroblasts**, which were then **reprogrammed into iPSC** by delivering the reprogramming factors OKSM(L). The vectors could survive reprogramming without getting silenced, as observed by persistent expression of the transgene GFP. The ability of S/MAR vectors as **reprogramming tools** was also addressed by comparing them to the current state-of-art for integrative and episomal reprogramming. Although less effective than EBNA-1 episomal vectors or lentivirus, S/MAR vectors were able to reprogram murine and human fibroblasts into *bona fide* iPSC, which expressed all pluripotency markers tested and were able to differentiate *in vitro* into derivatives of the three germ layers.

6.3 Improvement and refinement of S/MAR vectors

6.3.1 pEPI vector is not stable in mESC

Although several studies have shown that S/MAR vectors can transfect all cell lines tested [126], no previous work has reported the use of S/MAR vectors to modify stem cells persistently or to derive iPSC. Understanding on how S/MAR vectors function in SCs as well as their limitations, was necessary to set up the foundations of this work. Therefore, the suitability of the originally described pEPI minicircle was evaluated in mESC.

Although the pEPI minicircle can transfect mESC and GFP+ positive clones are obtained, the expression is heterogeneous and almost lost in most cases **(Figure 23).** Different reasons can explain this loss of expression:

1) Epigenetic silencing of the CMV promoter due to methylated CpG regions. This issue can be solved by depleting CpG regions as showed by the CpG vector pEPIto [170].

2) Epigenetic silencing of viral promoters in stem cells [303], which can be circumvented by placing an insulating UCOE element, as reported in [198, 226].

6.3.2 Evolution of S/MAR vector

The low expression of GFP in mESC suggests that the first described pEPI vector requires further development, regarding the choice of promoter, insulating elements and depletion of CpG motifs; as well as reconsideration about the general configuration of the vector. For that several improvements have been done:

1) Building of a vector backbone **devoid of SV40 origin of replication** and that relies exclusively on the S/MAR motif for episomal replication and maintenance. The fact that S/MAR minicircles can be episomally retained [295] makes the presence of the SV40 origin of replication not only redundant but also makes the vector more prone to silencing due to the recognition of viral sequences.

2) **Substitution of** the dual kanamycin/G418 **selection marker** for Puromycin and coupling to the transcription unit and the S/MAR motif, making the establishment of the vector part of an active process that can be directly controlled by the amount of selective pressure applied to the cells. In other words, the more Puromycin the cells receive, the more vector they need to survive

and therefore, the more they are forced to retain the S/MAR vector. The use and coupling of puromycin to the transcription unit result in a 5-fold increase in reporter gene expression **(Figure 24).**

3) Placing of **different molecular linkers** between transgene and Puromycin. When IRES or 2A separating elements are used to link GFP and puromycin, the establishment efficiency is higher with 2A sequences as compared to using an internal ribosome entry site (IRES) **(Figure 25),** probably due to the random binding of the ribosomes to the IRES versus an equimolar expression of both proteins due to the self-cleaving 2A sequence.

4) Introduction of **anti-repressive elements** such as Ubiquitous Chromatin-Opening Element (UCOE) or Element 40, which are *cis*-acting sequences known to increase the episomal vector expression and establishment [304, 305]; as well as to increase transcript and protein stability [306] by preventing the spreading of repressive chromatin marks (histone deacetylation and methylation patterns) to neighbouring DNA sequences. When these elements are placed upstream of the CMV promoter, the reporter gene expression is increased by 4-fold **(Figure 24).** Also, a statistical improvement on vector establishment is also observed in vectors containing UCOE or Element 40 **(Figure 25).** However, the use of UCOE elements was dropped due to their intellectual property protection.

5) **Reduction of bacterial backbone.** Bacterial sequences are recognised by the immune system, which results in vector silencing [117, 135]. Therefore, vectors with reduced or absent bacterial backbone are less prone to silencing and yield higher and more persistent levels of transgene expression [307 - 309]. According to work published by Argyros *et al.*, the minimally sized S/MAR minicircles yielded higher and more sustained levels of transgene expression in the absence of selection, both *in vivo* and *in vitro* [308]. Accordingly and in collaboration with Nature Technology Corporation (NTX), we have created the nano vector generation. These vectors rely on the RNA-OUT antibiotic free-selection system, which overall reduces the size of the bacterial backbone by almost 30% and the overall vector size by 1,5kb, which facilitates the DNA delivery in difficult cells such as primary human material. Last but not least, one of the most essential features of the Nanoplasmid™ technology is their approved use in humans by the Food and Drug Administration (FDA).

The removal of bacterial backbone and an overall decrease of CpG motifs results in not only statistically significant improvement of vector establishment **(Figure 26)** but more importantly, into a less impact on cell transcription **(Figure 27).** CpG islands are not only more frequent in prokaryotic genomes but they are also ordinarily unmethylated [170]. This pattern change is recognised by the human innate system via toll-like receptor 9 signalling [134], activating a cascade of inflammatory cytokines that elicit an inflammatory response against the vector and result in its silencing [137]. According to this, our expression profile data reveals that pSMARt vectors alter 1% of the cells' expression profile but more interestingly, it reflects the minimal impact of minimally sized nano vectors (nSMARt) into the cells' transcriptome, which accounts for only 0,05% of genes being altered.

6) **Replacing the long 2000bp S/MAR** sequence from the β-interferon gene by a smaller and more efficient sequence, which results in a better vector establishment efficiency **(Figure 28).**

7) **Splicing and removal of the S/MAR motif from the transcript**. It has been reported that the S/MAR motif has to be part of an actively transcribed unit to be kept active [149] and that its decoupling from transcription will result in either loss or integration of the vector. Although transcription is necessary for the S/MAR functioning and hence, is required for the S/MAR vectors to be kept episomal; the presence of a 2kb S/MAR region in the transcript is unnecessary. Therefore, its splicing and removal from the mRNA can potentially enhance the stability of the transcript and its expression. As expected, when the S/MAR motif is spliced out of the transcript, the RNA stability increases (data not shown), probably due to binding of splicing proteins which protect the transcript from degradation. The longer life of the transcript (expressing puromycin) in the cell results in increased vector establishment **(Figure 28).**

6.3.3 CAG promoter is suitable for stem cell work

The choice of promoter is mainly determined by the purpose of the vector and the cell type in which has to be expressed. Although in general viral promoters, such as CMV or SFFV, yield high levels of transgene expression, they eventually become silenced due to extensive methylation of CpG islands in pluripotent and differentiating cells, which makes them unsuitable for stem cell work [310]. Therefore, is necessary to find a suitable promoter that is not only able to keep high levels of transgene expression but also that its expression can be kept constant, independently of epigenetic events or cellular states.

Accordingly, we have generated a library of vectors containing a range of viral and mammalian promoters, which have been tested in HEK293T, MEFs and mESC. According to our comparison, the

chimeric CAG promoter, which consists on the cytomegalovirus early enhancer element (C), the promoter, first exon and intron of the chicken β-actin gene (A) and the splice acceptor of the rabbit β-globin gene (G); proves to be the most suitable promoter for both stem cell and fibroblast work. The CAG promoter not only provides the highest levels of expression but is the most stable over time. These findings go in line with those published in [311 - 314], in which the CAG promoter was found to be the most suitable promoter for ESC work.

Although the CAG promoter is suitable when the aim is to overexpress a gene, it might be too strong when the goal is to achieve physiological levels of a therapeutic gene. In such cases, a milder promoter such as UbC or PGK should be considered.

6.3.4 The new generations of S/MAR vectors are persistently expressed in MEF and mESC

After deciding in favour of the CAG promoter, the different vector generations were tested in murine fibroblasts (MEFs) and stem cells (mESC) to determine the most suitable vector configuration for pluripotent cells, including iPSC.

Not surprisingly, nSMARt_CAG has the best transfection efficiency, expression and homogeneity amongst all vectors tested (**Figure 30** to **Figure 32**). Its success can be attributed to the right combination of the promoter (CAG) together with the reduction in the bacterial backbone, which positively influences the transfection efficiency and the high and stable levels of expression, making the vector less prone to silencing. pSMARt_CAG is the second best performing vector for stem cell work, with slightly heterogenous and lower expression. nSpliced is a CMV-based vector which contains the anti-repressive Element40 to shield the promoter from silencing. This protective combination together with its reduced size and the introduction of splicing sites flanking the S/MAR motif results in a very competitive performance of the vector, comparable to that from nSMARt. However, pSMARter, which also has an insulating element shielding the CMV promoter from silencing, does not perform at the same level than the other vectors, probably due to the presence of bacterial backbone.

When mESC are subjected to hematopoietic differentiation, both nSpliced and nSMARt, provide sustained and high levels of expression throughout differentiation, outperforming the vectors containing bacterial-backbone (pSMARt and pSMARter) as shown in **Figure 61.** This suggests that differences in performance cannot be attributed to the promoter, since shielding the CMV promoter with Element 40 (nSpliced) results in comparable performance to the CAG promoter (nSMARt), but

instead suggests that the bacterial backbone is a hotspot for epigenetic modifications resulting in silencing of the vector. In other words, it is preferable to reduce the bacterial backbone rather than to change the promoter when stable levels of expression are to be achieved.

However, the application of the vector dictates the configuration that this should have. For example, when the purpose is to modify stem cells and their differentiated progeny or to generate transgenics, the vector should be highly and ubiquitously expressed. Therefore, a stable and high-performance vector, such as nSpliced or nSMARt, is preferable. However, if the purpose is to correct physiological levels of a therapeutic gene, a milder vector configuration and promoter should be chosen.

The vector series pSMARt, nSMARt (cloned early in this study) were used for further *in vitro* and *in vivo* studies, whereas the suitability of both nSMAR_spliced and pSMARter (cloned later) should be further evaluated.

6.4 S/MAR vectors survive differentiation and de-differentiation in a neuronal model

One unanswered question and critical point of this study was to elucidate whether S/MAR vectors were able to 'survive' differentiation and reprogramming, in other words, if S/MAR-labelled cells could retain the levels of transgene expression as well as to remain episomal throughout these processes. To gain insight, we used an ATRA-mediated differentiation and de-differentiation model in neuroblastoma cells.

6.4.1 Effects of ATRA on cell cycle and transcription

All-Trans Retinoic Acid (ATRA) or Vitamin A, is a morphogen known to direct neuronal differentiation in ESC as well as to promote neuronal differentiation and induce apoptosis in neuroblastoma cells [315]. These processes are triggered by the binding of retinoic acid (RA) to its receptors RAR and/or RXR, which dimerise and translocate to the nucleus where they bind to Retinoic Acid Response Elements (RAREs) and modulate gene expression [316, 317]. They also recruit protein complexes containing Histone Deacetylase (HDACs), which remove acetyl groups from histones making the overall chromatin charge more positive and increase the attraction to the negatively charges DNA, making the chromatin more compact. This condensation of chromatin limits the transcription of specific genes [318]. In addition to directing neuronal differentiation, ATRA is also known to up-regulate the CDK inhibitors p21 and p27 involved in the regulation of cell cycle checkpoints **(Figure 84).** This results in the arrest of the cell cycle [315] and prevents the cells from proliferating.

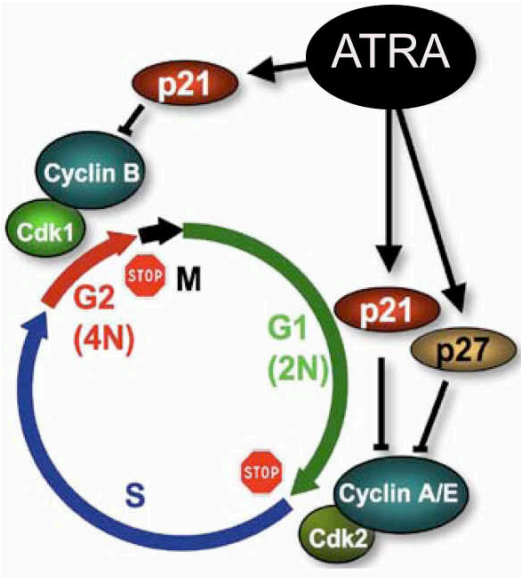


Figure 84: ATRA effects on cell cycle
Schematic depiction of ATRA effects on the cell cycle. Besides its morphogenetic activity and contribution to neuronal differentiation during development, ATRA also up-regulates p21 and p27, which in turn inhibit cdk/cyclin complexes, arresting the cell cycle. Therefore, upon ATRA administration, the cells will undergo cell cycle arrest and stop of proliferation. Modified from Taieb *et al.* [319].

6.4.2 S/MAR persistence through ATRA-mediated differentiation

Taking into consideration the aforementioned effects of ATRA on the cell cycle, and the need of transcription running through the S/MAR for the vector to be kept episomal, four possible scenarios need to be considered when looking at GFP expression and persistence of the vector during ATRA-mediated differentiation **(Figure 85).** After the establishment of the pSMARt_GFP vector, all Be2C cells contain between 2 to 4 copies of pSMARt tethered to the DNA through a protein complex. Upon ATRA administration, the cells could still divide for a while and progressively start differentiating, generating the following scenarios: (1) The vector survives differentiation and remains active and episomal throughout the process, and the cells differentiate into neurons with a functional episomal vector, expressing GFP. (2) The vector is initially episomal and tethered to the DNA, but as the cells keep differentiating, the vector detaches but does not get lost, since the cells do not divide. This result in neurons that are expressing GFP but in which the vector is not functional. (3) The vector does not survive the process from the beginning, resulting in loss or dilution of the vector (and hence GFP expression) in the differentiated population. (4) The vector integrates into the genome, which also results in GFP expressing differentiated cells.

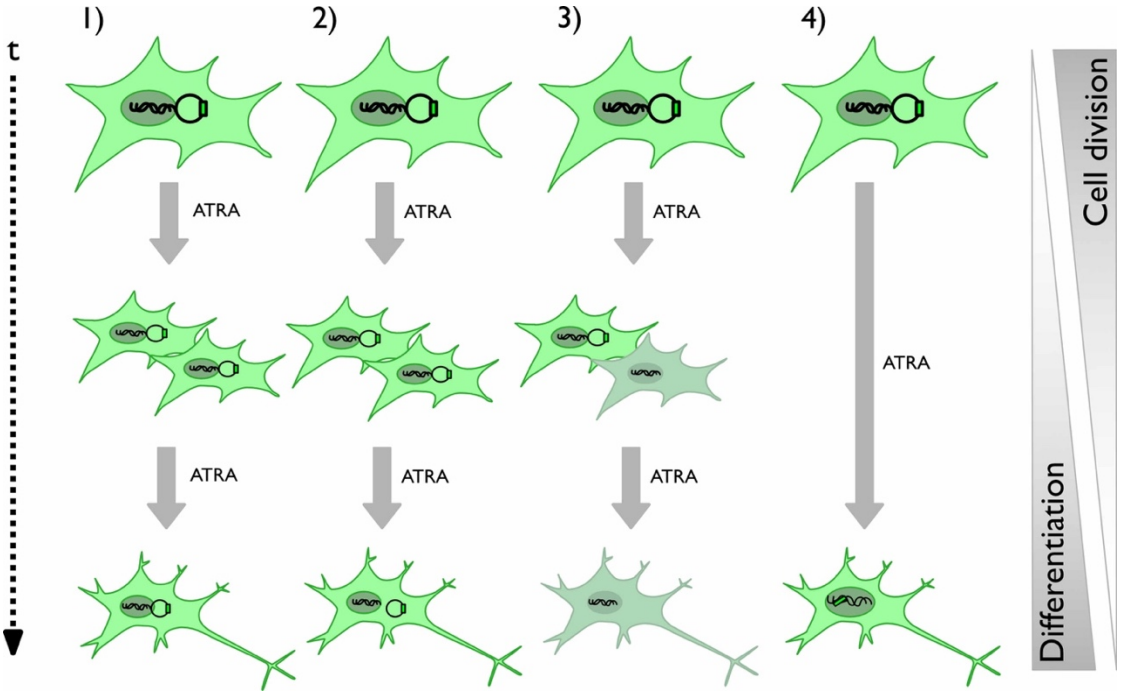


Figure 85: Possible outcomes after ATRA-mediated neuronal differentiation

Upon ATRA administration, neuroblastoma cells kept dividing for a while and progressively started differentiating, slowing cell division until they stopped dividing. Assuming the vector is initially episomal and tethered through a protein complex, there could be different scenarios: **(1)** The vector remains active throughout the process, and the cells differentiate into neurons with a functional episomal vector, expressing GFP. **(2)** The vector does not survive differentiation late in the process (when cells are slowing or not dividing), and cells differentiate into neurons which are expressing GFP, but the vector is not tethered to the DNA. **(3)** The vector does not survive the process from the beginning, resulting in loss or dilution of the vector (and hence GFP expression) in the differentiated population. **(4)** The vector integrates into the genome, which results in continuous expression of GFP.

After inducing 'differentiation' and 'de-differentiation', the cells kept expressing GFP **(Figure 43),** which suggests that the vector is either still tethered to the DNA and has not been lost during the process (outcome 1) or is integrated (outcome 4).

6.4.3 S/MAR does not modify neuroblastoma cells' behaviour

The effect of S/MAR presence in neuroblastoma cells was evaluated by first assessing the effect of the vector in the expression of neuronal markers. Immunofluorescent stainings show that the presence of S/MAR vectors does not modify the expression of N-Cam, Gap43 and β3-Tubulin; and that these markers are expressed both at the neuroblastoma and neuron stage, regardless of the treatment or presence of vector **(Figure 46).**

Then, the engrafting and tumorigenic capabilities of S/MAR-labelled cells were challenged in heterotopic injections into immunodeficient mice. Neuroblastoma is a type of cancer that forms into specific nervous tissues and is most frequently found in adrenal glands or the spine. Although orthotopic adrenal injections are preferred to investigate the biology and treatment of neuroblastoma [320], the injections are complex and require considerable skills to perform. Alternatively, heterotopic injections are easier to perform although differences in tumour environment and blood supply differ from the orthotopic model [321]. Taking advantage of neighbouring expertise in heterotopic injections, pSMARt-Be2C cells expressing GFP or Luciferase were injected either subcutaneously or intracranially into SCID mice, to address whether the cells would engraft and form tumours that expressed GFP. Indeed, the cells formed tumours regardless of being modified with S/MAR vectors and regardless of the transgene expressed (GFP or luciferase) **(Figure 47** and **Figure 48).** These tumours were isolated, homogenised and cultured, and remained expressing the reporter genes **(Figure 49** and **Figure 50).**

This data demonstrates that the pSMARt DNA vectors are capable of genetically modifying dividing and differentiating cells and can provide sustained robust expression of transgenes during 'differentiation' and 'de-differentiation' processes, therefore demonstrating that they are a valuable genetic tool that can be used to generate sophisticated isogenic cell lines without molecular or genetic damage.

6.4.4 S/MAR vectors are retained episomally and kept at low copy numbers in neuroblastoma cells

Up to this point, we focused on investigating the behavioural aspects of S/MAR vectors during differentiation and de-differentiation in the ATRA-mediated neuronal differentiation model, to understand how such processes, influenced the vector's expression and persistence. However, a closer examination of the vector status was missing.

To investigate the molecular state of the vector, a plasmid rescue experiment was performed. The results show that episomal forms of S/MAR vectors can be recovered intact from Be2C labelled cells **(Figure 45).** However, the recovery of episomal forms does not rule out the possibility of coexisting integrated forms of the vector, and therefore, further integration analysis required.

The number of vector copies per cell was also determined in stable Be2C clones. In agreement with published literature [149], we found that S/MAR vectors are kept at low copy number, ranging from 2-4 copies/cell **(Figure 44),** and more interestingly, that the copy number is stable during ATRA-mediated differentiation and de-differentiation (data not shown).

6.5 S/MAR vectors are expressed during reprogramming

The use of S/MAR DNA vectors for the generation of iPSC has never been described. Originally, iPSC derivation was accomplished by retroviral transduction of reprogramming factors Oct4, Klf4, Sox2 and cMyc [142]. Later, safety concerns regarding the oncogenicity of cMyc were addressed by using an excisable lentiviral polycistronic cassette [240], which allowed the removal of unnecessary reprogramming factors once pluripotency was achieved. Alternatively, transposons were also used to generate iPSC efficiently [242]. However, they rely on integrating factors that could leave a 'genomic scar' upon excision. Non-integrative methods such as Adenovirus [91], Sendai RNA virus [245], polycistronic minicircles [307] or direct delivery of mRNA and proteins [248] were used as safer alternatives, albeit compromising the efficiency of reprogramming. The current state-of-art episomal reprogramming system relies on the combined transfection of four EBNA-1 based plasmids [203]. Although able to efficiently generate iPSC, EBNA vectors are regarded as potentially oncogenic and might also influence the cells' expression profile [253].

The suitability of pSMARt vectors for labelling dividing and differentiating cells has been tested in a neuronal differentiation model using ATRA, suggesting that S/MAR-based vectors are capable of surviving both differentiation and a de-differentiation process. Then, the vectors were used in a real reprogramming process, from somatic cells (fibroblasts) to iPSC, in both murine and human cells. Not only do S/MAR-labelling vectors survive reprogramming but they also can themselves derive *bona fide* murine and human iPSC, although less efficiently than lentivirus or EBNA-1 vectors.

These results prove that S/MAR vectors are capable of surviving a real reprogramming process and can keep their expression throughout the process, even at the pluripotent state, suggesting that the vectors are not getting silenced or lost during the process. Taken altogether, these data point at S/MAR vectors as useful, cheap and safer alternative to integrative vectors or viral-based episomal vectors for both murine and human reprogramming.

6.5.1 pSMARt-vectors can immortalise murine fibroblasts

It has been reported that immortal cells are more straightforward to reprogram since they overcome replicative senescence [322, 323]. Based on this knowledge, we first labelled and **immortalised lung fibroblasts** by overexpression of SV40LT expressing pSMARt-GFP vectors **(Figure 51).** Although useful and efficient in generating immortal cells, the downside of SV40LT is that it promotes cellular transformation through suppression of p53 and pRb, which result in the cells entering S-phase and

dividing in a dysregulated manner [158]. Although suppressing tumour suppressor genes enhances the efficiency of the reprogramming due to evasion of DNA-damage responses and apoptosis [324] – which is especially relevant and useful for diseased or old patient cells – the manipulation of oncogenes rules out the possibility to use these cells for therapeutic applications. Therefore, a transient silencing of p53 would be desirable, increasing both efficiency and safety.

6.5.2 pSMARt-based vectors retain expression through reprogramming

Stably labelled pSMARt-SV40LT-GFP fibroblasts were reprogrammed via lentiviral delivery of OKSM factors and resulted in miPSCs expressing GFP. This indicates the presence and function of the S/MAR vector through reprogramming **(Figure 52).** However, the exogenous reprogramming factors are not switched off in immortalised miPSC colonies, which results in the inability of these cells to differentiate properly. Therefore, the use of SV40LT-immortal cells for reprogramming and differentiation experiments was dropped, and primary cells were used for further experiments.

Next, **human dermal fibroblasts** were co-transfected with both EBNA or S/MAR-reprogramming vectors containing the OKSML factors and the labelling vector pSMARt_GFP. Successfully reprogrammed GFP+ hiPSC colonies were obtained, which indicates that pSMARt vectors establish and remain active through reprogramming **(Figure 79b).** However, the frequency of GFP+ hiPSC is very low, which might be attributed to the simultaneous co-transfection of four EBNA reprogramming vectors or three S/MAR reprogramming vectors together with the labelling vector pSMARt. In another experiment, patient CLN474 fibroblasts were also co-transfected with nPOP and pSMARt_GFP, resulting in GFP+ hiPSC **(Figure 82),** which evidences again the ability of S/MAR vectors to survive reprogramming.

6.5.3 S/MAR vectors can reprogram murine and human fibroblasts

After confirming the ability of S/MAR vectors to establish and sustain transgene expression during reprogramming, we sought to address the capabilities of S/MAR vectors as reprogramming tools themselves. For that, two comparative experiments were performed. In the first one, the S/MAR-based vectors POP and nPOP were used in parallel with the integrative Lentivirus 4in1. In the second experiment, S/MAR-based vectors were compared to the episomal state-of-art EBNA-1 vectors.

6.5.3.1 S/MAR vectors reprogram miPSC although less efficiently than Lentivirus

First, **mouse embryonic fibroblasts (MEFs)** were transduced with polycistronic vectors coding for Oct4, Klf4, Sox2 and cMyc. The Lentivirus 4in1, which was used as positive control, could reprogram MEFs, whereas both POP and nPOP failed to do so in the first attempts. One possible explanation is that both S/MAR vectors have a very low transfection efficiency, probably due to their larger size (11kb) as compared to normal pSMARt-labelling vectors (around 7kb). Despite being low, the transfection efficiency of nPOP is slightly higher than POP, probably due to its reduced size and lack of bacterial backbone. In the next attempts, repeated transfections of S/MAR vectors resulted in enough levels of reprogramming factors to achieve reprogramming in nPOP-transfected cells, but not in cells transfected with POP **(Figure 53).** We hypothesise that the reprogramming success of nPOP is due to its reduced bacterial sequences, which results in the vector being less prone to silencing and therefore, in more prolonged and higher levels of reprogramming factors. On the contrary, POP cannot sustain enough levels of OKSM to induce reprogramming.

We then tried the S/MAR reprogramming vectors POP and nPOP in **human dermal fibroblasts**. First, we tried in **CLN474 cells** from a patient with Batten disease, due to unreported observations that these cells reprogram better than wildtype cells. Similarly to MEFs, we could reprogram CLN474 cells using nPOP but not POP **(Figure 82),** probably because of its smaller size and better transfection efficiency as well as its less tendency to silencing, hence providing higher levels of reprogramming factor expression and for a longer time. However, when we repeated the same experiment in **wild-type human dermal fibroblasts** (Promocell), we were unable to accomplish successful reprogramming neither with the lentivirus, POP or nPOP. Cells transduced with Lentivirus and transfected with nPOP, stained positive for AP (0.667% vs 0,042%) but showed aberrant morphology and failed to progress to a distinct hiPSC morphology **(Figure 83),** suggesting that the reprogramming factor combination (OKSM) is not suitable for human reprogramming or that the expression of reprogramming factors is too high or is expressed for too long, so the cells cannot switch off the exogenous factors and rely on their endogenous pluripotent machinery.

In summary, we show that S/MAR reprogramming vectors based on the OKSM expression cassette (nPOP) can induce pluripotency in murine fibroblasts, although less efficiently than the Lentivirus. Also, the reprogramming efficiency, as well as the miPSC quality, seem to be dependent on the levels

of expression of reprogramming factors, which in turn depends on the vector and the transfection efficiency, making nPOP but not POP a suitable tool for murine fibroblast reprogramming.

The results also show that only nPOP can reprogram patient-derived cells but not wildtype fibroblasts, suggesting that there must be a relationship between this particular patient-fibroblast line and reprogramming. However, this relation is out of the focus of this study. Not surprisingly and in agreement with our hypothesis, POP is not able to generate fully reprogrammed colonies, as observed by the remaining expression of exogenous factors (dTomato) and the aberrant colony morphology.

When used in wild-type fibroblasts, both lentiviral transduced and nPOP transfected cells develop into colonies with aberrant morphology, suggesting that the combination of OKSM factors might not be suitable for human fibroblast reprogramming.

6.5.3.2 S/MAR vectors can reprogram *bona fide* hiPSC but less efficiently than EBNA-1 based vectors

We then compared the reprogramming efficiencies of EBNA-1 and S/MAR-based vectors. The three S/MAR vectors or the four EBNA vectors (**Figure 37**), both systems expressing the reprogramming factors OKSML, were cotransfected together with a labelling pSMARt-GFP vector.

In collaboration with Prof. Mckay (MMU), we first tested and compared the reprogramming efficiencies of EBNA and S/MAR vectors in **CLN474 patient fibroblasts.** Although we observed a slight delay in reprogramming and a slightly lower amount of AP colonies in S/MAR transfected cells (data not shown), successful reprogramming was achieved with both EBNA and S/MAR vectors. Once reprogramed, both EBNA and S/MAR hiPSC are morphologically indistinguishable and stained positive for all pluripotent markers tested **(Figure 78).**

We then compared the two vector systems in **wildtype nHDFs.** Similarly to patient-derived cells, S/MAR vectors were also slower and less efficient at reprogramming than EBNA (0.005% vs 0.092%) **(Figure 79).** In this case, however, S/MAR transfected cells looked partially reprogrammed, although they stained positive for AP, but failed to progress to a distinct hiPSC morphology.

The differences in kinetics and reprogramming efficiency between S/MAR and EBNA-based vectors could have several explanations:

1) A potential reprogramming synergy of EBNA-1 by binding to the enhancer region of c-Myc [157], which is not present in the cocktail of delivered reprogramming factors. This could result in upregulation of the oncogene, adding a 'switched on' reprogramming factor, making the process faster and more efficient.

2) Binding of EBNA-1 to cellular promoters, resulting in dysregulation the cells' transcriptome and genes associated with cell growth [325].

3) Better expression of S/MAR vectors, due to their reduced size and absence of viral sequences, resulting in excessive exogenous factor delivery. The excess of reprogramming factors might difficult the 'switch off' of exogenous reprogramming factors and 'switch on' of the endogenous pluripotency machinery.

Overall, both morphology and immunofluorescence images suggest that there are no apparent differences between EBNA-1 and S/MAR reprogrammed hiPSC albeit a slight delay in the S/MAR reprogrammed cells. However, a closer look at the effects of EBNA in reprogramming is missing. We are currently conducting a microarray analysis in EBNA and S/MAR modified cells, which will provide a closer insight into the effect of both viral and mammalian sequences in the cells' transcriptome.

In summary, we show that S/MAR vectors can reprogram human fibroblasts and that once reprogrammed, both morphology and quality of both EBNA and S/MAR hiPSC are indistinguishable after few passages. The possible synergising effects of EBNA in reprogramming, together with its oncogenic potential, point at S/MAR-vectors as a safer reprogramming alternative. Finally, the generation of GFP-positive hiPSC colonies from the co-transfection of pSMARt-labeling vectors together with reprogramming vectors, suggests that S/MAR vectors are not only capable of reprogramming human cells but also that can survive the reprogramming process and establish as episomal entities.

6.6 S/MAR vectors are kept at low copy numbers and are found episomally in murine pluripotent cells

After assessing the abilities of S/MAR vectors to survive differentiation and reprogramming, the next logical step was to perform molecular analysis to determine the **number of copies per cell.** In agreement with the literature [195], molecular analysis of pSMARt and nSMARt labelled mESC show that the vectors are kept at low copy numbers, ranging from 1-2 copies/cell in the selected clones **(Figure 55).**

We then assessed the molecular status of the vector to exclude genomic integrations and prove the vectors' episomal maintenance. For that, a **plasmid rescue experiment**, which allowed recovery of episomal molecules, was performed in pSMARt labelled mESC. Our data shows that pSMARt could be rescued in its circular form both labelled MEFs, miPSC (data not shown) and labelled mESC **(Figure 56),** suggesting that the vectors can be retained episomally in stably labelled cells. Also, the presence of intact S/MAR and expression cassette was confirmed by PCR amplification and sequencing of the respective amplicons, suggesting that the vector remains stable.

Plasmid rescue experiments determine the presence of circular (episomal) forms in labelled cells but do not exclude the possibility of integrated forms. For that, **Southern Blot** was performed, in which both integrated and episomal forms could be detected. However, only nSpliced could be detected as an episome, and we could not rule out the possibility of other vectors being integrated as no band or smear could be detected in a Southern Blot **(Figure 57).** Further optimisation of protocols for episomal DNA isolation and detection is required.

Although the results are not conclusive, other pieces of evidence presented later in this discussion (*e.g.:* lack of episomal germ-line transmission as opposed to integration) contribute to the hypothesis that S/MAR vectors remain episomal. To definitively prove the episomal status of the vector, a thorough integration analysis should be performed. We are currently performing integration analysis of pSMARt-labelled mESC to address this issue. However, the results of these experiments could not be included in this thesis due to time constraints.

6.7 S/MAR vectors do not modify pluripotency of murine and human stem cells

One characteristic of S/MAR vectors that makes them distinct from other gene delivery systems is the absence of virally-derived sequences and proteins. Their design based mainly on human elements, results in minimal influence on labelled cells, with only 1% of dysregulated genes in stable cell lines **(Figure 27a).** When bacterial sequences are removed, the effect drops to only 5 up-regulated genes **(Figure 27b),** resulting in practically isogenic cell lines.

S/MAR-labelled neuroblastoma cells show expression of all neuronal markers tested and retain their engraftment potential, which contributes to evidence supporting the minimal impact of S/MAR vectors in the cells. Similarly, we addressed the effect of S/MAR vectors on modified mESC, miPSC and hiPSC, and most importantly, evaluated if the presence of an S/MAR vector would impair the pluripotent capabilities of stem cells.

First, we examined the **expression of pluripotent markers** in S/MAR-labelled mESC as well as miPSC. The cells were stained with Alkaline Phosphatase, a hydrolytic enzyme highly expressed in undifferentiated cells and one of the markers used for early detection of emerging colonies. Then, the cells were stained with more specific pluripotency markers, such as SSEA-1, Oct4 and Nanog **(Figure 58).** pSMARt-labelled mESC express all pluripotency markers and can, therefore, be considered pluripotent, indicating that the S/MAR vector does not modify or damage the cells. miPSC derived from lung fibroblast also show expression of all pluripotency markers. However, SV40LT-immortalised fibroblasts do not express Nanog, indicating that the internal pluripotency network is not switched on. Also, they keep the expression of the exogenous reprogramming factor (as observed by dTom expression), which reinforces the idea of aberrant reprogramming due to immortalisation and p53 dysregulation.

When the pluripotency was addressed in **human reprogrammed iPSC** labelled with pSMARt-GFP, both expression of AP as well as the human pluripotency markers Tra160, SSEA3, SSEA4 and Nanog **(Figure 78)** was observed, regardless of being reprogrammed with EBNA or S/MAR vectors. This suggests that reprogramming with S/MAR vectors was successfully achieved and that the presence of S/MAR vector does not alter the cells' behaviour.

Another feature of pluripotent cells is their ability to **differentiate into representatives of the three germ layers**. The differentiation potential of both murine and human stem cells was assessed in a random differentiation experiment *in vitro*. pSMARt-labelled cells can differentiate into ectoderm, mesoderm and endoderm representatives **(Figure 60** and **Figure 81),** just like unlabelled stem cells. Persistently expressing differentiated hematopoietic precursors can also be obtained from S/MAR-labelled mESC **(Figure 61).** A more challenging assessment of pluripotency consists of **generating chimaeras** upon injection of stem cells into early-stage embryos. pSMARt and nSMARt-labelled mESC can form chimaeras when injected into early-stage embryos **(Figure 62).** Random three-lineage differentiation, hematopoietic differentiation and chimaera formation experiments are discussed in more detail in the following sections.

The results prove that S/MAR vectors *per se*, do not modify the stem cells and does not prevent them from differentiating either *in vitro* or *in vivo*.

6.8 S/MAR vectors survive *in vitro* differentiation

Previous reports on episomal modification of stem cells failed to show persistent expression of the transgene, either because it was being silenced at the episomal state or during differentiation [26, 225, 226]. We challenged the improved S/MAR vectors were and subjected them to several differentiation tests, both *in vitro* and *in vivo*. We conclusively show that the vectors are not only able to persistently and safely modify stem cells but also that their expression can be maintained during differentiation.

6.8.1 Random differentiation of pSMARt-mESC and miPSC

First, pSMARt and nSMARt-labelled mESC were allowed to **spontaneously differentiate *in vitro*** in the absence LIF and 2i inhibitors. The results show that representatives of the three germ layers can be obtained from both unlabelled and S/MAR-labelled mESC and more importantly, that the expression of GFP, and therefore the S/MAR vector, is maintained during differentiation **(Figure 59).** Although miPSC reprogrammed from wildtype fibroblasts can differentiate into ectoderm and mesoderm (endodermal structures could not be found), pSMARt-SV40LT immortalised miPSC fail to differentiate. This experiments also highlight the importance of the SFFV promoter silencing and loss of exogenous OKSM expression in the pluripotent state as observed by the impossibility of miPSC to differentiate. However, the expression of pSMARt is retained, as observed by persistent expression of GFP.

6.8.2 Random differentiation of pSMARt-hiPSC

In a similar experiment, pSMARt-GFP-hiPSC or unlabelled-hiPSC, both generated with EBNA-1 vectors, were subjected to a random differentiation *in vitro*. Unlabelled and S/MAR-labelled hiPSC can differentiate into representatives of the three germ layers **(Figure 81)** and the differentiated progeny retains expression of GFP throughout differentiation **(Figure 80).**

6.8.3 Directed Hematopoietic differentiation

After assessing the S/MAR vector behaviour in random differentiation experiments, an *in vitro* hematopoietic differentiation was performed from mESC to hematopoietic progenitors. Although not entirely differentiated, HSC represent an intermediate step in between pluripotency and fully differentiated cell and can, therefore, provide some information about the vector behaviour and expression during differentiation and in a desired differentiated cell type.

mESC were stably transfected with different generations of S/MAR vectors, including pSMARt, nSMARt, nSpliced and pSMARter **(Figure 61).** At day 0, mESC cells were stained with a panel of endothelial (VE-cadherin and c-kit) as well as hematopoietic (CD41) markers. mESC were negative for CD41 and slightly positive for c-kit, which is both a stem cell as well as a hematopoietic precursor marker. A slight CD144 (VE-Cad) expression on day 0 is observed, which can be from some residual feeder cells.

At day 6 and following the addition of cytokines and growth factors, mESC acquire hematopoietic features: they acquire expression of the surface marker CD41 and lose expression of endothelial characteristics (c-kit and VE-cadherin). However, expression of c-kit and VE-cadherin should still be observed at the HSC stage [302], suggesting that these cells already acquired a further differentiated state [326]. Independently from the slightly more differentiated state of HSC, the data shows that both unlabelled mESC (negative control) and S/MAR-mESC follow the same marker expression profile and behave in the same way, suggesting once again that the S/MAR vectors do not modify the pluripotent behaviour and capabilities of mESC.

Not surprisingly, vectors devoid of bacterial sequences (nSMARt and nSpliced) result in the most stable and persistent expression of transgene during differentiation. Amongst them, vectors devoid of bacterial sequences provide stable and persistent expression of the transgene. nSMARt (which has a CAG promoter) is the most stable, followed by nSpliced. Vectors containing bacterial sequences (pSMARt and pSMARter) show a decrease in transgene expression, suggesting that the vectors are probably losing expression due to silencing events during differentiation.

6.9 S/MAR vectors survive *in vivo* differentiation

In a more representative and closer-to-development *in vivo* experiment, pSMARt- and nSMARt-mESC were injected into early-stage embryos, which developed into **chimeric mice,** as observed by coat chimerism. When these mice were analysed, GFP was expressed in their tissues, indicating the presence of the vector and its survival from stem cell to a fully developed organism. In another experiment, pSMARt vectors were directly **microinjected into the pronucleus** of 1-cell stage zygotes and were able to generate transgenic mice (in the absence of selection), which also expressed GFP in their tissues.

6.9.1 pSMARt-labelled mESC cells can form chimaeras

S/MAR-labelled mESC were injected into embryos and generated chimeric mice. First, high passage pSMARt-labelled BL6 mESC were injected into CD1 albino blastocysts, which resulted in low percentage chimerism (data not shown). One possible explanation for the low quality of chimaeras is the old passage and poor quality of the injected pSMARt-mESC, as a result of continued culturing to perform the previously described molecular analysis and pluripotency tests. To circumvent that, a second round of chimaeras was generated by injection of low passage pSMARt or nSMARt-labelled 129Ola mESC. This resulted in much better chimaeras, reaching in some cases complete chinchilla coat colour, which corresponds to practically 100% pSMARt-mESC contribution **(Figure 62).** By reducing the mESC passage number and improving the quality of the injected cells, the percentage of chimerism and the overall quality of the chimaeras obtained is significantly improved. Therefore, 129Ola/BL6 chimaeras were used for further molecular analysis and downstream experiments.

6.9.2 Microinjection of S/MAR vectors into 1-cell stage zygotes results in the generation of isotransgenic mice

Current methods to generate transgenic mice involve the use of integrative techniques, such as transposons [256], lentiviral transduction or gene editing techniques [260]. Differently, from these techniques, which might leave a genetic 'scar' on the genome, mice generated with S/MAR vectors should retain the same genomic sequences and information as wild-type animals and could be considered isogenic transgenics (isotransgenic) mice. The only difference between isotransgenics and the unmodified mice would then be the overexpression of supplementary genetic information without damage to the genomic DNA. The first attempt to generate transgenic mice using S/MAR vectors was made by pEPI delivery into pig embryos via sperm-mediated gene transfer [261], in which transgene

expression was achieved in foetuses, but they were not brought to term, and the germ-line transmission could not be addressed.

To investigate the potential of S/MAR vectors as a tool to generate isotransgenic mice, pSMARt vectors were microinjected directly into the pronucleus of 1-cell stage zygotes, which also allowed to gain a better understanding of the vectors' expression and establishment in a 'selection-free' environment.

For that, between 130 and 780 molecules of the pSMARt vector were injected into 1-cell stage zygotes of BL6 mice. 250 microinjected embryos were then transferred into 10 foster mothers and brought to term. 34 pups were born, and 3 died shortly after birth, which corresponded to approximately 13% survival rate. Taking into account that the survival rate of born embryos after DNA microinjection ranges between 10-25%, as compared to approximately 80% without microinjection [327], a survival rate of 13% should be considered normal. Although 95% of the embryos look viable after injection and before embryo transfer, the decrease in viability might be due to mechanical damage of the embryos with a glass capillary during the microinjection procedure itself *(source: Transgenic service, DKFZ).*

After microinjection, 12 embryos were kept in culture to check for toxicity of the vector and correct embryonic development **(Figure 63).** Considering that culturing of embryos *in vitro* is an artificial situation that generates differences in development between individual embryos, and that the DNA microinjection itself reduces the viability from 80% to 10-25%, a ratio of 8/12 embryos reaching the blastocyst stage *in vitro* is an indicative that the transgene (GFP) has no detrimental effect on the embryos. By monitoring the microinjected embryos, it can be observed that pSMARt is expressed throughout the first stages of embryonic development until hatching occurs.

Although all surviving embryos express GFP, even at the blastocyst stage, conclusions about the establishment of the vector in a 'selection-free' environment cannot be drawn at this point. The following assumptions should be considered:

1. **S/MAR vector dilution.** The vectors segregate during mitosis but do no establish (replicate), which results in vector dilution. Considering that between 1-2 picoliters of pSMARt (1-3ng/ul), corresponding to 130-780 copies of plasmid, were injected to 1-cell zygote, and that a late blastocyst has roughly 128 cells, which corresponds to 7 cell divisions; and assuming that in every cell division the vector is segregated in equal numbers to the daughter cells, that results in 1 to 5 copies of pSMARt vector in each cell of the blastocyst **(Table 49)**. In other words, the

GFP expression observed might be due to 'excess' of vector copies per cell, rather than due to the vector being established and replicating episomally. Assuming no vector establishment, the vector would start being diluted from the 8th cellular division, which would result in an adult mouse with 130-780 molecules of pSMARt (which corresponds to the initial number of molecules microinjected) distributed along its body. This practically equals no detectable GFP expression, as the chances of finding positive cells in the adult mouse are too low.

Table 49: Dilution of pSMARt vector copies in the first stages of embryonic development.

Cell division	Cell stage	Copies of vector
-	Zygote (1-cell)	130 – 780
1	2 cell	65 – 390
2	4 cell	32,5 – 195
3	8 cell	16,25 – 97,5
4	16 cell	8,12 – 48,75
5	Blastocyst 32 cell	4,06 – 23,38
6	Blastocyst 64 cell	2,03 – 11,69
7	Blastocyst 128 cell	1,02 – 5,85

2. **S/MAR vector establishment.** Assuming that the pSMARt vector is equally segregated and established (autonomously replicating) in a selection-free environment during the embryonic development, a complete and homogeneous transgenic (GFP expressing) mice is expected.

3. **S/MAR vector silencing.** One might also argue that S/MAR vectors can become silenced in the early phases of development and that the expression of GFP detected at the blastocyst stage is a residual expression from a vector that was originally 'active' in the zygote. However, considering that the turboGFP (a variant of coGFP) half-life is approximately 2h [328, 329], the transgene expression observed at the blastocyst stage (4-6 days after injection), has to be the consequence of the vector being active and transcribing.

However, no conclusions about vector establishment under selection-free conditions could be made until the pups were born and the GFP expression was confirmed. Later, we could confirm that out 31 born transgenic pups, only one expressed GFP in different tissues, therefore proving that, albeit low frequency (3,22%), S/MAR vectors can segregate and establish *in vivo* in the absence of selection.

6.9.3 S/MAR vectors are expressed in transgenic tissues

The expression of S/MAR vectors was evaluated in born transgenic and chimeric mice using fluorescence microscopy, flow cytometry, and PCR amplification of GFP.

The results show that S/MAR vectors can be detected in tissues from all germ layers; such as heart, kidney, liver, muscle, skin, blood, bone marrow and spleen. However, a tendency towards skeletal muscle is observed, which might be explained by the abundant actin presence in muscle and the presence of the first intron of the chicken ß-actin gene in the vector CAG promoter. On the contrary, expression in the kidney is rarely found. This differences in expression across tissues might have several explanations:

1) **In the case of microinjected pSMARt:** The vector does not segregate or establish equally in the embryonic cells early in the development, and therefore, there is some degree of mosaicism in the adult mice. This mosaicism can potentially explain heterogenous GFP expression and difficult detection via PCR.

2) **In the case of Chimeras:** pSMARt-mESC can have a preference towards a specific cell lineage *in vivo*. We observed a tendency of mESC towards ectodermal derivatives during *in vitro* random differentiation experiments, while mesodermal structures were less frequent and endodermal derivatives were hard to obtain.

3) S/MAR vectors get silenced in specific tissues (*e.g.*: kidney), due to the cell microenvironment or cell-specific epigenetic mechanisms.

6.10 S/MAR DNA vectors are transmitted to the gametes but fail to generate transgenic offspring

Little is known about the ability of S/MAR-labelled stem cells to contribute to the germ-line (stem cell transmission) or even more interestingly, about the inheritance of episomal DNA vectors (vector transmission). Was addressed these questions by breeding 129Ola/BL6 chimaeras or transgenic mice with C57BL/6 mice and investigating the presence and expression of S/MAR vectors into the offspring.

6.10.1 pSMARt and nSMARt-labelled mESC contribute to the germ-line

Chimeric males generated with both nSMARt (c39) and pSMARt (c44) resulted in 100% agouti litters **(Figure 70),** which indicates that S/MAR-labelled mESC contributed to the germ-line and the next generation. This is another piece of evidence that pSMARt and nSMARt stem cells retain their pluripotent capabilities and that their potential is not limited by the type of vector they contain.

6.10.2 pSMARt and nSMARt are not transmitted to the offspring

Next, we addressed whether S/MAR vectors, which could be detected in F0 chimeric and pronuclear injected mice, were also present and expressed in the F1 litters. However, flow cytometry analysis failed to detect relevant GFP expression **(Figure 71).** The next logical step was to investigate the presence of S/MAR vector in the offspring. Surprisingly, the S/MAR vectors could not be detected via PCR in any of the chimeric litters **(Figure 72),** suggesting that the vector is lost in the F1 generation. Intriguingly, when the same experiments were performed in the litters from pronuclear-injected mice, pSMARt could be amplified in two pups (T82 and 83) from mice T11 **(Figure 74).** Different possibilities explaining the one-off presence of S/MAR vectors in F1 generations are discussed **below.**

Overall, the absence of vectors in the F1 generation points at meiosis (spermatogenesis) as a significant hurdle for the transmission of S/MAR vectors into the offspring.

6.10.3 Spermatogenesis

Spermatogenesis is the process by which haploid sperm cells are obtained from diploid germ cells, located in the basal membrane of the seminiferous tubules. Spermatogonial stem cells (diploid) undergo mitosis to generate primary spermatocytes. Primary Spermatocytes undergo one round of cellular replication and two subsequent rounds of cellular division. Each primary spermatocyte first

divides meiotically into two secondary spermatocytes, which in turn divide again to generate two haploid spermatids each. The biological processes in which gametes are produced, cannot be fully understood without taking into account the epigenetic mechanisms involved. These are mainly relevant during spermatogenesis, in which DNA is highly methylated, the majority of histones are substituted by protamines and the remaining ones are highly modified [330].

Considering that high DNA methylation results in a substantial reduction of transcription and that S/MAR vectors need transcription running through the S/MAR to be kept episomal [163], it is reasonable to think that S/MAR vectors can be strongly influenced by epigenetic changes happening during spermatogenesis, and that these changes can explain the lack of transgene expression in the offspring. Therefore, we had a closer look into spermatogenesis and evaluated the vector presence and expression before (testes) and after (sperm) meiosis.

As observed in **(Figure 75a)** fluorescent microscopy shows expression of the transgene in testes of transgenic (T11) and chimeric (C39, C44) mice. To exclude that the fluorescence is emitted by the external testicular membrane or *tunica albuginea*, and that fluorescence is coming from meiotic cells, the *seminiferous tubules* were homogenised to isolate the germinal cells, comprised by *spermatogonias*, *spermatocytes* and *spermatids.* The same GFP expression pattern was observed in the cellular compartment via Flow cytometry **(Figure 75b),** suggesting that the vectors are present and expressed in germinal cells. Finally, the end stage of meiosis, which corresponds to mature sperm, was analysed. However, no GFP could be detected in sperm from T11, C39 or C44 via fluorescent microscopy; except for the sperm from the positive control mice **(Figure 75c).** Taking into account that the positive control mice contain UBC::GFP integrated and that S/MAR vectors need active transcription to remain episomal, there are two possible explanations accounting for the lack of episomal GFP detection in sperm:

1) Sperm cells are characterised by highly packaged and compact DNA and a reduction of transcription. Considering that the expression levels of GFP in UBC:GFP testes and germ cells are much higher than those from T11, C39 and C44, this could result remaining GFP in the cytoplasm of positive control sperm, a phenomenon that could not have been observed in the cytoplasm of sperm with lower levels of GFP expression.

2) Episome is inactive and lost during meiosis, while integrated GFP is not. However, PCR amplification revealed that GFP is present in both testes and sperm **(Figure 76)** suggesting that

the vector is not lost but silenced in sperm cells, probably due to high DNA compaction and reduction of transcription.

We then investigated at which stage of spermatogenesis are the vectors getting silenced. Immunohistochemistry stainings of GFP from transgenic and chimeric testes reveal that GFP can be found in spermatogonias but not in intermediate or end-stage meiotic cells, suggesting that silencing occurs at the *spermatogonia-primary spermatid* transition **(Figure 77).**

6.10.4 Fertilisation

Considering that S/MAR vectors are present, although not expressed, in sperm cells; we propose several options that can explain the lack of S/MAR detection in the F1 generation; taking part before, during or shortly after fertilisation **(Figure 86).** We also provide evidence suggesting that the vectors are not integrated in the S/MAR-modified F0 generation, although they can become integrated after fertilisation.

1. **S/MAR vectors are integrated in the F0 generation**. Integration in transgenic and chimeric F0 could result in GFP expression in the sperm cells and offspring, similarly to the integrated UBC::GFP positive control mice. Since sperm cells from S/MAR transgenic animals do not express GFP and the F1 generation does not contain the vectors, it seems unlikely that the vectors are integrated.

2. **S/MAR vectors become integrated during spermatogenesis**. Epigenetic events happening during spermatogenesis, such as DNA methylation or histone modifications, could result in a decrease of transcription, which could lead to the integration of S/MAR vectors. Also, during the first meiotic division, homologous recombination between non-sister chromatids occurs. The homology between plasmid and endogenous S/MAR motifs could lead to vector rearrangements or integration. This might explain the two pSMARt-GFP detections in F1 mice from T11 **(Figure 74).**

3. **S/MAR becomes silenced during meiosis.** The same epigenetic events could lead to silencing of S/MAR vectors during meiosis, as observed by the loss of GFP expression in the germinal epithelia. Although not expressed, the vector remains episomal in the sperm cell.

4. **Differential segregation vs equal segregation of S/MAR vectors.** Between 1 to 2 copies of S/MAR vectors are established per cell. Meiosis involves one round of replication and two

rounds of cell division. Assuming equal segregation of the vector, all sperm cells could get at least one copy of the vector, while a differential segregation scenario could result in ¼ or ½ of sperm cells having the vector.

5. **Detrimental effect of S/MAR-sperm.** In the case of a mixed pSMARt sperm population (differential segregation setting), the presence of pSMARt vectors would increase the sperm cell weight by 0,007 femtograms (7,55 $x10^{-18}$ g, corresponding to one pSMARt molecule of 7Kb), which could result into slowing sperm swimming and in a fertilisation disadvantage.

6. **Immunity of female reproductive tract towards S/MAR- sperm.** Expression of GFP antigens in the sperm could be recognised in the female reproductive tract, resulting in only the GFP-negative sperm being able to fertilise. However, there are studies showing that seminal plasma suppresses immune responses in the female reproductive tract, making it tolerant to sperm [331].

7. **Immunity of zygote/embryo against pSMARt-sperm**. Recognition and silencing of the vector at the embryonic stage seems unlikely since pSMARt vectors were microinjected into 1-cell stage zygotes to generate transgenics, were expressed during the first stages of development and contributed to the generation of pSMARt-expressing isotransgenic mice.

8. **Vector re-establishment.** The establishment efficiency of pSMARt *in vitro* is approximately 5% and 3,22% *in vivo*, after DNA microinjection. Assuming that sperm cells deliver between 1-2 copies of pSMARt into the zygote and that the vectors have to re-establish in the new embryo, the chances of 1-2 copies of DNA being established after fertilisation are too low, resulting in the vector being diluted during cell divisions.

9. **DNA repair mechanisms after fertilisation leading to vector integration.** DNA repair mechanisms are absent in late phases of spermatogenesis [332] but become very active after fertilisation [333]. The DNA repair mechanisms, involving in some cases homologous recombination, could lead to rearrangements of the vector and integrations. This might explain the few pSMARt-GFP detections in F1 mice from T11 **(Figure 74).**

6.10.4.1 Working hypothesis

The absence of pSMARt vectors in the F1 generation is likely to be a consequence of a low-efficiency vector re-establishment after fertilisation. The sperm cells, which contain inactive episomal DNA, deliver 1 - 2 copies of S/MAR vectors into the egg. Considering that the *in vitro* and *in vivo* establishment efficiencies are around 5% and 3% respectively and that vector establishment is a stochastic event, the chances of vector re-establishment after fertilisation are very low. Therefore, the few episomal vector copies would dilute as the embryo develops, which would result in loss of vector transmission in the F1 generation. Despite the establishment chances being low, it is still possible to detect S/MAR vectors in the F1 generation, as observed by PCR amplification of GFP **(Figure 74).** However, the molecular state of this detected vectors remains unclear.

Although with the present data, 'the re-establishment hypothesis' seems the most plausible, further experiments are needed to validate or discard it. For example, a combination of *in vitro* fertilisation of pSMARt-sperm cells with subsequent FISH analysis on fertilised embryos would provide some hints about the number of vector copies 'delivered' to the egg. Tracking those embryos with fluorescent microscopy would shed some light on the establishment and expression of S/MAR vectors after fertilisation. Also, investigation of S/MAR vector behaviour in female gametogenesis (oogenesis) could determine possible differences between meiotic processes. Chimeric females were bred but stem cell transmission was not found, and the vector transmission could not be addressed.

We could not discard the infrequent but possible S/MAR integrations before, during or after meiosis, which could also result in detection of S/MAR vectors in the F1 pups. Likewise, we did not have enough evidence supporting equal vector segregation during meiosis, which could be investigated via Fluorescent *In Situ* Hybridisation (FISH) of S/MAR vectors in the germinal epithelia, and in sperm cells in particular. This would allow to track the vector segregation during meiosis and to detect how many sperm cells receive DNA vectors.

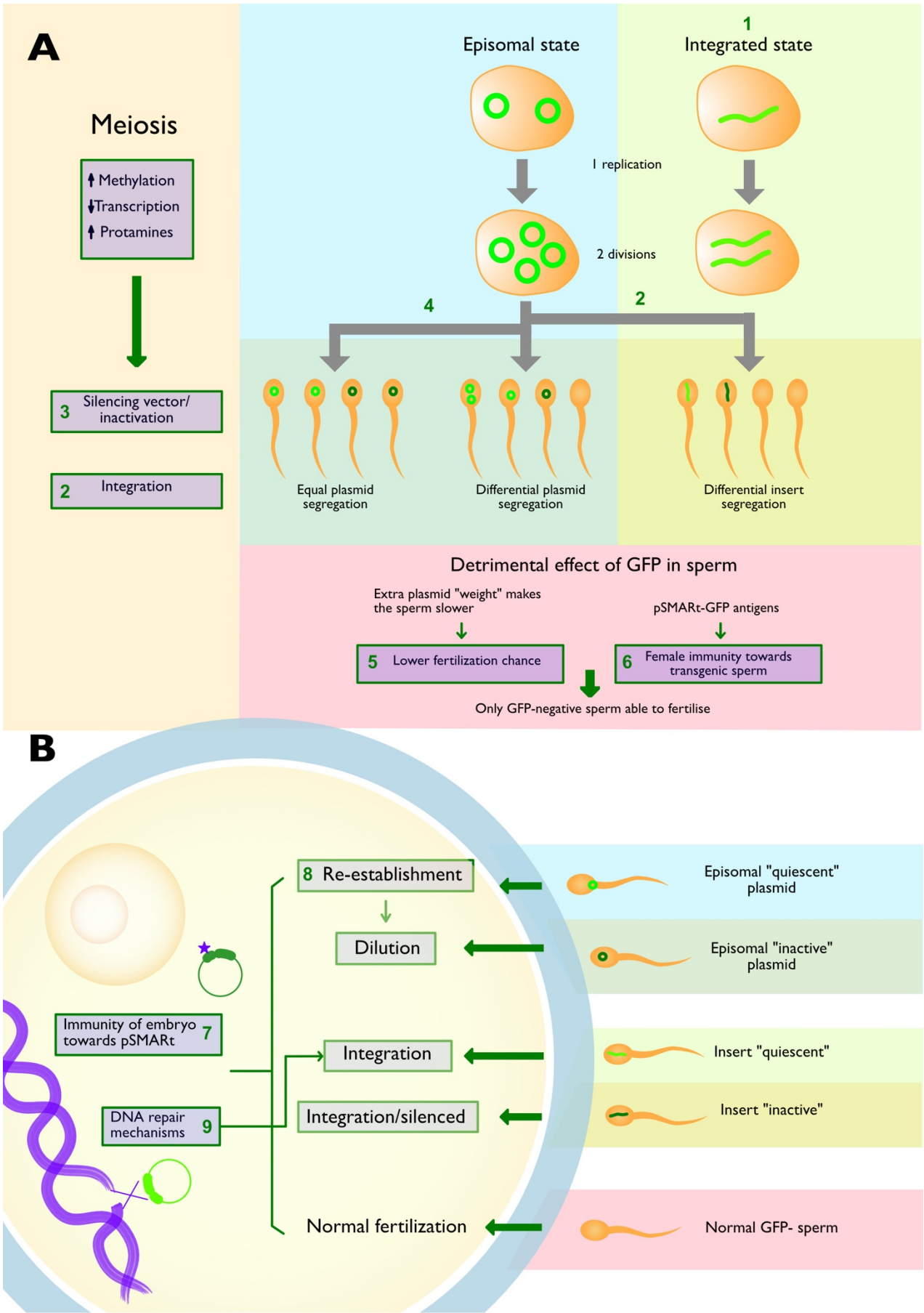


Figure 86: S/MAR vector fate during meiosis and fertilisation

7. CONCLUDING REMARKS

This thesis shows the development and improvement of S/MAR DNA vectors from the originally described pEPI vector, which we demonstrated to be incapable of sustaining transgene expression in pluripotent cells; through a range of vector generations. The new vectors have improved all aspects of their functionality and are capable of persistently modifying dividing and differentiating cells for an unlimited period while providing sustained and high levels of transgene expression in their differentiated progeny, without causing molecular damage to the cells they modify.

This work shows for the first time that S/MAR DNA vectors can:

- Persistently express high levels of the transgene in rapidly proliferating cells, including but not limited to neuroblastoma cells, murine embryonic stem cells (mESC), murine induced pluripotent stem cells (miPSC) and human induced pluripotent stem cells (hiPSC).
- Retain transgene expression during ATRA-mediated differentiation and de-differentiation in neuroblastoma cells.
- Generate minimal impact on modified neuroblastoma cells, as they retain the expression of neuronal markers and are able to engraft when injected into immunodeficient mice (NB), generating transgene-expressing tumours.
- Establish at low copy numbers, ranging from 2 - 4 copies/cell in neuroblastoma cells and 1 - 2 copies/cell in mESC.
- Remain extrachromosomal as an episomal replicon in both neuroblastoma and mESC.
- Preserve the pluripotent capabilities of modified mESC, miPSC and hiPSC.
- Retain transgene expression during *in vitro* differentiation, from mESC and hiPSC into representatives of the three germ layers.
- Retain high levels of transgene expression during *in vitro* hematopoietic differentiation, from mESC to CD34+ hematopoietic precursors.

- Survive *in vivo* differentiation and generating transgenic animals either by pronuclear injection of DNA vectors or injection of genetically modified mESC into embryos.
- Persistently express high levels of the transgene in transgenic organs and tissues.
- Be transmitted to the gametes but fail to generate transgenic offspring.
- Serve as reprogramming tools for the derivation of murine and human primary cells into miPSC and hiPSC.
- Retain transgene expression during cellular reprogramming, in both miPSC and hiPSC.
- Immortalise primary cells, such as murine lung primary fibroblasts.

8. FUTURE DIRECTIONS

The development and improvement of the S/MAR vectors' functionality result in the persistent and robust modification of dividing and differentiating cells, which broadens the horizon of future applications of this vectors, including their application for gene and cell therapy, the generation of transgenics and the immortalisation of cells.

8.1 Gene and cell therapy

An ideal application S/MAR vectors would be the generation of disease-corrected and transgene-free iPSC for autologous transplantation with the aim to treat monogenic diseases. For that purpose, a prototype of a multifunctional S/MAR vector, named pAmpel, was designed and tested. This vector is composed of 1) constitutively expressed labelling cassette, which drives the expression of a reporter gene, the selection marker and the S/MAR, maintaining the episomal status of the vector; a 2) reprogramming cassette containing the reprogramming factors (OKSM for mice or OKSML for human) and a red reporter gene (dTom). This cassette can be removed upon transient addition of Cre recombinase, either in a vector or delivered as mRNA; and 3) a therapeutic cassette containing the gene of interest and the reporter gene luciferase. The combination of three functional units would allow the delivery of one unique multifunctional DNA molecule to the cells, avoiding the low efficiency of corrected-hiPSC obtained when four EBNA vectors are co-transfected with pSMARt-GFP. However, the large size of this multifunctional vector compromises the efficiency of DNA delivery.

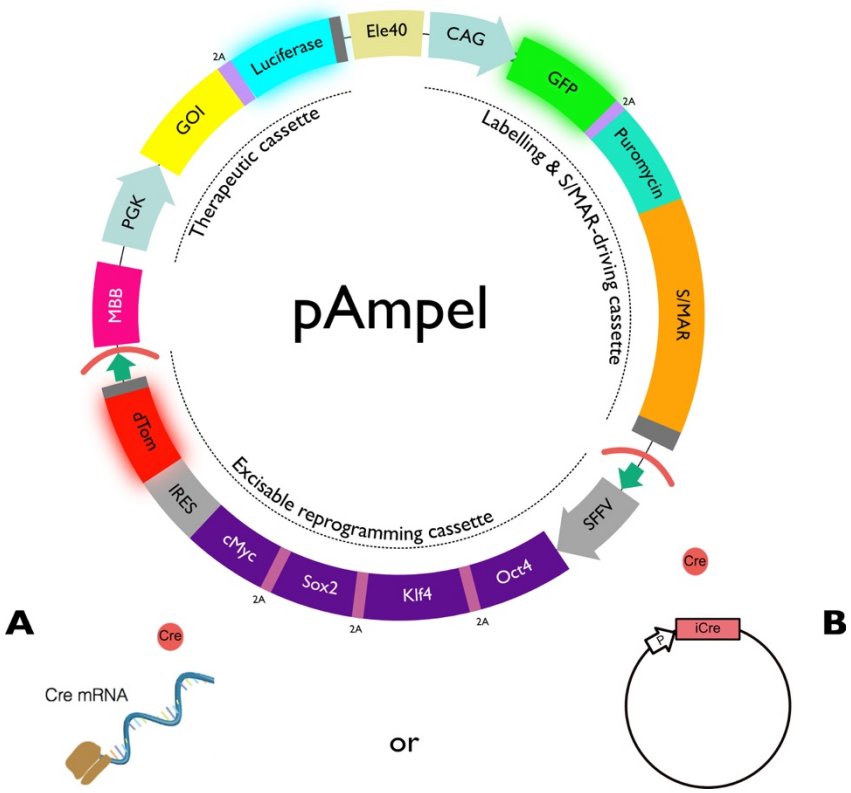


Figure 87: Cartoon representing the multifunctional vector pAmpel

To generate pAmpel, the different units were first cloned and tested separately. pSMARt-labelling vectors **(Figure 33),** reprogramming vectors **(Figure 35),** and therapeutic vectors **(Figure 39** and **Figure 40)** are shown in **Section 5.1.8.** Then, both labelling and reprogramming units were combined into one vector (pBeast vector series) and tested in MEFs. pBeast constitutively expressed S/MAR driving cassette containing the reporter gene coGFP with the selective marker Puromycin, and a reprogramming cassette containing the OKSM factors and the reporter gene dTom. Two LoxP sites are flanking the reprogramming cassette, which can be excised upon addition of Cre recombinase; leaving behind the functional labelling GFP-2A-Puro_S/MAR cassette. Other versions of pBeast with a combination of other promoters and chromosomal elements were cloned but not tested. Although pBeast demonstrated that the co-existence of two expression cassettes in one S/MAR vector is possible, the pBeast vector series was left aside in this study due to its large size and the difficulty of delivering large DNA into MEFs [334]. Therefore, the idea of using a combined pAmpel multifunctional vector was dropped and refocused on exploiting the individual parts of the vector separately.

However, the improved vector spliced design in combination with minimally sized nanovectors, together with an easier transfection efficiency in human fibroblasts, opens up the possibility to revisit the multifunctional pAmpel vector idea.

Therapeutic vectors coding for FancA **(Figure 39)** and Rep1 **(Figure 40)** were cloned, tested, and are ready to be cloned into pAmpel or used in combination with S/MAR reprogramming vectors to reprogram and genetically correct patient-derived cells.

8.1.1 Simultaneous correction and reprogramming of Fanconi Anaemia patient-derived cells

Fanconi Anaemia is an ideal candidate to be treated with gene therapy because the supplementation of the correct gene is enough to restore the functionality of the pathways as well as to rescue the phenotype. Several gene therapy attempts have been performed using retrovirally-mediated gene transfer of autologous HSC, although success has not yet been achieved. One reason is that HSCs are hard to mobilise from the bone marrow and hard to transduce with lentivirus. A possible solution to overcome these problems is the derivation and correction of patients' cells.

Müller *et al.* showed that Correction of FancA is necessary to overcome reprogramming resistance of Fanca-/- cells [275]. We demonstrated that our improved S/MAR vectors are capable of reprogramming patient-derived fibroblasts as well as to persistently label mESC and sustain high levels of transgene expression through hematopoietic differentiation **(Figure 61).** Therefore, pSMARt-FancA vectors could be inserted into pAmpel to transfect patient fibroblasts, which could be corrected by supplementation of FancA as well as reprogrammed into hiPSC. Corrected hiPSC could then be differentiated into HSC and transplanted back into the patient.

8.1.2 Genetic correction, derivation and autologous hiPSC transplantation for the correction of degenerative eye diseases

Retinal degenerative diseases are a group of debilitating conditions that affect the photoreceptor cell layer of the retina, also called retinal pigmented epithelia (RPE). The most common degenerative disease is threatening the vision in older populations of developed countries is **age-related macular degeneration (AMD)**. AMD is characterised by the development neovasculature in the subretinal space of the centre of the retina (macula), which disrupts and damages the RPE. Current treatments involve the intraocular injection of anti-vascular endothelial growth factor (VEGF) drugs, which ameliorate but do not solve the underlying issue; or surgical removal of the neovasculature. Regenerative therapies were first based on allotransplantation of RPE derived from human embryonic stem cells, which resulted in graft rejection. Alternatively, autologous transplantation of patient-derived-hiPSC showed promising results [335]. However, the derivation of hiPSC was done either by using retroviral delivery

of reprogramming factors [336], or EBNA-1 derived vectors [203]. Alternatively, S/MAR reprogramming vectors could be used to derive patient cells and generate hiPSC, which could be then differentiated into RPE and transplanted back to the patients.

Other afflictions of the RPE are inherited forms of blindness, such as **Choroideremia.** Choroideremia (CHM) is an X-linked recessive chorioretinal dystrophy caused by mutations in the Rab Escort Protein 1 (REP1) gene. Mutations that lead to loss of REP1 function disrupt normal intracellular trafficking and post-translational lipid modification of Rab small GTPases (Rab proteins) leading to progressive degeneration of the retinal pigment epithelium, photoreceptors, and choroid.

prenylation

Previous attempts to correct deficient cells were based either on *in situ* delivery of lentiviral [28] or adenoviral [278] particles containing the therapeutic gene Rep1. However, inflammatory responses led to severe side effects in treated patients.

Alternatively, pSMARt-Rep-1 vectors could be used in pAmpel or combination with S/MAR reprogramming vectors to correct the Rep1 genetic defect of patient-derived hiPSC. Directed differentiation and autologous transplantation of the hiPSC-derived RPE would represent an alternative to current gene therapy approaches to cure choroideremia.

8.2 Transgenics

The new generations of S/MAR vectors have been designed and refined to sustain high levels of ubiquitous expression when combined with an appropriate promoter. S/MAR vectors can retain expression of the transgene across different tissues for the lifetime of the mice when they are microinjected into the pronucleus or used for labelling mESC. Although the vector seems to be present in some F1 pups, the episomal state of the vectors in the progeny is not yet clear. Therefore, one missing point and potential direction of this work is the presence assessment of vector episomes after female gametogenesis (oogenesis).

Although vector transmission is not yet completely understood, the S/MAR vector technology could still be used as a safer alternative to the current transgenic technology, which relies on the use of integrative techniques or viral-derived episomal vectors. For example, S/MAR vectors could be used to generate transgenic mouse lines containing a specific GOI. These modified mice could be used as a source of modified cells for *in vitro* applications.

Another application of S/MAR vectors could be the reprogramming of human cells for the generation of xeno-transplantable organs. This idea is was coined by the group of Izpisua-Belmonte, which generated inter-species chimaeras via blastocyst complementation between human iPSC and pig embryos to generate xeno-transplantable human organs [337]. Human fibroblasts were reprogrammed using the aforementioned EBNA vectors described by Okita *et al.* [298], and hiPSC were then injected into pig blastocysts, which resulted in the formation of human-pig chimaeras. With optimised DNA delivery protocols, S/MAR vectors could reprogram (and even correct) either wildtype or patient-derived fibroblast into hiPSC, which could then be implanted into pig embryos to generate (corrected) xeno-transplantable organs.

8.3 Cellular immortalisation

Immortal cells can evade cellular senescence and undergo an indefinite number of cell divisions. Their indefinite propagation has made the culturing of immortal cells a tool for biochemical, cell biology and toxicology analysis. Typically, immortal cells can be isolated from 1) certain types of cancer cells with specific mutations that confer them immortality or 2) cells infected with viral genes that dysregulate cell cycle (*e.g.*: HEK293T cells infected with the adenoviral protein E1 or HeLa cells infected with HPV). Immortal cells can also be generated artificially via expression of proteins required to overcome senescence (*e.g.*: TERT), or that dysregulate the cell cycle (*e.g.*: SV40 large T antigen).

Another potential application of S/MAR vectors is their use as immortalising tools for culturing of primary cells. S/MAR-immortalisation can be achieved by a typical transfection of primary cells with a pSMARt vector containing SV40LT or TERT. For instance, patient-derived cells could be immortalised and cultured for drug screenings *in vitro*, but the indirect manipulation of oncogenes rules out the possibility to use these cells in therapeutic applications.

9. REFERENCES

9.1 Online Resources

- http://biochemicalc.nuim.ie/calculators/
- http://nebiocalculator.neb.com/#!/ligation
- http://cels.uri.edu/gsc/cndna.html
- http://primer3.ut.ee
- https://www.thermofisher.com/de/de/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo-scientific-web-tools/tm-calculator.html
- http://genome.ucsc.edu/cgi-bin/hgPcr
- https://www.thermofisher.com/de/de/home/brands/thermo-scientific/molecular-biology/molecular-biology-learning-center/molecular-biology-resource-library/thermo-scientific-web-tools/multiple-primer-analyzer.html
- http://www.idtdna.com/pages/products/nextgen/target-capture.
- https://www.fda.gov/BiologicsBloodVaccines/CellularGeneTherapyProducts/ApprovedProducts/default.htm
- https://www.technologyreview.com/s/609643/2017-was-the-year-of-gene-therapy-breakthroughs/

9.2 Literature

[1] *Gene Therapy Clinical Trials Worldwide*. 2017.

[2] "Approved Cellular and Gene Therapy Products." pp. 1–3, 2017.

[3] E. Mullin, "2017 was the year of gene-therapy breakthroughs," vol. 376, no. 9. pp. 848–855, 03-Jan-2018.

[4] A. W. Nienhuis, C. E. Dunbar, and B. P. Sorrentino, "Genotoxicity of retroviral integration in hematopoietic cells.," *Mol Ther*, vol. 13, no. 6, pp. 1031–1049, Jun. 2006.

[5] S. Jenks, "Gene therapy death--"everyone has to share in the guilt"..," *J. Natl. Cancer Inst.*, vol. 92, no. 2, pp. 98–100, Jan. 2000.

[6] S. Hacein-Bey-Abina, C. von Kalle, M. Schmidt, F. Le Deist, N. Wulffraat, E. McIntyre, I. Radford, J.-L. Villeval, C. C. Fraser, M. Cavazzana-Calvo, and A. Fischer, "A serious adverse event after successful gene therapy for X-linked severe combined immunodeficiency.," *N. Engl. J. Med.*, vol. 348, no. 3, pp. 255–256, Jan. 2003.

[7] M. Schleef, *Minicircle and Miniplasmid DNA Vectors*. John Wiley & Sons, 2013.

[8] L. Huang, D. Liu, and E. Wagner, Eds., *Advances in Genetics, Volume 89*, vol. 89. Elsevier Inc., 2015, pp. 1–297.

[9] R. I. Mahato, K. Kawabata, T. Nomura, Y. Takakura, and M. Hashida, "Physicochemical and pharmacokinetic characteristics of plasmid DNA/cationic liposome complexes.," *J Pharm Sci*, vol. 84, no. 11, pp. 1267–1271, Nov. 1995.

[10] D. V. Goeddel, D. G. Kleid, F. Bolivar, H. L. Heyneker, D. G. Yansura, R. Crea, T. Hirose, A. Kraszewski, K. Itakura, and A. D. Riggs, "Expression in Escherichia coli of chemically synthesized genes for human insulin.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 76, no. 1, pp. 106–110, Jan. 1979.

[11] A. C. A. Roque, C. R. Lowe, and M. A. Taipa, "Antibodies and genetically engineered related molecules: production and purification.," *Biotechnol. Prog.*, vol. 20, no. 3, pp. 639–654, May 2004.

[12] C. Han and B. S. Kwon, "Chimeric antigen receptor T-cell therapy for cancer: a basic research-oriented perspective.," *Immunotherapy*, vol. 10, no. 3, pp. 221–234, Mar. 2018.

[13] R. D. Palmiter, H. Y. Chen, and R. L. Brinster, "Differential regulation of metallothionein-thymidine kinase fusion genes in transgenic mice and their offspring.," *Cell*, vol. 29, no. 2, pp. 701–710, Jun. 1982.

[14] A. Fischer, S. Hacein-Bey-Abina, and M. Cavazzana-Calvo, "20 years of gene therapy for SCID.," *Nat. Immunol.*, vol. 11, no. 6, pp. 457–460, Jun. 2010.

[15] A. Black, V. Vasireddy, D. C. Chung, A. M. Maguire, R. Gaddameedi, T. Tolmachova, M. Seabra, and J. Bennett, "Adeno-associated virus 8-mediated gene therapy for choroideremia: preclinical studies in in vitro and in vivo models.," *The Journal of Gene Medicine*, vol. 16, no. 5, pp. 122–130, May 2014.

[16] N. Cereso, M. O. Pequignot, L. Robert, F. Becker, V. De Luca, N. Nabholz, V. Rigau, J. De Vos, C. P. Hamel, and V. Kalatzis, "Proof of concept for AAV2/5-mediated gene therapy in iPSC-derived retinal pigment epithelium of a choroideremia patient," *Molecular Therapy - Methods & Clinical Development*, vol. 1, pp. 14011–13, 2014.

[17] A. Geiselhart, A. Lier, D. Walter, and M. D. Milsom, "Disrupted Signaling through the Fanconi Anemia Pathway Leads to Dysfunctional Hematopoietic Stem Cell Biology: Underlying Mechanisms and Potential Therapeutic Strategies," *Anemia*, vol. 2012, no. 1, pp. 1–18, 2012.

[18] M. D. Milsom and D. A. Williams, "Gaining the hard yard: pre-clinical evaluation of lentiviral-mediated gene therapy for the treatment of β-thalassemia," *EMBO Mol Med*, vol. 2, no. 8, pp. 291–293, Jul. 2010.

[19] B. Alberts, *Molecular Biology of the Cell*. Garland Science, 2008.

[20] G. L. Rosano and E. A. Ceccarelli, "Recombinant protein expression in Escherichia coli: advances and challenges.," *Front Microbiol*, vol. 5, no. 631607, p. 172, 2014.

[21] J. Ihssen, M. Kowarik, S. Dilettoso, C. Tanner, M. Wacker, and L. Thöny-Meyer, "Production of glycoprotein vaccines in Escherichia coli.," *Microb. Cell Fact.*, vol. 9, no. 1, p. 61, Aug. 2010.

[22] V. Koppolu and V. K. Vasigala, "Role of Escherichia coli in Biofuel Production.," *Microbiol Insights*, vol. 9, pp. 29–35, 2016.

[23] Z. K. Punja, "Genetic engineering of plants to enhance resistance to fungal pathogens—a review of progress and future prospects," *Canadian Journal of Plant Pathology*, vol. 23, no. 3, pp. 216–235, Sep. 2001.

[24] P. Krimpenfort, A. Rademakers, W. Eyestone, A. van der Schans, S. van den Broek, P. Kooiman, E. Kootwijk, G. Platenburg, F. Pieper, and R. Strijker, "Generation of transgenic dairy cattle using 'in vitro' embryo production.," *Biotechnology (N.Y.)*, vol. 9, no. 9, pp. 844–847, Sep. 1991.

[25] L. Lai, J. X. Kang, R. Li, J. Wang, W. T. Witt, H. Y. Yong, Y. Hao, D. M. Wax, C. N. Murphy, A. Rieke, M. Samuel, M. L. Linville, S. W. Korte, R. W. Evans, T. E. Starzl, R. S. Prather, and Y. Dai, "Generation of cloned transgenic pigs rich in omega-3 fatty acids.," *Nat Biotechnol*, vol. 24, no. 4, pp. 435–436, Apr. 2006.

[26] "Disease-corrected haematopoietic progenitors from Fanconi anaemia induced pluripotent stem cells," vol. 460, no. 7251, pp. 53–59, Feb. 2009.

[27] K. A. Gush, K. L. Fu, M. Grompe, and C. E. Walsh, "Phenotypic correction of Fanconi anemia group C knockout mice.," *Blood*, vol. 95, no. 2, pp. 700–704, Jan. 2000.

[28] T. Tolmachova, O. E. Tolmachov, S. T. Wavre Shapton, D. Tracey White, C. E. Futter, and M. C. Seabra, "CHM/REP1 cDNA delivery by lentiviral vectors provides functional expression of the transgene in the retinal pigment epithelium of choroideremia mice," *The Journal of Gene Medicine*, vol. 14, no. 3, pp. 158–168, Mar. 2012.

[29] T. Tolmachova, O. E. Tolmachov, A. R. Barnard, S. R. de Silva, D. M. Lipinski, N. J. Walker, R. E. MacLaren, and M. C. Seabra, "Functional expression of Rab escort protein 1 following AAV2-mediated gene delivery in the retina of choroideremia mice and human cells ex vivo," *J. Mol. Med.*, vol. 91, no. 7, pp. 825–837, 2013.

[30] F. D. Urnov, J. C. Miller, Y.-L. Lee, C. M. Beausejour, J. M. Rock, S. Augustus, A. C. Jamieson, M. H. Porteus, P. D. Gregory, and M. C. Holmes, "Highly efficient endogenous human gene correction using designed zinc-finger nucleases.," *Nature*, vol. 435, no. 7042, pp. 646–651, Jun. 2005.

[31] M. Morange, "What history tells us XIX. The notion of the episome," *J Biosci*, vol. 34, no. 6, pp. 845–848, 2009.

[32] J. Boch, H. Scholze, S. Schornack, A. Landgraf, S. Hahn, S. Kay, T. Lahaye, A. Nickstadt, and U. Bonas, "Breaking the code of DNA binding specificity of TAL-type III effectors.," *Science*, vol. 326, no. 5959, pp. 1509–1512, Dec. 2009.

[33] M. Jinek, K. Chylinski, I. Fonfara, M. Hauer, J. A. Doudna, and E. Charpentier, "A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.," *Science*, vol. 337, no. 6096, pp. 816–821, Aug. 2012.

[34] A. Fire, S. Xu, M. K. Montgomery, S. A. Kostas, S. E. Driver, and C. C. Mello, "Potent and specific genetic interference by double-stranded RNA in Caenorhabditis elegans.," *Nature*, vol. 391, no. 6669, pp. 806–811, Feb. 1998.

[35] M. A. Matzke and J. A. Birchler, "RNAi-mediated pathways in the nucleus.," *Nat. Rev. Genet.*, vol. 6, no. 1, pp. 24–35, Jan. 2005.

[36] M. Wassenegger, "The role of the RNAi machinery in heterochromatin formation.," *Cell*, vol. 122, no. 1, pp. 13–16, Jul. 2005.

[37] P. D. Zamore, T. Tuschl, P. A. Sharp, and D. P. Bartel, "RNAi: double-stranded RNA directs the ATP-dependent cleavage of mRNA at 21 to 23 nucleotide intervals.," *Cell*, vol. 101, no. 1, pp. 25–33, Mar. 2000.

[38] D. P. Bartel and C.-Z. Chen, "Micromanagers of gene expression: the potentially widespread influence of metazoan microRNAs.," *Nat. Rev. Genet.*, vol. 5, no. 5, pp. 396–400, May 2004.

[39] J. Liu, M. A. Valencia-Sanchez, G. J. Hannon, and R. Parker, "MicroRNA-dependent localization of targeted mRNAs to mammalian P-bodies.," *Nat. Cell Biol.*, vol. 7, no. 7, pp. 719–723, Jul. 2005.

[40] V. Bitko, A. Musiyenko, O. Shulyayeva, and S. Barik, "Inhibition of respiratory viruses by nasally administered siRNA.," *Nat. Med.*, vol. 11, no. 1, pp. 50–55, Jan. 2005.

[41] C. Raoul, S. D. Barker, and P. Aebischer, "Viral-based modelling and correction of neurodegenerative diseases by RNA interference.," *Gene Ther.*, vol. 13, no. 6, pp. 487–495, Mar. 2006.

[42] S. I. Pai, Y.-Y. Lin, B. Macaes, A. Meneshian, C.-F. Hung, and T.-C. Wu, "Prospects of RNA interference therapy for cancer.," *Gene Ther.*, vol. 13, no. 6, pp. 464–477, Mar. 2006.

[43] T. J. McFarland, Y. Zhang, B. Appukuttan, and J. T. Stout, "Gene therapy for proliferative ocular diseases.," *Expert Opin. Biol. Ther.*, vol. 4, no. 7, pp. 1053–1058, Jul. 2004.

[44] K. V. Morris and J. J. Rossi, "Lentiviral-mediated delivery of siRNAs for antiviral therapy.," *Gene Ther.*, vol. 13, no. 6, pp. 553–558, Mar. 2006.

[45] D. Grimm and M. A. Kay, "Therapeutic short hairpin RNA expression in the liver: viral targets and vectors.," *Gene Ther.*, vol. 13, no. 6, pp. 563–575, Mar. 2006.

[46] D. H. Kim and J. J. Rossi, "Strategies for silencing human disease using RNA interference.," *Nat. Rev. Genet.*, vol. 8, no. 3, pp. 173–184, Mar. 2007.

[47] L. Zhang, L. Li, G. A. Hoffmann, and R. M. Hoffman, "Depth-targeted efficient gene delivery and expression in the skin by pulsed electric fields: an approach to gene therapy of skin aging and other diseases.," *Biochemical and Biophysical Research Communications*, vol. 220, no. 3, pp. 633–636, Mar. 1996.

[48] H. Fischer, J. Scherz, S. Szabo, M. Mildner, C. Benarafa, A. Torriglia, E. Tschachler, and L. Eckhart, "DNase 2 is the main DNA-degrading enzyme of the stratum corneum.," *PLoS ONE*, vol. 6, no. 3, p. e17581, Mar. 2011.

[49] K. Kawabata, Y. Takakura, and M. Hashida, "The fate of plasmid DNA after intravenous injection in mice: involvement of scavenger receptors in its hepatic uptake.," *Pharm. Res.*, vol. 12, no. 6, pp. 825–830, Jun. 1995.

[50] R. J. Mumper, J. G. Duguid, K. Anwer, M. K. Barron, H. Nitta, and A. P. Rolland, "Polyvinyl derivatives as novel interactive polymers for controlled gene delivery to muscle.," *Pharm. Res.*, vol. 13, no. 5, pp. 701–709, May 1996.

[51] K. Miyata, N. Nishiyama, and K. Kataoka, "Rational design of smart supramolecular assemblies for gene delivery: chemical challenges in the creation of artificial viruses.," *Chem Soc Rev*, vol. 41, no. 7, pp. 2562–2574, Apr. 2012.

[52] C. M and H. O, "Cancer Gene Therapy – Key Biological Concepts in the Design of Multifunctional Non-Viral Delivery Systems," in *Gene Therapy - Tools and Potential Applications*, no. 9, InTech, 2013, pp. 1–37.

[53] I. A. Khalil, K. Kogure, H. Akita, and H. Harashima, "Uptake pathways and subsequent intracellular trafficking in nonviral gene delivery.," *Pharmacol. Rev.*, vol. 58, no. 1, pp. 32–45, Mar. 2006.

[54] A. P. Lam and D. A. Dean, "Progress and prospects: nuclear import of nonviral vectors.," *Gene Ther.*, vol. 17, no. 4, pp. 439–447, Apr. 2010.

[55] J. Mercer, M. Schelhaas, and A. Helenius, "Virus entry by endocytosis.," *Annu. Rev. Biochem.*, vol. 79, no. 1, pp. 803–833, 2010.

[56] F. Melchior and L. Gerace, "Mechanisms of nuclear protein import.," *Curr. Opin. Cell Biol.*, vol. 7, no. 3, pp. 310–318, Jun. 1995.

[57] H. Kamiya, Y. Fujimura, I. Matsuoka, and H. Harashima, "Visualization of intracellular trafficking of exogenous DNA delivered by cationic liposomes.," *Biochemical and Biophysical Research Communications*, vol. 298, no. 4, pp. 591–597, Nov. 2002.

[58] D. S. Anson, "The use of retroviral vectors for gene therapy-what are the risks? A review of retroviral pathogenesis and its relevance to retroviral vector-mediated gene delivery.," *Genet Vaccines Ther*, vol. 2, no. 1, p. 9, Aug. 2004.

[59] S. E. Raper, N. Chirmule, F. S. Lee, N. A. Wivel, A. Bagg, G.-P. Gao, J. M. Wilson, and M. L. Batshaw, "Fatal systemic inflammatory response syndrome in a ornithine transcarbamylase deficient patient following adenoviral gene transfer.," *Mol. Genet. Metab.*, vol. 80, no. 1, pp. 148–158, Sep. 2003.

[60] F. Mingozzi, M. V. Maus, D. J. Hui, D. E. Sabatino, S. L. Murphy, J. E. J. Rasko, M. V. Ragni, C. S. Manno, J. Sommer, H. Jiang, G. F. Pierce, H. C. J. Ertl, and K. A. High, "CD8(+) T-cell responses to adeno-associated virus capsid in humans.," *Nat. Med.*, vol. 13, no. 4, pp. 419–422, Apr. 2007.

[61] F. Mingozzi and K. A. High, "Immune responses to AAV vectors: overcoming barriers to successful gene therapy.," *Blood*, vol. 122, no. 1, pp. 23–36, Jul. 2013.

[62] *advances-in-genetics-non-viral-vectors-for-gene-therapy*, vol. 89. Elsevier Inc., 2015, pp. 1–297.

[63] O. Danos and R. C. Mulligan, "Safe and efficient generation of recombinant retroviruses with amphotropic and ecotropic host ranges.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 85, no. 17, pp. 6460–6464, Sep. 1988.

[64] L. Naldini, U. Blömer, P. Gallay, D. Ory, R. Mulligan, F. H. Gage, I. M. Verma, and D. Trono, "In vivo gene delivery and stable transduction of nondividing cells by a lentiviral vector.," *Science*, vol. 272, no. 5259, pp. 263–267, Apr. 1996.

[65] S. Daya and K. I. Berns, "Gene therapy using adeno-associated virus vectors.," *Clin. Microbiol. Rev.*, vol. 21, no. 4, pp. 583–593, Oct. 2008.

[66] P. L. Felgner, T. R. Gadek, M. Holm, R. Roman, H. W. Chan, M. Wenz, J. P. Northrop, G. M. Ringold, and M. Danielsen, "Lipofection: a highly efficient, lipid-mediated DNA-transfection procedure.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 84, no. 21, pp. 7413–7417, Nov. 1987.

[67] M. Ogris, S. Brunner, S. Schüller, R. Kircheis, and E. Wagner, "PEGylated DNA/transferrin-PEI complexes: reduced interaction with blood components, extended circulation in blood and potential for systemic gene delivery.," *Gene Ther.*, vol. 6, no. 4, pp. 595–605, Apr. 1999.

[68] A. L. Parker, C. Newman, S. Briggs, L. Seymour, and P. J. Sheridan, "Nonviral gene delivery: techniques and implications for molecular medicine.," *Expert Reviews in Molecular Medicine*, vol. 5, no. 22, pp. 1–15, Sep. 2003.

[69] M. R. Capecchi, "High efficiency transformation by direct microinjection of DNA into cultured mammalian cells.," *Cell*, vol. 22, no. 2, pp. 479–488, Nov. 1980.

[70] R. M. Klein, E. D. Wolf, R. Wu, and J. C. Sanford, *High-velocity microprojectiles for delivering nucleic acids into living cells. 1987.*, vol. 24. 1992, pp. 384–386.

[71] E. Neumann, M. Schaefer-Ridder, Y. Wang, and P. H. Hofschneider, "Gene transfer into mouse lyoma cells by electroporation in high electric fields.," *EMBO J.*, vol. 1, no. 7, pp. 841–845, 1982.

[72] M. Zeitelhofer, J. P. Vessey, Y. Xie, F. Tübing, S. Thomas, M. Kiebler, and R. Dahm, "High-efficiency transfection of mammalian neurons via nucleofection.," *Nat Protoc*, vol. 2, no. 7, pp. 1692–1704, 2007.

[73] D. Bommannan, H. Okuyama, P. Stauffer, and R. H. Guy, "Sonophoresis. I. The use of high-frequency ultrasound to enhance transdermal drug delivery.," *Pharm. Res.*, vol. 9, no. 4, pp. 559–564, Apr. 1992.

[74] K. J. Widder and A. E. Senyei, "Magnetic microspheres: a vehicle for selective targeting of drugs.," *Pharmacol. Ther.*, vol. 20, no. 3, pp. 377–395, 1983.

[75] A. Sharei, J. Zoldan, A. Adamo, W. Y. Sim, N. Cho, E. Jackson, S. Mao, S. Schneider, M.-J. Han, A. Lytton-Jean, P. A. Basto, S. Jhunjhunwala, J. Lee, D. A. Heller, J. W. Kang, G. C. Hartoularos, K.-S. Kim, D. G. Anderson, R. Langer, and K. F. Jensen, "A vector-free microfluidic platform for intracellular delivery.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 110, no. 6, pp. 2082–2087, Feb. 2013.

[76] R. C. Rao and D. N. Zacks, "Cell and gene therapy.," *Dev Ophthalmol*, vol. 53, pp. 167–177, 2014.

[77] C. Hardee, L. Arévalo-Soliz, B. Hornstein, and L. Zechiedrich, "Advances in Non-Viral DNA Vectors for Gene Therapy," *Genes*, vol. 8, no. 2, pp. 65–22, Feb. 2017.

[78] M. S. Al-Dosari and X. Gao, "Nonviral gene delivery: principle, limitations, and recent progress.," *AAPS J*, vol. 11, no. 4, pp. 671–681, Dec. 2009.

[79] R. Remus, C. Kämmer, H. Heller, B. Schmitz, G. Schell, and W. Doerfler, "Insertion of foreign DNA into an established mammalian genome can alter the methylation of cellular DNA sequences.," *J. Virol.*, vol. 73, no. 2, pp. 1010–1022, Feb. 1999.

[80] D. Garrick, S. Fiering, D. I. Martin, and E. Whitelaw, "Repeat-induced gene silencing in mammals.," *Nat. Genet.*, vol. 18, no. 1, pp. 56–59, Jan. 1998.

[81] A. D. Miller, J. V. Garcia, N. von Suhr, C. M. Lynch, C. Wilson, and M. V. Eiden, "Construction and properties of retrovirus packaging cells based on gibbon ape leukemia virus.," *J. Virol.*, vol. 65, no. 5, pp. 2220–2224, May 1991.

[82] X. Wu, Y. Li, B. Crise, and S. M. Burgess, "Transcription start regions in the human genome are favored targets for MLV integration.," *Science*, vol. 300, no. 5626, pp. 1749–1751, Jun. 2003.

[83] E. Montini, D. Cesana, M. Schmidt, F. Sanvito, C. C. Bartholomae, M. Ranzani, F. Benedicenti, L. S. Sergi, A. Ambrosi, M. Ponzoni, C. Doglioni, C. Di Serio, C. von Kalle, and L. Naldini, "The genotoxic potential of retroviral

vectors is strongly modulated by vector design and integration site selection in a mouse model of HSC gene therapy.," *J. Clin. Invest.*, vol. 119, no. 4, pp. 964–975, Apr. 2009.

[84] T. Maetzig, M. Galla, C. Baum, and A. Schambach, "Gammaretroviral vectors: biology, technology and application.," *Viruses*, vol. 3, no. 6, pp. 677–713, Jun. 2011.

[85] L. Naldini, U. Blömer, F. H. Gage, D. Trono, and I. M. Verma, "Efficient transfer, integration, and sustained long-term expression of the transgene in adult rat brains injected with a lentiviral vector.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 93, no. 21, pp. 11382–11388, Oct. 1996.

[86] Y. D. Korin and J. A. Zack, "Progression to the G1b phase of the cell cycle is required for completion of human immunodeficiency virus type 1 reverse transcription in T cells.," *J. Virol.*, vol. 72, no. 4, pp. 3161–3168, Apr. 1998.

[87] K. Breckpot, D. Escors, F. Arce, L. Lopes, K. Karwacz, S. Van Lint, M. Keyaerts, and M. Collins, "HIV-1 lentiviral vector immunogenicity is mediated by Toll-like receptor 3 (TLR3) and TLR7.," *J. Virol.*, vol. 84, no. 11, pp. 5627–5636, Jun. 2010.

[88] A. R. W. Schröder, P. Shinn, H. Chen, C. Berry, J. R. Ecker, and F. Bushman, "HIV-1 integration in the human genome favors active genes and local hotspots.," *Cell*, vol. 110, no. 4, pp. 521–529, Aug. 2002.

[89] L. Apolonia, S. N. Waddington, C. Fernandes, N. J. Ward, G. Bouma, M. P. Blundell, A. J. Thrasher, M. K. Collins, and N. J. Philpott, "Stable gene transfer to muscle using non-integrating lentiviral vectors.," *Mol Ther*, vol. 15, no. 11, pp. 1947–1954, Nov. 2007.

[90] F. Kreppel and S. Kochanek, "Modification of adenovirus gene transfer vectors with synthetic polymers: a scientific review and technical guide.," *Mol. Ther.*, vol. 16, no. 1, pp. 16–29, Jan. 2008.

[91] M. Stadtfeld, M. Nagaya, J. Utikal, G. Weir, and K. Hochedlinger, "Induced pluripotent stem cells generated without viral integration.," *Science*, vol. 322, no. 5903, pp. 945–949, Nov. 2008.

[92] T. Reid, R. Warren, and D. Kirn, "Intravascular adenoviral agents in cancer patients: lessons from clinical trials.," *Cancer Gene Ther.*, vol. 9, no. 12, pp. 979–986, Dec. 2002.

[93] E. Marshall, "Gene therapy death prompts review of adenovirus vector.," *Science*, vol. 286, no. 5448, pp. 2244–2245, Dec. 1999.

[94] K. L. Molnar-Kimber, D. H. Sterman, M. Chang, E. H. Kang, M. ElBash, M. Lanuti, A. Elshami, K. Gelfand, J. M. Wilson, L. R. Kaiser, and S. M. Albelda, "Impact of preexisting and induced humoral and cellular immune responses in an adenovirus-based gene therapy phase I clinical trial for localized mesothelioma.," *Hum. Gene Ther.*, vol. 9, no. 14, pp. 2121–2133, Sep. 1998.

[95] A. Asokan, D. V. Schaffer, and R. J. Samulski, "The AAV vector toolkit: poised at the clinical crossroads.," *Mol. Ther.*, vol. 20, no. 4, pp. 699–708, Apr. 2012.

[96] J. F. Wright, "Transient transfection methods for clinical adeno-associated viral vector production.," *Hum. Gene Ther.*, vol. 20, no. 7, pp. 698–706, Jul. 2009.

[97] C. Summerford and R. J. Samulski, "Membrane-associated heparan sulfate proteoglycan is a receptor for adeno-associated virus type 2 virions.," *J. Virol.*, vol. 72, no. 2, pp. 1438–1445, Feb. 1998.

[98] C. Summerford, J. S. Bartlett, and R. J. Samulski, "AlphaVbeta5 integrin: a co-receptor for adeno-associated virus type 2 infection.," *Nat. Med.*, vol. 5, no. 1, pp. 78–82, Jan. 1999.

[99] A. Asokan, J. B. Hamra, L. Govindasamy, M. Agbandje-McKenna, and R. J. Samulski, "Adeno-associated virus type 2 contains an integrin alpha5beta1 binding domain essential for viral cell entry.," *J. Virol.*, vol. 80, no. 18, pp. 8961–8969, Sep. 2006.

[100] K. Qing, C. Mah, J. Hansen, S. Zhou, V. Dwarki, and A. Srivastava, "Human fibroblast growth factor receptor 1 is a co-receptor for infection by adeno-associated virus 2.," *Nat. Med.*, vol. 5, no. 1, pp. 71–77, Jan. 1999.

[101] B. Akache, D. Grimm, K. Pandey, S. R. Yant, H. Xu, and M. A. Kay, "The 37/67-kilodalton laminin receptor is a receptor for adeno-associated virus serotypes 8, 2, 3, and 9.," *J. Virol.*, vol. 80, no. 19, pp. 9831–9836, Oct. 2006.

[102] M. Nakanishi and M. Otsu, "Development of Sendai virus vectors and their potential applications in gene therapy and regenerative medicine.," *Curr Gene Ther*, vol. 12, no. 5, pp. 410–416, Oct. 2012.

[103] V. Hukkanen, "Herpesvirus vectors in gene therapy.," *Open Virol J*, vol. 4, no. 3, pp. 94–95, Jun. 2010.

[104] E. Hellebrand, J. Mautner, G. Reisbach, F. Nimmerjahn, M. Hallek, R. Mocikat, and W. Hammerschmidt, "Epstein-Barr virus vector-mediated gene transfer into human B cells: potential for antitumor vaccination.," *Gene Ther.*, vol. 13, no. 2, pp. 150–162, Jan. 2006.

[105] S. Stoll, "Epstein–Barr Virus/Human Vector Provides High-Level, Long-Term Expression of α1-Antitrypsin in Mice," *Mol Ther*, vol. 4, no. 2, pp. 122–129, Aug. 2001.

[106] S. Humme, G. Reisbach, R. Feederle, H.-J. Delecluse, K. Bousset, W. Hammerschmidt, and A. Schepers, "The EBV nuclear antigen 1 (EBNA1) enhances B cell immortalization several thousandfold.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 100, no. 19, pp. 10989–10994, Sep. 2003.

[107] M. J. Mastrangelo, L. C. Eisenlohr, L. Gomella, and E. C. Lattime, "Poxvirus vectors: orphaned and underappreciated.," *J. Clin. Invest.*, vol. 105, no. 8, pp. 1031–1034, Apr. 2000.

[108] D. Moroziewicz and H. L. Kaufman, "Gene therapy with poxvirus vectors.," *Curr. Opin. Mol. Ther.*, vol. 7, no. 4, pp. 317–325, Aug. 2005.

[109] B. Moss, "Poxvirus DNA replication.," *Cold Spring Harb Perspect Biol*, vol. 5, no. 9, pp. a010199–a010199, Sep. 2013.

[110] A. M. Darquet, R. Rangara, P. Kreiss, B. Schwartz, S. Naimi, P. Delaère, J. Crouzet, and D. Scherman, "Minicircle: an improved DNA molecule for in vitro and in vivo gene transfer.," *Gene Ther.*, vol. 6, no. 2, pp. 209–218, Feb. 1999.

[111] B. W. Bigger, O. Tolmachov, J. M. Collombet, M. Fragkos, I. Palaszewski, and C. Coutelle, "An araC-controlled bacterial cre expression system to produce DNA minicircle vectors for nuclear and mitochondrial gene therapy.," *J. Biol. Chem.*, vol. 276, no. 25, pp. 23018–23027, Jun. 2001.

[112] K. Nehlsen, S. Broll, and J. Bode, "Replicating minicircles: Generation of nonviral episomes for the efficient modification of dividing cells," *Gene Therapy and Molecular Biology*, vol. 10, pp. 233–244, Sep. 2006.

[113] Z. Y. Chen, C.-Y. He, A. Ehrhardt, and M. A. Kay, "Minicircle DNA vectors devoid of bacterial DNA result in persistent and high-level transgene expression in vivo.," *Mol Ther*, vol. 8, no. 3, pp. 495–500, Sep. 2003.

[114] A. Barahona, *Barbara McClintock and the transposition concept.*, vol. 46, no. 137. 1997, pp. 309–329.

[115] Z. Ivics, P. B. Hackett, R. H. Plasterk, and Z. Izsvák, "Molecular reconstruction of Sleeping Beauty, a Tc1-like transposon from fish, and its transposition in human cells.," *Cell*, vol. 91, no. 4, pp. 501–510, Nov. 1997.

[116] J. A. Wolff, J. J. Ludtke, G. Acsadi, P. Williams, and A. Jani, "Long-term persistence of plasmid DNA and foreign gene expression in mouse muscle.," *Hum. Mol. Genet.*, vol. 1, no. 6, pp. 363–369, Sep. 1992.

[117] G. McLachlan, B. J. Stevenson, D. J. Davidson, and D. J. Porteous, "Bacterial DNA is implicated in the inflammatory response to delivery of DNA/DOTAP to mouse lungs.," *Gene Ther.*, vol. 7, no. 5, pp. 384–392, Mar. 2000.

[118] Z. Izsvák, M. K. L. Chuah, T. VandenDriessche, and Z. Ivics, "Efficient stable gene transfer into human cells by the Sleeping Beauty transposon vectors.," *Methods*, vol. 49, no. 3, pp. 287–297, Nov. 2009.

[119] B. D. Hornstein, D. Roman, L. M. Arévalo-Soliz, M. A. Engevik, and L. Zechiedrich, "Effects of Circular DNA Length on Transfection Efficiency by Electroporation into HeLa Cells.," *PLoS ONE*, vol. 11, no. 12, p. e0167537, 2016.

[120] A. M. Darquet, B. Cameron, P. Wils, D. Scherman, and J. Crouzet, "A new DNA vehicle for nonviral gene delivery: supercoiled minicircle.," *Gene Ther.*, vol. 4, no. 12, pp. 1341–1349, Dec. 1997.

[121] S. S. Heinzel, P. J. Krysan, C. T. Tran, and M. P. Calos, "Autonomous DNA replication in human cells is affected by the size and the source of the DNA.," *Mol. Cell. Biol.*, vol. 11, no. 4, pp. 2263–2272, Apr. 1991.

[122] P. J. Krysan and M. P. Calos, "Replication initiates at multiple locations on an autonomously replicating plasmid in human cells.," *Mol. Cell. Biol.*, vol. 11, no. 3, pp. 1464–1472, Mar. 1991.

[123] J. Wang and B. Sugden, "Origins of bidirectional replication of Epstein-Barr virus: Models for understanding mammalian origins of DNA synthesis," *J. Cell. Biochem.*, vol. 94, no. 2, pp. 247–256, 2005.

[124] M. M. P. Lufino, R. Manservigi, and R. Wade-Martins, "An S/MAR-based infectious episomal genomic DNA expression vector provides long-term regulated functional complementation of LDLR deficiency," *Nucleic Acids Research*, vol. 35, no. 15, pp. e98–e98, Jul. 2007.

[125] S. P. Wong, O. Argyros, C. Coutelle, and R. P. Harbottle, "Strategies for the episomal modification of cells.," *Curr. Opin. Mol. Ther.*, vol. 11, no. 4, pp. 433–441, Aug. 2009.

[126] C. Hagedorn, A. Baiker, J. Postberg, A. Ehrhardt, and H. J. Lipps, "Handling S/MAR vectors.," *Cold Spring Harb Protoc*, vol. 2012, no. 6, pp. 657–663, Jun. 2012.

[127] O. Mazda, E. Satoh, K. Yasutomi, and J. Imanishi, "Extremely efficient gene transfection into lympho-hematopoietic cell lines by Epstein-Barr virus-based vectors.," *J. Immunol. Methods*, vol. 204, no. 2, pp. 143–151, May 1997.

[128] O. Argyros, S. P. Wong, C. Coutelle, and R. P. Harbottle, "Utilising Minicircle Vectors for the Episomal Modication of Cells," in *Minicircle and Miniplasmid DNA Vectors*, no. 7, M. Schleef, Ed. 2013, pp. 1–25.

[129] S. Stenler, P. Blomberg, and C. I. E. Smith, "Safety and efficacy of DNA vaccines: plasmids vs. minicircles.," *Hum Vaccin Immunother*, vol. 10, no. 5, pp. 1306–1308, 2014.

[130] S. P. Wong, O. Argyros, and R. P. Harbottle, "Sustained expression from DNA vectors.," *Adv. Genet.*, vol. 89, pp. 113–152, 2015.

[131] B. L. Hodges, K. M. Taylor, M. F. Joseph, S. A. Bourgeois, and R. K. Scheule, "Long-term transgene expression from plasmid DNA gene therapy vectors is negatively affected by CpG dinucleotides.," *Mol Ther*, vol. 10, no. 2, pp. 269–278, Aug. 2004.

[132] Z. Siegfried, S. Eden, M. Mendelsohn, X. Feng, B. Z. Tsuberi, and H. Cedar, "DNA methylation represses transcription in vivo.," *Nat. Genet.*, vol. 22, no. 2, pp. 203–206, Jun. 1999.

[133] E. Riu, Z. Y. Chen, H. Xu, C.-Y. He, and M. A. Kay, "Histone modifications are associated with the persistence or silencing of vector-mediated transgene expression in vivo.," *Mol. Ther.*, vol. 15, no. 7, pp. 1348–1355, Jul. 2007.

[134] H. Hemmi, O. Takeuchi, T. Kawai, T. Kaisho, S. Sato, H. Sanjo, M. Matsumoto, K. Hoshino, H. Wagner, K. Takeda, and S. Akira, "A Toll-like receptor recognizes bacterial DNA.," *Nature*, vol. 408, no. 6813, pp. 740–745, Dec. 2000.

[135] H.-S. Zhou, D.-P. Liu, and C.-C. Liang, "Challenges and strategies: the immune responses in gene therapy.," *Med Res Rev*, vol. 24, no. 6, pp. 748–761, Nov. 2004.

[136] G. Gribaudo, S. Ravaglia, A. Caliendo, R. Cavallo, M. Gariglio, M. G. Martinotti, and S. Landolfo, "Interferons inhibit onset of murine cytomegalovirus immediate-early gene transcription.," *Virology*, vol. 197, no. 1, pp. 303–311, Nov. 1993.

[137] H. Zhao, H. Hemmi, S. Akira, S. H. Cheng, R. K. Scheule, and N. S. Yew, "Contribution of Toll-like receptor 9 signaling to the acute inflammatory response to nonviral vectors.," *Mol Ther*, vol. 9, no. 2, pp. 241–248, Feb. 2004.

[138] A. Sarukhan, S. Camugli, B. Gjata, H. von Boehmer, O. Danos, and K. Jooss, "Successful interference with cellular immune responses to immunogenic proteins encoded by recombinant viral vectors.," *J. Virol.*, vol. 75, no. 1, pp. 269–277, Jan. 2001.

[139] C. Plank, K. Mechtler, F. C. Szoka, and E. Wagner, "Activation of the complement system by synthetic DNA complexes: a potential barrier for intravenous gene delivery.," *Hum. Gene Ther.*, vol. 7, no. 12, pp. 1437–1446, Aug. 1996.

[140] S. Seisenberger, S. Andrews, F. Krueger, J. Arand, J. Walter, F. Santos, C. Popp, B. Thienpont, W. Dean, and W. Reik, "The dynamics of genome-wide DNA methylation reprogramming in mouse primordial germ cells.," *Mol. Cell*, vol. 48, no. 6, pp. 849–862, Dec. 2012.

[141] S. P. Medvedev, E. A. Pokushalov, and S. M. Zakian, "Epigenetics of pluripotent cells.," *Acta Naturae*, vol. 4, no. 4, pp. 28–46, Oct. 2012.

[142] K. Takahashi and S. Yamanaka, "Induction of Pluripotent Stem Cells from Mouse Embryonic and Adult Fibroblast Cultures by Defined Factors," *Cell*, vol. 126, no. 4, pp. 663–676, Aug. 2006.

[143] B. Papp and K. Plath, "Epigenetics of Reprogramming to Induced Pluripotency," *Cell*, vol. 152, no. 6, pp. 1324–1343, 2013.

[144] J. Ohgane, S. Yagi, and K. Shiota, "Epigenetics: the DNA methylation profile of tissue-dependent and differentially methylated regions in cells.," *Placenta*, vol. 29, pp. S29–35, Mar. 2008.

[145] M. J. Boland, K. L. Nazor, and J. F. Loring, "Epigenetic regulation of pluripotency and differentiation.," *Circ. Res.*, vol. 115, no. 2, pp. 311–324, Jul. 2014.

[146] A. Fiorini, F. de S. Gouveia, and M. A. Fernandez, "Scaffold/Matrix Attachment Regions and intrinsic DNA curvature.," *Biochemistry Mosc.*, vol. 71, no. 5, pp. 481–488, May 2006.

[147] P.-A. Girod and N. Mermod, *Use of scaffold/matrix-attachment regions for protein production*, vol. 38. Elsevier, 2003, pp. 359–379.

[148] M. J. Deutsch, E. Ott, P. Papior, and A. Schepers, "The latent origin of replication of Epstein-Barr virus directs viral genomes to active regions of the nucleus.," *J. Virol.*, vol. 84, no. 5, pp. 2533–2546, Mar. 2010.

[149] I. M. Stehle, J. Postberg, S. Rupprecht, T. Cremer, D. A. Jackson, and H. J. Lipps, "Establishment and mitotic stability of an extra-chromosomal mammalian replicon," *BMC Cell Biol.*, vol. 8, no. 1, p. 33, 2007.

[150] C. Hagedorn, A. Gogol-Döring, S. Schreiber, J. T. Epplen, and H. J. Lipps, "Genome-wide profiling of S/MAR-based replicon contact sites," *Nucleic Acids Research*, vol. 45, no. 13, pp. 7841–7854, Jun. 2017.

[151] P. R. Cook, "The organization of replication and transcription.," *Science*, vol. 284, no. 5421, pp. 1790–1795, Jun. 1999.

[152] A. Ehrhardt, R. Haase, A. Schepers, M. J. Deutsch, H. J. Lipps, and A. Baiker, "Episomal vectors for gene therapy.," *Curr Gene Ther*, vol. 8, no. 3, pp. 147–161, Jun. 2008.

[153] H. Nakai, E. Montini, S. Fuess, T. A. Storm, M. Grompe, and M. A. Kay, "AAV serotype 2 vectors preferentially integrate into active genes in mice.," *Nat. Genet.*, vol. 34, no. 3, pp. 297–302, Jul. 2003.

[154] D. Li, R. Zhao, W. Lilyestrom, D. Gai, R. Zhang, J. A. DeCaprio, E. Fanning, A. Jochimiak, G. Szakonyi, and X. S. Chen, "Structure of the replicative helicase of the oncoprotein SV40 large tumour antigen.," *Nature*, vol. 423, no. 6939, pp. 512–518, May 2003.

[155] H. Wu, D. F. Ceccarelli, and L. Frappier, "The DNA segregation mechanism of Epstein-Barr virus nuclear antigen 1.," *EMBO Rep.*, vol. 1, no. 2, pp. 140–144, Aug. 2000.

[156] D. K. Snudden, J. Hearing, P. R. Smith, F. A. Grässer, and B. E. Griffin, "EBNA-1, the major nuclear antigen of Epstein-Barr virus, resembles 'RGG' RNA binding proteins.," *EMBO J.*, vol. 13, no. 20, pp. 4840–4847, Oct. 1994.

[157] N. S. Sung and J. S. Pagano, "Molecular Mechanisms of Transformation by Epstein-Barr Virus," in *DNA Tumor Viruses*, G. Barbanti-Brodano, M. Bendinelli, and H. Friedman, Eds. Boston, MA: Springer US, 1995, pp. 1–43.

[158] S. H. Ali, J. S. Kasper, T. Arai, and J. A. DeCaprio, "Cul7/p185/p193 binding to simian virus 40 large T antigen has a role in cellular transformation.," *J. Virol.*, vol. 78, no. 6, pp. 2749–2757, Mar. 2004.

[159] B. P. Duncker, I. N. Chesnokov, and B. J. McConkey, "The origin recognition complex protein family.," *Genome Biol.*, vol. 10, no. 3, p. 214, 2009.

[160] D. T. Stinchcomb, M. Thomas, J. Kelly, E. Selker, and R. W. Davis, "Eukaryotic DNA segments capable of autonomous replication in yeast.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 77, no. 8, pp. 4559–4563, Aug. 1980.

[161] T. O. Nielsen, N. H. Cossons, M. Zannis-Hadjopoulos, and G. B. Price, "Circular YAC vectors containing short mammalian origin sequences are maintained under selection as HeLa episomes.," *J. Cell. Biochem.*, vol. 76, no. 4, pp. 674–685, Jan. 2000.

[162] C. Piechaczek, C. Fetzer, A. Baiker, J. Bode, and H. J. Lipps, "A vector based on the SV40 origin of replication and chromosomal S/MARs replicates episomally in CHO cells.," *Nucleic Acids Research*, vol. 27, no. 2, pp. 426–428, Jan. 1999.

[163] I. M. Stehle, M. F. Scinteie, A. Baiker, A. C. W. Jenke, and H. J. Lipps, "Exploiting a minimal system to study the epigenetic control of DNA replication: the interplay between transcription and replication.," *Chromosome Res.*, vol. 11, no. 5, pp. 413–421, 2003.

[164] C. Hagedorn, A. Baiker, J. Postberg, A. Ehrhardt, and H. J. Lipps, "Handling S/MAR Vectors," *Cold Spring Harb Protoc*, vol. 2012, no. 6, pp. pdb.top068262–pdb.top068262, Jun. 2012.

[165] E. P. Papapetrou, P. G. Ziros, I. D. Micheva, N. C. Zoumbos, and A. Athanassiadou, "Gene transfer into human hematopoietic progenitor cells with an episomal vector carrying an S/MAR element.," *Gene Ther.*, vol. 13, no. 1, pp. 40–51, Jan. 2006.

[166] C. Hagedorn and H. J. Lipps, "pEPI for Gene Therapy Non viral episomes and their Application in Somatic Gene Therapy," *J Cell Sci Ther*, vol. 4, no. 2, pp. 1–8, 2013.

[167] S. Manzini, A. Vargiolu, D. Seruggia, M. Cerrito, and M. Busnelli, "S/MAR trek: A new generation of non-viral episomal vectors to be used in gene transfer experiments

," *Transgenic Res.*, no. 19, pp. 307–355, 2010.

[168] S. P. Wong, O. Argyros, C. Coutelle, and R. P. Harbottle, "Non-viral S/MAR vectors replicate episomally in vivo when provided with a selective advantage.," *Gene Ther.*, vol. 18, no. 1, pp. 82–87, Jan. 2011.

[169] O. Argyros, S. P. Wong, M. Niceta, S. N. Waddington, S. J. Howe, C. Coutelle, A. D. Miller, and R. P. Harbottle, "Persistent episomal transgene expression in liver following delivery of a scaffold/matrix attachment region containing non-viral vector.," *Gene Ther.*, vol. 15, no. 24, pp. 1593–1605, Dec. 2008.

[170] R. Haase, O. Argyros, S. P. Wong, R. P. Harbottle, H. J. Lipps, M. Ogris, T. Magnusson, M. Pinto, J. Haas, and A. Baiker, "pEPito: a significantly improved non-viral episomal expression vector for mammalian cells," *BMC Biotechnol.*, vol. 10, no. 1, pp. 20–14, 2010.

[171] O. Argyros, S. P. Wong, C. Fedonidis, O. Tolmachov, S. N. Waddington, S. J. Howe, M. Niceta, C. Coutelle, and R. P. Harbottle, "Development of S/MAR minicircles for enhanced and persistent transgene expression in the mouse liver," *J. Mol. Med.*, vol. 89, no. 5, pp. 515–529, Feb. 2011.

[172] J. Mirkovitch, M. E. Mirault, and U. K. Laemmli, "Organization of the higher-order chromatin loop: specific DNA attachment sites on nuclear scaffold.," *Cell*, vol. 39, no. 1, pp. 223–232, Nov. 1984.

[173] J. Bode, C. Benham, A. Knopp, and C. Mielke, "Transcriptional augmentation: modulation of gene expression by scaffold/matrix-attached regions (S/MAR elements).," *Crit. Rev. Eukaryot. Gene Expr.*, vol. 10, no. 1, pp. 73–90, 2000.

[174] J. P. von Kries, H. Buhrmester, and W. H. Strätling, "A matrix/scaffold attachment region binding protein: identification, purification, and mode of binding.," *Cell*, vol. 64, no. 1, pp. 123–135, Jan. 1991.

[175] D. A. Jackson and P. R. Cook, "The structural basis of nuclear function.," *Int. Rev. Cytol.*, vol. 162, pp. 125–149, 1995.

[176] J. Bode, Y. Kohwi, L. Dickinson, T. Joh, D. Klehr, C. Mielke, and T. Kohwi-Shigematsu, "Biological significance of unwinding capability of nuclear matrix-associating DNAs.," *Science*, vol. 255, no. 5041, pp. 195–197, Jan. 1992.

[177] S. Goetze, A. Baer, S. Winkelmann, K. Nehlsen, J. Seibler, K. Maass, and J. Bode, "Performance of genomic bordering elements at predefined genomic loci.," *Mol. Cell. Biol.*, vol. 25, no. 6, pp. 2260–2272, Mar. 2005.

[178] M. Frisch, K. Frech, A. Klingenhoff, K. Cartharius, I. Liebich, and T. Werner, "In silico prediction of scaffold/matrix attachment regions in large genomic sequences.," *Genome Res.*, vol. 12, no. 2, pp. 349–354, Feb. 2002.

[179] I. de Belle, S. Cai, and T. Kohwi-Shigematsu, "The genomic sequences bound to special AT-rich sequence-binding protein 1 (SATB1) in vivo in Jurkat T cells are tightly associated with the nuclear matrix at the bases of the chromatin loops.," *J Cell Biol*, vol. 141, no. 2, pp. 335–348, Apr. 1998.

[180] A. C. W. Jenke, M. F. Scinteie, I. M. Stehle, and H. J. Lipps, "Expression of a transgene encoded on a non-viral episomal vector is not subject to epigenetic silencing by cytosine methylation.," *Mol. Biol. Rep.*, vol. 31, no. 2, pp. 85–90, Jun. 2004.

[181] S. Rupprecht, C. Hagedorn, D. Seruggia, T. Magnusson, E. Wagner, M. Ogris, and H. J. Lipps, "Controlled removal of a nonviral episomal vector from transfected cells," *Gene*, vol. 466, no. 1, pp. 36–42, Oct. 2010.

[182] Q. Dang, J. Auten, and I. Plavec, "Human beta interferon scaffold attachment region inhibits de novo methylation and confers long-term, copy number-dependent expression to a retroviral vector.," *J. Virol.*, vol. 74, no. 6, pp. 2671–2678, Mar. 2000.

[183] B. H. C. Jenke, C. P. Fetzer, I. M. Stehle, F. Jönsson, F. O. Fackelmayer, H. Conradt, J. Bode, and H. J. Lipps, "An episomally replicating vector binds to the nuclear matrix protein SAF-A in vivo.," *EMBO Rep.*, vol. 3, no. 4, pp. 349–354, Apr. 2002.

[184] H. J. Lipps and J. Bode, "Exploiting chromosomal and viral strategies: the design of safe and efficient non-viral gene transfer systems.," *Curr. Opin. Mol. Ther.*, vol. 3, no. 2, pp. 133–141, Apr. 2001.

[185] A. Stief, D. M. Winter, W. H. Strätling, and A. E. Sippel, "A nuclear DNA attachment element mediates elevated and position-independent gene activity.," *Nature*, vol. 341, no. 6240, pp. 343–345, Sep. 1989.

[186] D. Klehr, T. Schlake, K. Maass, and J. Bode, "Scaffold-attached regions (SAR elements) mediate transcriptional effects due to butyrate.," *Biochemistry*, vol. 31, no. 12, pp. 3222–3229, Mar. 1992.

[187] C. Bonifer, M. Vidal, F. Grosveld, and A. E. Sippel, "Tissue specific and position independent expression of the complete gene domain for chicken lysozyme in transgenic mice.," *EMBO J.*, vol. 9, no. 9, pp. 2843–2848, Sep. 1990.

[188] M. Lichtenstein, G. Keini, H. Cedar, and Y. Bergman, "B cell-specific demethylation: a novel role for the intronic kappa chain enhancer sequence.," *Cell*, vol. 76, no. 5, pp. 913–923, Mar. 1994.

[189] W. C. Forrester, L. A. Fernández, and R. Grosschedl, "Nuclear matrix attachment regions antagonize methylation-dependent repression of long-range enhancer-promoter interactions.," *Genes & Development*, vol. 13, no. 22, pp. 3003–3014, Nov. 1999.

[190] T. Boulikas, "Chromatin domains and prediction of MAR sequences.," *Int. Rev. Cytol.*, vol. 162, pp. 279–388, 1995.

[191] D. Ottaviani, E. Lever, P. Takousis, and D. Sheer, "Anchoring the genome.," *Genome Biol.*, vol. 9, no. 1, p. 201, Jan. 2008.

[192] N. Harraghy, A. Gaussin, and N. Mermod, "Sustained transgene expression using MAR elements.," *Curr Gene Ther*, vol. 8, no. 5, pp. 353–366, Oct. 2008.

[193] C. Mielke, Y. Kohwi, T. Kohwi-Shigematsu, and J. Bode, "Hierarchical binding of DNA fragments derived from scaffold-attached regions: correlation of properties in vitro and function in vivo.," *Biochemistry*, vol. 29, no. 32, pp. 7475–7485, Aug. 1990.

[194] G. C. Allen, G. Hall, S. Michalowski, W. Newman, S. Spiker, A. K. Weissinger, and W. F. Thompson, "High-level transgene expression in plant cells: effects of a strong scaffold attachment region from tobacco.," *Plant Cell*, vol. 8, no. 5, pp. 899–913, May 1996.

[195] I. M. Stehle, J. Postberg, S. Rupprecht, T. Cremer, D. A. Jackson, and H. J. Lipps, "Establishment and mitotic stability of an extra-chromosomal mammalian replicon," *BMC Cell Biol.*, vol. 8, no. 1, pp. 33–12, 2007.

[196] F. Tessadori, K. Zeng, E. Manders, M. Riool, D. Jackson, and R. van Driel, "Stable S/MAR-based episomal vectors are regulated at the chromatin level.," *Chromosome Res.*, vol. 18, no. 7, pp. 757–775, Nov. 2010.

[197] C. Hagedorn, M. N. Antoniou, and H. J. Lipps, "Genomic cis-acting Sequences Improve Expression and Establishment of a Nonviral Vector," *Mol Ther Nucleic Acids*, vol. 2, pp. e118–9, 2013.

[198] C. Hagedorn, M. N. Antoniou, and H. J. Lipps, "Genomic cis-acting Sequences Improve Expression and Establishment of a Nonviral Vector," *Mol Ther Nucleic Acids*, vol. 2, pp. e118–9, 2013.

[199] K. Schwanke, S. Merkert, H. Kempf, S. Hartung, M. Jara-Avaca, C. Templin, G. Göhring, A. Haverich, U. Martin, and R. Zweigerdt, "Fast and Efficient Multitransgenic Modification of Human Pluripotent Stem Cells," *Human Gene Therapy Methods*, vol. 25, no. 2, pp. 136–153, Apr. 2014.

[200] S. He, D. Nakada, and S. J. Morrison, "Mechanisms of stem cell self-renewal.," *Annu. Rev. Cell Dev. Biol.*, vol. 25, no. 1, pp. 377–406, 2009.

[201] K. O'Donoghue and N. M. Fisk, "Fetal stem cells.," *Best Pract Res Clin Obstet Gynaecol*, vol. 18, no. 6, pp. 853–875, Dec. 2004.

[202] K. Takahashi, K. Tanabe, M. Ohnuki, M. Narita, T. Ichisaka, K. Tomoda, and S. Yamanaka, "Induction of pluripotent stem cells from adult human fibroblasts by defined factors.," *Cell*, vol. 131, no. 5, pp. 861–872, Nov. 2007.

[203] K. Okita, Y. Matsumura, Y. Sato, A. Okada, A. Morizane, S. Okamoto, H. Hong, M. Nakagawa, K. Tanabe, K.-I. Tezuka, T. Shibata, T. Kunisada, M. Takahashi, J. Takahashi, H. Saji, and S. Yamanaka, "A more efficient method to generate integration-free human iPS cells," *Nat. Methods*, vol. 8, no. 5, pp. 409–412, Apr. 2011.

[204] K. Takahashi and S. Yamanaka, "Induced pluripotent stem cells in medicine and biology.," *Development*, vol. 140, no. 12, pp. 2457–2461, Jun. 2013.

[205] O. Forostyak, G. Dayanithi, and S. Forostyak, "CNS Regenerative Medicine and Stem Cells," *Opera Med Physiol*, vol. 2, no. 1, pp. 55–62, Aug. 2016.

[206] X.-Y. Zhao, W. Li, Z. Lv, L. Liu, M. Tong, T. Hai, J. Hao, C.-L. Guo, Q.-W. Ma, L. Wang, F. Zeng, and Q. Zhou, "iPS cells produce viable mice through tetraploid complementation.," *Nature*, vol. 461, no. 7260, pp. 86–90, Sep. 2009.

[207] L. Kang, J. Wang, Y. Zhang, Z. Kou, and S. Gao, "iPS cells can support full-term development of tetraploid blastocyst-complemented embryos.," *Cell Stem Cell*, vol. 5, no. 2, pp. 135–138, Aug. 2009.

[208] B. Nashun, P. W. S. Hill, and P. Hajkova, "Reprogramming of cell fate: epigenetic memory and the erasure of memories past.," *EMBO J.*, vol. 34, no. 10, pp. 1296–1308, May 2015.

[209] Y. Zhou, J. Kim, X. Yuan, and T. Braun, "Epigenetic modifications of stem cells: a paradigm for the control of cardiac progenitor cells.," *Circ. Res.*, vol. 109, no. 9, pp. 1067–1081, Oct. 2011.

[210] M. Bibikova, E. Chudin, B. Wu, L. Zhou, E. W. Garcia, Y. Liu, S. Shin, T. W. Plaia, J. M. Auerbach, D. E. Arking, R. Gonzalez, J. Crook, B. Davidson, T. C. Schulz, A. Robins, A. Khanna, P. Sartipy, J. Hyllner, P. Vanguri, S. Savant-Bhonsale, A. K. Smith, A. Chakravarti, A. Maitra, M. Rao, D. L. Barker, J. F. Loring, and J.-B. Fan, "Human embryonic stem cells have a unique epigenetic signature.," *Genome Res.*, vol. 16, no. 9, pp. 1075–1083, Sep. 2006.

[211] N. Maherali, R. Sridharan, W. Xie, J. Utikal, S. Eminli, K. Arnold, M. Stadtfeld, R. Yachechko, J. Tchieu, R. Jaenisch, K. Plath, and K. Hochedlinger, "Directly reprogrammed fibroblasts show global epigenetic remodeling and widespread tissue contribution.," *Cell Stem Cell*, vol. 1, no. 1, pp. 55–70, Jun. 2007.

[212] M. G. Guenther, G. M. Frampton, F. Soldner, D. Hockemeyer, M. Mitalipova, R. Jaenisch, and R. A. Young, "Chromatin structure and gene expression programs of human embryonic and induced pluripotent stem cells.," *Cell Stem Cell*, vol. 7, no. 2, pp. 249–257, Aug. 2010.

[213] "Highly Efficient miRNA-Mediated Reprogramming of Mouse and Human Somatic Cells to Pluripotency," vol. 8, no. 4, pp. 376–388, Apr. 2011.

[214] J. B. GURDON, T. R. ELSDALE, and M. FISCHBERG, "Sexually mature individuals of Xenopus laevis from the transplantation of single somatic nuclei.," *Nature*, vol. 182, no. 4627, pp. 64–65, Jul. 1958.

[215] R. A. Miller and F. H. Ruddle, "Pluripotent teratocarcinoma-thymus somatic cell hybrids.," *Cell*, vol. 9, no. 1, pp. 45–55, Sep. 1976.

[216] "Powering Reprogramming with Vitamin C," vol. 6, no. 1, pp. 1–2, Jan. 2010.

[217] A. M. Drozd, M. P. Walczak, S. Piaskowski, E. Stoczynska-Fidelus, P. Rieske, and D. P. Grzela, "Generation of human iPSCs from cells of fibroblastic and epithelial origin by means of the oriP/EBNA-1 episomal reprogramming system," *Stem Cell Res Ther*, vol. 6, no. 1, pp. 663–18, Jun. 2015.

[218] K. Okita, M. Nakagawa, H. Hyenjong, T. Ichisaka, and S. Yamanaka, "Induced pluripotent stem cell," *Science*, vol. 322, no. 5903. Wikipedia, pp. 949–953, 07-Nov-2008.

[219] A. G. Sommer, S. S. Rozelle, S. Sullivan, J. A. Mills, S.-M. Park, B. W. Smith, A. M. Iyer, D. L. French, D. N. Kotton, P. Gadue, G. J. Murphy, and G. Mostoslavsky, "Generation of human induced pluripotent stem cells from peripheral blood using the STEMCCA lentiviral vector.," *J Vis Exp*, no. 68, pp. e4327–e4327, Oct. 2012.

[220] X.-Y. Zhang, V. F. La Russa, L. Bao, J. Kolls, P. Schwarzenberger, and J. Reiser, "Lentiviral vectors for sustained transgene expression in human bone marrow-derived stromal cells.," *Mol Ther*, vol. 5, no. 5, pp. 555–565, May 2002.

[221] C. Laker, J. Meyer, A. Schopen, J. Friel, C. Heberlein, W. Ostertag, and C. Stocking, "Host cis-mediated extinction of a retrovirus permissive for expression in embryonal stem cells during differentiation.," *J. Virol.*, vol. 72, no. 1, pp. 339–348, Jan. 1998.

[222] S. Hacein-Bey-Abina, F. Le Deist, F. Carlier, C. Bouneaud, C. Hue, J.-P. De Villartay, A. J. Thrasher, N. Wulffraat, R. Sorensen, S. Dupuis-Girod, A. Fischer, E. G. Davies, W. Kuis, L. Leiva, and M. Cavazzana-Calvo, "Sustained correction of X-linked severe combined immunodeficiency by ex vivo gene therapy.," *N. Engl. J. Med.*, vol. 346, no. 16, pp. 1185–1193, Apr. 2002.

[223] S. Hacein-Bey-Abina, J. Hauer, A. Lim, C. Picard, G. P. Wang, C. C. Berry, C. Martinache, F. Rieux-Laucat, S. Latour, B. H. Belohradsky, L. Leiva, R. Sorensen, M. Debré, J. L. Casanova, S. Blanche, A. Durandy, F. D. Bushman, A. Fischer, and M. Cavazzana-Calvo, "Efficacy of gene therapy for X-linked severe combined immunodeficiency.," *N. Engl. J. Med.*, vol. 363, no. 4, pp. 355–364, Jul. 2010.

[224] R. Zufferey, T. Dull, R. J. Mandel, A. Bukovsky, D. Quiroz, L. Naldini, and D. Trono, "Self-inactivating lentivirus vector for safe and efficient in vivo gene delivery.," *J. Virol.*, vol. 72, no. 12, pp. 9873–9880, Dec. 1998.

[225] F. Herbst, C. R. Ball, F. Tuorto, A. Nowrouzi, W. Wang, O. Zavidij, S. M. Dieter, S. Fessler, F. van der Hoeven, U. Kloz, F. Lyko, M. Schmidt, C. von Kalle, and H. Glimm, "Extensive Methylation of Promoter Sequences Silences Lentiviral Transgene Expression During Stem Cell Differentiation In Vivo," *Mol. Ther.*, vol. 20, no. 5, pp. 1014–1021, Nov. 2009.

[226] N. Pfaff, N. Lachmann, M. Ackermann, S. Kohlscheen, C. Brendel, T. Maetzig, H. Niemann, M. N. Antoniou, M. Grez, A. Schambach, T. Cantz, and T. Moritz, "A ubiquitous chromatin opening element prevents transgene silencing in pluripotent stem cells and their differentiated progeny.," *Stem Cells*, vol. 31, no. 3, pp. 488–499, Mar. 2013.

[227] N. Brown, L. Song, N. R. Kollu, and M. L. Hirsch, "Adeno-Associated Virus Vectors and Stem Cells: Friends or Foes?," *Hum. Gene Ther.*, vol. 28, no. 6, pp. 450–463, Jun. 2017.

[228] C. E. Walsh, A. W. Nienhuis, R. J. Samulski, M. G. Brown, J. L. Miller, N. S. Young, and J. M. Liu, "Phenotypic correction of Fanconi anemia in human hematopoietic cells with a recombinant adeno-associated virus vector.," *J. Clin. Invest.*, vol. 94, no. 4, pp. 1440–1448, Oct. 1994.

[229] S. Z. Zhou, S. Cooper, L. Y. Kang, L. Ruggieri, S. Heimfeld, A. Srivastava, and H. E. Broxmeyer, "Adeno-associated virus 2-mediated high efficiency gene transfer into immature and mature subsets of hematopoietic progenitor cells in human umbilical cord blood.," *J Exp Med*, vol. 179, no. 6, pp. 1867–1875, Jun. 1994.

[230] B. L. Ellis, M. L. Hirsch, J. C. Barker, J. P. Connelly, R. J. Steininger, and M. H. Porteus, "A survey of ex vivo/in vitro transduction efficiency of mammalian primary cells and cell lines with Nine natural adeno-associated virus (AAV1-9) and one engineered adeno-associated virus serotype.," *Virol. J.*, vol. 10, no. 1, p. 74, Mar. 2013.

[231] S. Knaän-Shanzer, M. J. M. van de Watering, I. van der Velde, M. A. F. V. Gonçalves, D. Valerio, and A. A. F. de Vries, "Endowing human adenovirus serotype 5 vectors with fiber domains of species B greatly enhances gene transfer into human mesenchymal stem cells.," *Stem Cells*, vol. 23, no. 10, pp. 1598–1607, Nov. 2005.

[232] A. Zaldumbide, F. Carlotti, M. A. Gonçalves, S. Knaän-Shanzer, S. J. Cramer, B. O. Roep, E. J. H. J. Wiertz, and R. C. Hoeben, "Adenoviral vectors stimulate glucagon transcription in human mesenchymal stem cells expressing pancreatic transcription factors.," *PLoS ONE*, vol. 7, no. 10, p. e48093, 2012.

[233] Y. Yang, F. A. Nunes, K. Berencsi, E. E. Furth, E. Gönczöl, and J. M. Wilson, "Cellular immunity to viral antigens limits E1-deleted adenoviruses for gene therapy.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 91, no. 10, pp. 4407–4411, May 1994.

[234] S.-W. Han, C. Nakamura, N. Kotobuki, I. Obataya, H. Ohgushi, T. Nagamune, and J. Miyake, "High-efficiency DNA injection into a single human mesenchymal stem cell using a nanoneedle and atomic force microscopy.," *Nanomedicine*, vol. 4, no. 3, pp. 215–225, Sep. 2008.

[235] J. Gehl, "Electroporation: theory and methods, perspectives for drug delivery, gene therapy and research.," *Acta Physiol. Scand.*, vol. 177, no. 4, pp. 437–447, Apr. 2003.

[236] O. Gresch, F. B. Engel, D. Nesic, T. T. Tran, H. M. England, E. S. Hickman, I. Körner, L. Gan, S. Chen, S. Castro-Obregon, R. Hammermann, J. Wolf, H. Müller-Hartmann, M. Nix, G. Siebenkotten, G. Kraus, and K. Lun, "New non-viral method for gene transfer into primary cells.," *Methods*, vol. 33, no. 2, pp. 151–163, Jun. 2004.

[237] X. Cao, W. Deng, Y. Wei, W. Su, Y. Yang, Y. Wei, J. Yu, and X. Xu, "Encapsulation of plasmid DNA in calcium phosphate nanoparticles: stem cell uptake and gene transfer efficiency.," *Int J Nanomedicine*, vol. 6, pp. 3335–3349, 2011.

[238] J.-W. Cho, C.-Y. Lee, and Y. Ko, "Therapeutic potential of mesenchymal stem cells overexpressing human forkhead box A2 gene in the regeneration of damaged liver tissues.," *J. Gastroenterol. Hepatol.*, vol. 27, no. 8, pp. 1362–1370, Aug. 2012.

[239] S. Y. Jeon, J. S. Park, H. N. Yang, D. G. Woo, and K.-H. Park, "Co-delivery of SOX9 genes and anti-Cbfa-1 siRNA coated onto PLGA nanoparticles for chondrogenesis of human MSCs.," *Biomaterials*, vol. 33, no. 17, pp. 4413–4423, Jun. 2012.

[240] C. A. Sommer, A. G. Sommer, T. A. Longmire, C. Christodoulou, D. D. Thomas, M. Gostissa, F. W. Alt, G. J. Murphy, D. N. Kotton, and G. Mostoslavsky, "Excision of reprogramming transgenes improves the differentiation potential of iPS cells generated with a single excisable vector.," *Stem Cells*, vol. 28, no. 1, pp. 64–74, Jan. 2010.

[241] T. Maetzig, J. Kuehle, A. Schwarzer, S. Turan, M. Rothe, A. Chaturvedi, M. Morgan, T. C. Ha, M. Heuser, W. Hammerschmidt, C. Baum, and A. Schambach, "All-in-One inducible lentiviral vector systems based on drug controlled FLP recombinase," *Biomaterials*, vol. 35, no. 14, pp. 4345–4356, May 2014.

[242] T. R. Talluri, D. Kumar, S. Glage, W. Garrels, Z. Ivics, K. Debowski, R. Behr, and W. A. Kues, "Non-viral reprogramming of fibroblasts into induced pluripotent stem cells by Sleeping Beauty and piggyBac transposons.," *Biochemical and Biophysical Research Communications*, vol. 450, no. 1, pp. 581–587, Jul. 2014.

[243] I. Grabundzija, J. Wang, A. Sebe, Z. Erdei, R. Kajdi, A. Devaraj, D. Steinemann, K. Szuhai, U. Stein, T. Cantz, A. Schambach, C. Baum, Z. Izsvák, B. Sarkadi, and Z. Ivics, "Sleeping Beauty transposon-based system for cellular reprogramming and targeted gene insertion in induced pluripotent stem cells.," *Nucleic Acids Research*, vol. 41, no. 3, pp. 1829–1847, Feb. 2013.

[244] K. Woltjen, I. P. Michael, P. Mohseni, R. Desai, M. Mileikovsky, R. Hämäläinen, R. Cowling, W. Wang, P. Liu, M. Gertsenstein, K. Kaji, H.-K. Sung, and A. Nagy, "piggyBac transposition reprograms fibroblasts to induced pluripotent stem cells," *Nature*, vol. 458, no. 7239, pp. 766–770, Mar. 2009.

[245] N. Fusaki, H. Ban, A. Nishiyama, K. Saeki, and M. Hasegawa, "Efficient induction of transgene-free human pluripotent stem cells using a vector based on Sendai virus, an RNA virus that does not integrate into the host genome.," *Proc. Jpn. Acad., Ser. B, Phys. Biol. Sci.*, vol. 85, no. 8, pp. 348–362, 2009.

[246] F. Jia, K. D. Wilson, N. Sun, D. M. Gupta, M. Huang, Z. Li, N. J. Panetta, Z. Y. Chen, R. C. Robbins, M. A. Kay, M. T. Longaker, and J. C. Wu, "A nonviral minicircle vector for deriving human iPS cells," *Nat. Methods*, vol. 7, no. 3, pp. 197–199, Feb. 2010.

[247] N. Montserrat, E. Garreta, F. González, J. Gutiérrez, C. Eguizábal, V. Ramos, S. Borrós, and J. C. Izpisua Belmonte, "Simple generation of human induced pluripotent stem cells using poly-beta-amino esters as the non-viral gene delivery system.," *J. Biol. Chem.*, vol. 286, no. 14, pp. 12417–12428, Apr. 2011.

[248] D. Kim, C.-H. Kim, J.-I. Moon, Y.-G. Chung, M.-Y. Chang, B.-S. Han, S. Ko, E. Yang, K. Y. Cha, R. Lanza, and K.-S. Kim, "Generation of human induced pluripotent stem cells by direct delivery of reprogramming proteins.," *Cell Stem Cell*, vol. 4, no. 6, pp. 472–476, Jun. 2009.

[249] K. Van Craenenbroeck, P. Vanhoenacker, and G. Haegeman, "Episomal vectors for gene expression in mammalian cells.," *Eur. J. Biochem.*, vol. 267, no. 18, pp. 5665–5678, Sep. 2000.

[250] O. Argyros, S. P. Wong, and R. P. Harbottle, "Non-viral episomal modification of cells using S/MAR elements.," *Expert Opin. Biol. Ther.*, vol. 11, no. 9, pp. 1177–1191, Sep. 2011.

[251] J. Yu, K. Hu, K. Smuga-Otto, S. Tian, R. Stewart, I. I. Slukvin, and J. A. Thomson, "Human induced pluripotent stem cells free of vector and transgene sequences.," *Science*, vol. 324, no. 5928, pp. 797–801, May 2009.

[252] K. Okita, T. Yamakawa, Y. Matsumura, Y. Sato, N. Amano, A. Watanabe, N. Goshima, and S. Yamanaka, "An efficient nonviral method to generate integration-free human-induced pluripotent stem cells from cord blood and peripheral blood cells.," *Stem Cells*, vol. 31, no. 3, pp. 458–466, Mar. 2013.

[253] G. Coppotelli, N. Mughal, S. Callegari, R. Sompallae, L. Caja, M. S. Luijsterburg, N. P. Dantuma, A. Moustakas, and M. G. Masucci, "The Epstein-Barr virus nuclear antigen-1 reprograms transcription by mimicry of high mobility group A proteins.," *Nucleic Acids Research*, vol. 41, no. 5, pp. 2950–2962, Mar. 2013.

[254] L.-M. Houdebine, "Methods to generate Transgenic Animals," in *Genetic Engineering in Livestock*, M. Engelhard, K. Hagen, and M. Boysen, Eds. 2009, pp. 1–19.

[255] S. Ding, X. Wu, G. Li, M. Han, Y. Zhuang, and T. Xu, "Efficient transposition of the piggyBac (PB) transposon in mammalian cells and mice.," *Cell*, vol. 122, no. 3, pp. 473–483, Aug. 2005.

[256] Z. Ivics, W. Garrels, L. Mátés, T. Y. Yau, S. Bashir, V. Zidek, V. Landa, A. Geurts, M. Pravenec, T. Rülicke, W. A. Kues, and Z. Izsvák, "Germline transgenesis in pigs by cytoplasmic microinjection of Sleeping Beauty transposons.," *Nat Protoc*, vol. 9, no. 4, pp. 810–827, Apr. 2014.

[257] A. Pfeifer, "Lentiviral transgenesis--a versatile tool for basic research and gene therapy.," *Curr Gene Ther*, vol. 6, no. 4, pp. 535–542, Aug. 2006.

[258] C. Lois, E. J. Hong, S. Pease, E. J. Brown, and D. Baltimore, "Germline transmission and tissue-specific expression of transgenes delivered by lentiviral vectors.," *Science*, vol. 295, no. 5556, pp. 868–872, Feb. 2002.

[259] A. Hofmann, B. Kessler, S. Ewerling, A. Kabermann, G. Brem, E. Wolf, and A. Pfeifer, "Epigenetic regulation of lentiviral transgene vectors in a large animal model.," *Mol Ther*, vol. 13, no. 1, pp. 59–66, Jan. 2006.

[260] M. R. Capecchi, "Gene targeting in mice: functional analysis of the mammalian genome for the twenty-first century.," *Nat. Rev. Genet.*, vol. 6, no. 6, pp. 507–512, Jun. 2005.

[261] S. Manzini, A. Vargiolu, I. M. Stehle, M. L. Bacci, M. G. Cerrito, R. Giovannoni, A. Zannoni, M. R. Bianco, M. Forni, P. Donini, M. Papa, H. J. Lipps, and M. Lavitrano, "Genetically modified pigs produced with a nonviral episomal vector.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 103, no. 47, pp. 17672–17677, Nov. 2006.

[262] C. E. Dunbar, K. A. High, J. K. Joung, D. B. Kohn, K. Ozawa, and M. Sadelain, "Gene therapy comes of age.," *Science*, vol. 359, no. 6372, p. eaan4672, Jan. 2018.

[263] National Institutes of Health, "Regenerative Medicine 2006," pp. 1–106, Aug. 2006.

[264] N. Savić and G. Schwank, "Advances in therapeutic CRISPR/Cas9 genome editing.," *Transl Res*, vol. 168, pp. 15–21, Feb. 2016.

[265] A. C. Nathwani, U. M. Reiss, E. G. D. Tuddenham, C. Rosales, P. Chowdary, J. McIntosh, M. Della Peruta, E. Lheriteau, N. Patel, D. Raj, A. Riddell, J. Pie, S. Rangarajan, D. Bevan, M. Recht, Y.-M. Shen, K. G. Halka, E. Basner-Tschakarjan, F. Mingozzi, K. A. High, J. Allay, M. A. Kay, C. Y. C. Ng, J. Zhou, M. Cancio, C. L. Morton, J. T. Gray, D. Srivastava, A. W. Nienhuis, and A. M. Davidoff, "Long-term safety and efficacy of factor IX gene therapy in hemophilia B.," *N. Engl. J. Med.*, vol. 371, no. 21, pp. 1994–2004, Nov. 2014.

[266] J. Bennett, J. Wellman, K. A. Marshall, S. McCague, M. Ashtari, J. DiStefano-Pappas, O. U. Elci, D. C. Chung, J. Sun, J. F. Wright, D. R. Cross, P. Aravand, L. L. Cyckowski, J. L. Bennicelli, F. Mingozzi, A. Auricchio, E. A. Pierce, J. Ruggiero, B. P. Leroy, F. Simonelli, K. A. High, and A. M. Maguire, "Safety and durability of effect of contralateral-eye administration of AAV2 gene therapy in patients with childhood-onset blindness caused by RPE65 mutations: a follow-on phase 1 trial.," *Lancet*, vol. 388, no. 10045, pp. 661–672, Aug. 2016.

[267] S. G. Jacobson, A. V. Cideciyan, A. J. Roman, A. Sumaroka, S. B. Schwartz, E. Heon, and W. W. Hauswirth, "Improvement and decline in vision with gene therapy in childhood blindness.," *N. Engl. J. Med.*, vol. 372, no. 20, pp. 1920–1926, May 2015.

[268] S.-I. Muramatsu, K.-I. Fujimoto, S. Kato, H. Mizukami, S. Asari, K. Ikeguchi, T. Kawakami, M. Urabe, A. Kume, T. Sato, E. Watanabe, K. Ozawa, and I. Nakano, "A phase I study of aromatic L-amino acid decarboxylase gene therapy for Parkinson's disease.," *Mol. Ther.*, vol. 18, no. 9, pp. 1731–1735, Sep. 2010.

[269] M. Körbling, "Peripheral Blood Stem Cells: A Novel Source for Allogeneic Transplantation.," *Oncologist*, vol. 2, no. 2, pp. 104–113, 1997.

[270] S. Ma, N. Xie, W. Li, B. Yuan, Y. Shi, and Y. Wang, "Immunobiology of mesenchymal stem cells.," *Cell Death Differ.*, vol. 21, no. 2, pp. 216–225, Feb. 2014.

[271] J. L. M. Ferrara, J. E. Levine, P. Reddy, and E. Holler, "Graft-versus-host disease.," *Lancet*, vol. 373, no. 9674, pp. 1550–1561, May 2009.

[272] M. Otsu, "Perspectives on stem cell gene therapy for genetic disorders ," *ISBT Science Series*, vol. 10, no. 1, pp. 231–234, Apr. 2015.

[273] N. J. Brown and M. L. Hirsch, "Adeno-associated virus (AAV) gene delivery in stem cell therapy.," *Discov Med*, vol. 20, no. 111, pp. 333–342, Nov. 2015.

[274] K. Rapti, F. Stillitano, I. Karakikes, M. Nonnenmacher, T. Weber, J.-S. Hulot, and R. J. Hajjar, "Effectiveness of gene delivery systems for pluripotent and differentiated cells.," *Molecular Therapy - Methods & Clinical Development*, vol. 2, p. 14067, 2015.

[275] L. U. W. Müller, M. D. Milsom, C. E. Harris, R. Vyas, K. M. Brumme, K. Parmar, L. A. Moreau, A. Schambach, I.-H. Park, W. B. London, K. Strait, T. Schlaeger, A. L. Devine, E. Grassman, A. D'Andrea, G. Q. Daley, and D. A. Williams, "Overcoming reprogramming resistance of Fanconi anemia cells.," *Blood*, vol. 119, no. 23, pp. 5449–5457, Jun. 2012.

[276] J. M. Liu, S. Kim, E. J. Read, M. Futaki, I. Dokal, C. S. Carter, S. F. Leitman, M. Pensiero, N. S. Young, and C. E. Walsh, "Engraftment of hematopoietic progenitor cells transduced with the Fanconi anemia group C gene (FANCC).," *Hum. Gene Ther.*, vol. 10, no. 14, pp. 2337–2346, Sep. 1999.

[277] A. M. Maguire, F. Simonelli, E. A. Pierce, E. N. Pugh, F. Mingozzi, J. Bennicelli, S. Banfi, K. A. Marshall, F. Testa, E. M. Surace, S. Rossi, A. Lyubarsky, V. R. Arruda, B. Konkle, E. Stone, J. Sun, J. Jacobs, L. Dell'Osso, R. Hertle, J.-X. Ma, T. M. Redmond, X. Zhu, B. Hauck, O. Zelenaia, K. S. Shindler, M. G. Maguire, J. F. Wright, N. J. Volpe, J. W. McDonnell, A. Auricchio, K. A. High, and J. Bennett, "Safety and efficacy of gene transfer for Leber's congenital amaurosis.," *N. Engl. J. Med.*, vol. 358, no. 21, pp. 2240–2248, May 2008.

[278] R. E. MacLaren, M. Groppe, A. R. Barnard, C. L. Cottriall, T. Tolmachova, L. Seymour, K. R. Clark, M. J. During, F. P. M. Cremers, G. C. M. Black, A. J. Lotery, S. M. Downes, A. R. Webster, and M. C. Seabra, "Retinal gene therapy in patients with choroideremia: initial findings from a phase 1/2 clinical trial.," *Lancet*, vol. 383, no. 9923, pp. 1129–1137, Mar. 2014.

[279] M. P. Simunovic, K. Xue, J. K. Jolly, and R. E. MacLaren, "Structural and Functional Recovery Following Limited Iatrogenic Macular Detachment for Retinal Gene Therapy.," *JAMA Ophthalmol*, vol. 135, no. 3, pp. 234–241, Mar. 2017.

[280] I. S. Dimopoulos, C. Tseng, and I. M. MacDonald, "Microperimetry as an Outcome Measure in Choroideremia Trials: Reproducibility and Beyond.," *Invest. Ophthalmol. Vis. Sci.*, vol. 57, no. 10, pp. 4151–4161, Aug. 2016.

[281] S. P. Wong and R. P. Harbottle, "Genetic modification of dividing cells using episomally maintained S/MAR DNA vectors.," *Mol Ther Nucleic Acids*, vol. 2, no. 8, p. e115, 2013.

[282] W. S. Pear, G. P. Nolan, M. L. Scott, and D. Baltimore, "Production of high-titer helper-free retroviruses by transient transfection.," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 90, no. 18, pp. 8392–8396, Jul. 1993.

[283] W. F. Scherer, J. T. Syverton, and G. O. Gey, "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 epidermioid carcinoma of the cervix," *J Exp Med*, vol. 97, no. 5, pp. 695–710, Jan. 1953.

[284] Y. Okita and K. I. Nakayama, "UPS Delivers Pluripotency," *Stem Cell*, vol. 11, no. 6, pp. 728–730, Dec. 2012.

[285] "Lentiviral Vector Design and Imaging Approaches to Visualize the Early Stages of Cellular Reprogramming," vol. 19, no. 4, pp. 782–789, Nov. 2009.

[286] S. CHANDRA, O. LEVRAN, I. JURICKOVA, C. MAAS, R. KAPUR, D. SCHINDLER, R. HENRY, K. MILTON, S. BATISH, and J. CANCELAS, "A Rapid Method for Retrovirus-Mediated Identification of Complementation Groups in Fanconi Anemia Patients," *Mol Ther*, vol. 12, no. 5, pp. 976–984, Nov. 2005.

[287] R. Zufferey, D. Nagy, R. J. Mandel, L. Naldini, and D. Trono, "Multiply attenuated lentiviral vector achieves efficient gene delivery in vivo.," *Nat Biotechnol*, vol. 15, no. 9, pp. 871–875, Sep. 1997.

[288] J. Schindelin, I. Arganda-Carreras, E. Frise, V. Kaynig, M. Longair, T. Pietzsch, S. Preibisch, C. Rueden, S. Saalfeld, B. Schmid, J.-Y. Tinevez, D. J. White, V. Hartenstein, K. Eliceiri, P. Tomancak, and A. Cardona, "Fiji: an open-source platform for biological-image analysis," *Nat. Methods*, vol. 9, no. 7, pp. 676–682, Jun. 2012.

[289] A. Untergasser, I. Cutcutache, T. Koressaar, J. Ye, B. C. Faircloth, M. Remm, and S. G. Rozen, "Primer3--new capabilities and interfaces," *Nucleic Acids Research*, vol. 40, no. 15, pp. e115–e115, Aug. 2012.

[290] T. Koressaar and M. Remm, "Enhancements and modifications of primer design program Primer3," *Bioinformatics*, vol. 23, no. 10, pp. 1289–1291, May 2007.

[291] G. Koutsoudakis, A. Kaul, E. Steinmann, S. Kallis, V. Lohmann, T. Pietschmann, and R. Bartenschlager, "Characterization of the early steps of hepatitis C virus infection by using luciferase reporter viruses.," *J. Virol.*, vol. 80, no. 11, pp. 5308–5320, Jun. 2006.

[292] O. Varisli, C. Agca, and Y. Agca, "Short-term storage of rat sperm in the presence of various extenders.," *J. Am. Assoc. Lab. Anim. Sci.*, vol. 52, no. 6, pp. 732–737, Nov. 2013.

[293] T. Garcia and M.-C. Hofmann, "Isolation of undifferentiated and early differentiating type A spermatogonia from Pou5f1-GFP reporter mice.," *Methods Mol. Biol.*, vol. 825, no. 3, pp. 31–44, 2012.

[294] M. Bozza, "The development of a novel S/MAR DNA vector platform for the stable, persistent and safe Genetic Engineering of Dividing Cells," Oct. 2017.

[295] O. Argyros, S. P. Wong, C. Fedonidis, O. Tolmachov, S. N. Waddington, S. J. Howe, M. Niceta, C. Coutelle, and R. P. Harbottle, "Development of S/MAR minicircles for enhanced and persistent transgene expression in the mouse liver," *J. Mol. Med.*, vol. 89, no. 5, pp. 515–529, Feb. 2011.

[296] J. M. Luke, A. E. Carnes, and J. A. Williams, "Development of antibiotic-free selection system for safer DNA vaccination.," *Methods Mol. Biol.*, vol. 1143, no. 6, pp. 91–111, 2014.

[297] J. Kuehle, S. Turan, T. Cantz, D. Hoffmann, J. D. Suerth, T. Maetzig, D. Zychlinski, C. Klein, D. Steinemann, C. Baum, J. Bode, and A. Schambach, "Modified Lentiviral LTRs Allow Flp Recombinase–mediated Cassette Exchange and In Vivo Tracing of 'Factor-free' Induced Pluripotent Stem Cells," *Mol Ther*, vol. 22, no. 5, pp. 919–928, May 2014.

[298] K. Okita, Y. Matsumura, Y. Sato, A. Okada, A. Morizane, S. Okamoto, H. Hong, M. Nakagawa, K. Tanabe, K.-I. Tezuka, T. Shibata, T. Kunisada, M. Takahashi, J. Takahashi, H. Saji, and S. Yamanaka, "A more efficient method to generate integration-free human iPS cells," *Nat. Methods*, vol. 8, no. 5, pp. 409–412, Apr. 2011.

[299] J. Hanna, K. Saha, B. Pando, J. van Zon, C. J. Lengner, M. P. Creyghton, A. van Oudenaarden, and R. Jaenisch, "Direct cell reprogramming is a stochastic process amenable to acceleration.," *Nature*, vol. 462, no. 7273, pp. 595–601, Dec. 2009.

[300] K. Hawkins, S. Joy, and T. McKay, "Cell signalling pathways underlying induced pluripotent stem cell reprogramming.," *World J Stem Cells*, vol. 6, no. 5, pp. 620–628, Nov. 2014.

[301] Y. Zhao, X. Yin, H. Qin, F. Zhu, H. Liu, W. Yang, Q. Zhang, C. Xiang, P. Hou, Z. Song, Y. Liu, J. Yong, P. Zhang, J. Cai, M. Liu, H. Li, Y. Li, X. Qu, K. Cui, W. Zhang, T. Xiang, Y. Wu, Y. Zhao, C. Liu, C. Yu, K. Yuan, J. Lou, M. Ding, and H. Deng, "Two Supporting Factors Greatly Improve the Efficiency of Human iPSC Generation," *Cell Stem Cell*, vol. 3, no. 5, pp. 475–479, Nov. 2008.

[302] C. Lancrin, P. Sroczynska, A. G. Serrano, A. Gandillet, C. Ferreras, V. Kouskoff, and G. Lacaud, "Blood cell generation from the hemangioblast.," *J. Mol. Med.*, vol. 88, no. 2, pp. 167–172, Feb. 2010.

[303] H. Kawasaki, I. Kosugi, Y. Arai, T. Iwashita, and Y. Tsutsui, "Mouse embryonic stem cells inhibit murine cytomegalovirus infection through a multi-step process.," *PLoS ONE*, vol. 6, no. 3, p. e17492, 2011.

[304] C. Hagedorn, M. N. Antoniou, and H. J. Lipps, "Genomic cis-acting Sequences Improve Expression and Establishment of a Nonviral Vector," *Mol Ther Nucleic Acids*, vol. 2, no. 8, p. e118, Sep. 2013.

[305] U. Muller-Kuller, M. Ackermann, S. Kolodziej, C. Brendel, J. Fritsch, N. Lachmann, H. Kunkel, J. Lausen, A. Schambach, T. Moritz, and M. Grez, "A minimal ubiquitous chromatin opening element (UCOE) effectively prevents silencing of juxtaposed heterologous promoters by epigenetic remodeling in multipotent and pluripotent stem cells," *Nucleic Acids Research*, vol. 43, no. 3, pp. 1577–1592, Feb. 2015.

[306] T. H. J. Kwaks, P. Barnett, W. Hemrika, T. Siersma, R. G. A. B. Sewalt, D. P. E. Satijn, J. F. Brons, R. van Blokland, P. Kwakman, A. L. Kruckeberg, A. Kelder, and A. P. Otte, "Identification of anti-repressor elements that confer high and stable protein production in mammalian cells," *Nat Biotechnol*, vol. 21, no. 5, pp. 553–558, Apr. 2003.

[307] F. Jia, K. D. Wilson, N. Sun, D. M. Gupta, M. Huang, Z. Li, N. J. Panetta, Z. Y. Chen, R. C. Robbins, M. A. Kay, M. T. Longaker, and J. C. Wu, "A nonviral minicircle vector for deriving human iPS cells.," *Nat. Methods*, vol. 7, no. 3, pp. 197–199, Mar. 2010.

[308] O. Argyros, S. P. Wong, C. Fedonidis, O. Tolmachov, S. N. Waddington, S. J. Howe, M. Niceta, C. Coutelle, and R. P. Harbottle, "Development of S/MAR minicircles for enhanced and persistent transgene expression in the mouse liver.," *J. Mol. Med.*, vol. 89, no. 5, pp. 515–529, May 2011.

[309] V. Gaspar, D. de Melo-Diogo, E. Costa, A. Moreira, J. Queiroz, C. Pichon, I. Correia, and F. Sousa, "Minicircle DNA vectors for gene therapy: advances and applications.," *Expert Opin. Biol. Ther.*, vol. 15, no. 3, pp. 353–379, Mar. 2015.

[310] "Extensive Methylation of Promoter Sequences Silences Lentiviral Transgene Expression During Stem Cell Differentiation In Vivo," vol. 20, no. 5, pp. 1014–1021, Nov. 2009.

[311] A. N. Alexopoulou, J. R. Couchman, and J. R. Whiteford, "The CMV early enhancer/chicken beta actin (CAG) promoter can be used to drive transgene expression during the differentiation of murine embryonic stem cells into vascular progenitors.," *BMC Cell Biol.*, vol. 9, no. 1, p. 2, 2008.

[312] S. Chung, T. Andersson, K.-C. Sonntag, L. Björklund, O. Isacson, and K.-S. Kim, "Analysis of different promoter systems for efficient transgene expression in mouse embryonic stem cell lines.," *Stem Cells*, vol. 20, no. 2, pp. 139–145, 2002.

[313] J. Y. Qin, L. Zhang, K. L. Clift, I. Hulur, A. P. Xiang, B.-Z. Ren, and B. T. Lahn, "Systematic comparison of constitutive promoters and the doxycycline-inducible promoter.," *PLoS ONE*, vol. 5, no. 5, p. e10611, May 2010.

[314] R. Wang, J. Liang, H. Jiang, L.-J. Qin, and H.-T. Yang, "Promoter-dependent EGFP expression during embryonic stem cell propagation and differentiation.," *Stem Cells Dev.*, vol. 17, no. 2, pp. 279–289, Apr. 2008.

[315] J. Qiao, P. Paul, S. Lee, L. Qiao, E. Josifi, J. R. Tiao, and D. H. Chung, "PI3K/AKT and ERK regulate retinoic acid-induced neuroblastoma cellular differentiation," *Biochemical and Biophysical Research Communications*, vol. 424, no. 3, pp. 421–426, Aug. 2012.

[316] J. Bastien and C. Rochette-Egly, "Nuclear retinoid receptors and the transcription of retinoid-target genes.," *Gene*, vol. 328, pp. 1–16, Mar. 2004.

[317] M. M. McGrane, "Vitamin A regulation of gene expression: molecular mechanism of a prototype gene.," *J. Nutr. Biochem.*, vol. 18, no. 8, pp. 497–508, Aug. 2007.

[318] M. Kishimoto, R. Fujiki, S. Takezawa, Y. Sasaki, T. Nakamura, K. Yamaoka, H. Kitagawa, and S. Kato, "Nuclear receptor mediated gene regulation through chromatin remodeling and histone modifications.," *Endocr. J.*, vol. 53, no. 2, pp. 157–172, Apr. 2006.

[319] F. Taieb, J.-P. Nougayrède, and E. Oswald, "Cycle inhibiting factors (cifs): cyclomodulins that usurp the ubiquitin-dependent degradation pathway of host cells.," *Toxins (Basel)*, vol. 3, no. 4, pp. 356–368, Apr. 2011.

[320] C. Khanna, J. J. Jaboin, E. Drakos, M. Tsokos, and C. J. Thiele, "Biologically relevant orthotopic neuroblastoma xenograft models: primary adrenal tumor growth and spontaneous distant metastasis.," *In Vivo*, vol. 16, no. 2, pp. 77–85, Mar. 2002.

[321] M. A. Hayat, *Neuroblastoma*, vol. 1. Dordrecht: Springer Science & Business Media, 2011.

[322] J. M. Polo and K. Hochedlinger, "When fibroblasts MET iPSCs.," *Cell Stem Cell*, vol. 7, no. 1, pp. 5–6, Jul. 2010.

[323] J. Utikal, J. M. Polo, M. Stadtfeld, N. Maherali, W. Kulalert, R. M. Walsh, A. Khalil, J. G. Rheinwald, and K. Hochedlinger, "Immortalization eliminates a roadblock during cellular reprogramming into iPS cells.," *Nature*, vol. 460, no. 7259, pp. 1145–1148, Aug. 2009.

[324] R. M. Marión, K. Strati, H. Li, M. Murga, R. Blanco, S. Ortega, O. Fernandez-Capetillo, M. Serrano, and M. A. Blasco, "A p53-mediated DNA damage response limits reprogramming to ensure iPS cell genomic integrity.," *Nature*, vol. 460, no. 7259, pp. 1149–1153, Aug. 2009.

[325] A. Canaan, I. Haviv, A. E. Urban, V. P. Schulz, S. Hartman, Z. Zhang, D. Palejev, A. B. Deisseroth, J. Lacy, M. Snyder, M. Gerstein, and S. M. Weissman, "EBNA1 regulates cellular gene expression by binding cellular promoters.,"

Proceedings of the National Academy of Sciences of the United States of America, vol. 106, no. 52, pp. 22421–22426, Dec. 2009.

[326] C. Robin, K. Ottersbach, J.-C. Boisset, A. Oziemlak, and E. Dzierzak, "CD41 is developmentally regulated and differentially expressed on mouse hematopoietic stem cells.," *Blood*, vol. 117, no. 19, pp. 5088–5091, May 2011.

[327] T. J. Fielder, L. Barrios, and L. Montoliu, "A survey to establish performance standards for the production of transgenic mice.," *Transgenic Res.*, vol. 19, no. 4, pp. 675–681, Aug. 2010.

[328] X. Li, X. Zhao, Y. Fang, X. Jiang, T. Duong, C. Fan, C. C. Huang, and S. R. Kain, "Generation of destabilized green fluorescent protein as a transcription reporter.," *J. Biol. Chem.*, vol. 273, no. 52, pp. 34970–34975, Dec. 1998.

[329] A. G. Evdokimov, M. E. Pokross, N. S. Egorov, A. G. Zaraisky, I. V. Yampolsky, E. M. Merzlyak, A. N. Shkoporov, I. Sander, K. A. Lukyanov, and D. M. Chudakov, "Structural basis for the fast maturation of Arthropoda green fluorescent protein.," *EMBO Rep.*, vol. 7, no. 10, pp. 1006–1012, Oct. 2006.

[330] U. Schagdarsurengin, A. Paradowska, and K. Steger, "Analysing the sperm epigenome: roles in early embryogenesis and assisted reproduction.," *Nat Rev Urol*, vol. 9, no. 11, pp. 609–619, Nov. 2012.

[331] G. F. Clark and D. J. Schust, "Manifestations of immune tolerance in the human female reproductive tract.," *Front Immunol*, vol. 4, p. 26, 2013.

[332] F. Marchetti, J. Bishop, J. Gingerich, and A. J. Wyrobek, "Meiotic interstrand DNA damage escapes paternal repair and causes chromosomal aberrations in the zygote by maternal misrepair.," *Sci. Rep.*, vol. 5, no. 1, p. 7689, Jan. 2015.

[333] Y. Ménézo, B. Dale, and M. Cohen, "DNA damage and repair in human oocytes and embryos: a review.," *Zygote*, vol. 18, no. 4, pp. 357–365, Nov. 2010.

[334] K. Canté-Barrett, R. D. Mendes, W. K. Smits, Y. M. van Helsdingen-van Wijk, R. Pieters, and J. P. P. Meijerink, "Lentiviral gene transfer into human and murine hematopoietic stem cells: size matters.," *BMC Res Notes*, vol. 9, no. 1, p. 312, Jun. 2016.

[335] M. Mandai, A. Watanabe, Y. Kurimoto, Y. Hirami, C. Morinaga, T. Daimon, M. Fujihara, H. Akimaru, N. Sakai, Y. Shibata, M. Terada, Y. Nomiya, S. Tanishima, M. Nakamura, H. Kamao, S. Sugita, A. Onishi, T. Ito, K. Fujita, S. Kawamata, M. J. Go, C. Shinohara, K.-I. Hata, M. Sawada, M. Yamamoto, S. Ohta, Y. Ohara, K. Yoshida, J. Kuwahara, Y. Kitano, N. Amano, M. Umekage, F. Kitaoka, A. Tanaka, C. Okada, N. Takasu, S. Ogawa, S. Yamanaka, and M. Takahashi, "Autologous Induced Stem-Cell–Derived Retinal Cells for Macular Degeneration," *http://dx.doi.org/10.1056/NEJMoa1608368*, vol. 376, no. 11, pp. 1038–1046, Mar. 2017.

[336] M. Nakagawa, M. Koyanagi, K. Tanabe, K. Takahashi, T. Ichisaka, T. Aoi, K. Okita, Y. Mochiduki, N. Takizawa, and S. Yamanaka, "Generation of induced pluripotent stem cells without Myc from mouse and human fibroblasts.," *Nat Biotechnol*, vol. 26, no. 1, pp. 101–106, Jan. 2008.

[337] J. Wu, A. Platero-Luengo, M. Sakurai, A. Sugawara, M. A. Gil, T. Yamauchi, K. Suzuki, Y. S. Bogliotti, C. Cuello, M. M. Valencia, D. Okumura, J. Luo, M. Vilariño, I. Parrilla, D. A. Soto, C. A. Martinez, T. Hishida, S. Sánchez-Bautista, M. L. Martinez-Martinez, H. Wang, A. Nohalez, E. Aizawa, P. Martinez-Redondo, A. Ocampo, P. Reddy, J. Roca, E. A. Maga, C. R. Esteban, W. T. Berggren, E. N. Delicado, J. Lajara, I. Guillen, P. Guillen, J. M. Campistol, E. A. Martinez, P. J. Ross, and J. C. I. Belmonte, "Interspecies Chimerism with Mammalian Pluripotent Stem Cells," *Cell*, vol. 168, no. 3, pp. 473–486.e15, Jan. 2017.