Dissertation

Submitted to the

Combined Faculties for the Natural Science and for Mathematics

of the Ruperto-Carola University of Heidelberg, Germany

for the degree of

Doctor of Natural Sciences

presented by

Matthias Bozza

Born in Bolzano, Italy

Date of examination: 06-12-2017

The development of a novel S/MAR DNA vector platform for the stable, persistent and safe Genetic Engineering of Dividing Cells

Referees: Prof. Dr. Peter Angel

Prof. Dr. Rienk Offringa

Declaration according to § 8 (3) b) and c) of the doctoral degree regulations:

b) I hereby declare that I have written the submitted dissertation myself and in this process have used no other sources or materials than those expressly indicated,

c) I hereby declare that I have not applied to be examined at any other institution, nor have I used the dissertation in this or any other form at any other institution as an examination paper, nor submitted it to any other faculty as a dissertation.

Heidelberg, October 9th, 2017

ABSTRACT

The development of episomally maintained DNA vectors that are capable of providing safe, persistent and stable modification of cells whilst avoiding the risk of integration-mediated genotoxicity would provide a valuable tool for genetic research.

DNA vectors harboring a Scaffold/Matrix Attachment Region (S/MAR) can provide persistent and robust transgene expression in human cancer cell lines which can be used in in vitro, in vivo and ex vivo studies. A prototype S/MAR DNA vector with which we initiated this study replicates episomally, remains unsilenced and unmethylated following the genetic modification of cells. Although, it showed great promise it does have significant limitations which restricts its application. The establishment rate of the original DNA vector is an inefficient passive process and the selection procedure is lengthy and often produces drug-resistant but non-expressing colonies generating a mosaic of cells with differential transgene expression. Thus, these vectors represent a reasonable tool for simple studies such as cell labelling with reporter genes but are not suitable for more sophisticated work such as gene-rescue experiments or for the genetic engineering of primary human cells. In this project, we have refined and enhanced the S/MAR DNA vector system. The range of next-generation DNA vectors that we have produced provide several advances over the original vectors. We have demonstrated that this new S/MAR DNA vector platform is more efficient and stable with improved efficiency in establishing stable cell-lines. We have also demonstrated that the persistence of transgene expression and the molecular integrity of the vector has been improved in a range of cancer cell lines as well as in primary human cells.

We have used this next-generation of DNA vector to generate labelled cells suitable for *in vitro* and *in vivo* drug screening. We have also generated isogenic tumor models which provide insights into the mechanism of pancreatic cancer development by restoring crucial tumor suppressor genes to the cells without altering the molecular or biochemical integrity of the cells. Additionally, we have

utilized the vector system to persistently modify a range of dividing cell types including primary mouse embryonic stem cells and embryonic fibroblasts and primary human fibroblasts.

As an ultimate demonstration of the efficacy of this DNA vector we have used it to genetically modify human T-cells for immunotherapy and have demonstrated it to be capable of expressing transgenes in these cells for over 1 month with minimal toxicity.

We have demonstrated that this novel class of DNA vector can be used to persistently modify every cell-type tested providing sustained and high levels of transgene expression whilst avoiding the risk of insertional mutagenesis induced by the random integration of genetic material with minimal impact to the cell.

ZUSAMMENFASSUNG

Die Entwicklung von episomal aufrecht erhaltenen DNA Vektoren, würde ein brauchbares Werkzeug für die Forschung liefern. Während sie das Risiko von Genotoxizität vermeiden, sind diese in der Lage, sichere, dauerhafte und stabile Modifizierung von Zellen zu ermöglichen.
DNA Vektoren mit einer Scaffold/Matrix Attachment Region (S/MAR) können stabile und robuste transgen Expression in humanen Krebszelllinien liefern, welche daraufhin in *in vitro*, *in vivo* und *ex vivo* Studien verwendet werden können. Zu Beginn dieser Studie arbeiteten wir mit einem Prototyp S/MAR DNA Vektor, welcher episomal replizierend ist und nach stabiler Modifizierung von Zellen unsilenced und unmethyliert bleibt. Trotz seiner vielversprechenden Eigenschaften, zeigten sich bald Limitierungen in seinen Anwendungsmöglichkeiten. Zum einen ist die Etablierungsrate dieses originalen DNA Vektors ist ein ineffizienter, passiver Prozess. Außerdem ist die Selektion zeitaufwändig und liefert oft resistente, aber nicht exprimierende Kolonien, was in einem Mosaik aus Zellen mit verschiedenster Transgen Expression resultiert. Dementsprechend stellen diese Vektoren ein angemessenes Werkzeug für einfache Studien wie Zellmarkierung mit Reporter Genen dar. Für anspruchsvollere Arbeiten wie Gen-Rescue Experimente oder für die genetische Modifizierung von humanen Primärzellen sind sie allerdings weniger geeignet.
In diesem Projekt wurde daraufhin das vorliegende S/MAR DNA Vektorsystem verfeinert und optimiert. Das Sortiment an von uns hergestellten next-generation DNA Vektoren liefern verschiedenste Vorteile gegenüber dem ursprünglichen System. Wir konnten beweisen, dass diese neue S/MAR DNA Vektor Plattform effizienter und stabiler ist, und zusätzlich verbesserte Effizienz in der Etablierung von stabilen Zelllinien aufweist. Des weiteren wurde die Aufrechterhaltung der Transgen Expression und die molekulare Integrität der Vektoren in vielen humanen Krebs- und Primärzelllinien verbessert.

Zusammenfassung

Wir verwendeten diese next-generation DNA Vektoren um markierte Zellen, die für *in vitro* und *in vivo* Medikament screenings geeignet sind, herzustellen. Außerdem haben wir isogenetische Tumor Modelle generiert, in denen wir, durch die Wiederherstellung der Funktionalität von Tumor Suppressor Genen, die Entwicklung von Bauchspeicheldrüsenkrebs studierten. Die Vektoren wurden auch erfolgreich verwendet um verschiedenste sich teilende Zellen wie primäre embryonale Maus Fibroblasten und Stammzellen, sowie primäre humane Fibroblasten dauerhaft zu verändern. Als wichtigste Veranschaulichung der Leistungsfähigkeit dieser Vektoren wurden sie für die genetische Modifizierung von humanen T-Zellen für die Immuntherapie verwendet. Hierbei konnten wir die Expression der Transgene für eine Dauer von über einem Monat mit minimaler Toxizität demonstrieren.

Wir zeigten auf, dass unsere neue Klasse an DNA Vektoren verwendet werden kann, um dauerhaft jegliche untersuchten Zelltypen zu verändern. Dabei liefern sie nachhaltige hohe Transgen Expression mit minimalem Einfluss auf die Zelle, ohne das Risiko willkürlicher Integration des genetischen Materials.

ACKNOWLEDGEMENTS

This work would have not been possible without the help and the support of the people around me.

The biggest heartfelt thanks goes, undoubtedly to Elena with whom I have the privilege to share my life. Thank you, for your extensive support and patience in every single moment.

Thanks to my family that although far away is always present. Thank you for all the times you put lots of effort in taking a trip to Heidelberg just for a short week end. Thank you to my parents and my sister that, just to spend a couple of days together made very exhausting trips. I can't forget to mention the daily incredible support of my parents-in-law, I needed to never give up.

I have to thank my supervisor Richard for giving me the opportunity of being part of his staff. Thank you for the extensive freedom and support you gave me for this project and more importantly thank you for your friendship and help in the last years.

Thank you Alica (ups, Ali) for sharing the good and the bad moments in the lab. You are undoubtedly the person everybody would like to work with. I don't like to just call you my colleague because, more importantly, you are my friend. Thanks for taking care of the organization of all the events and for the great time we spent outside the lab. It will always stay in my memories our first conference in Ulm, the EMBL, and several others, parties and the great trip to New York-Washington.

Thank you and a big "good luck" goes to Tamara. I really loved our philosophical conversations that were surely one of the highest moments during our lab life.

Acknowledgements

Thank you to all the lab members, it has been a great pleasure to work with you: Matthias for the incredible help, Sveta, Daniel, Rita and Manuela (I'm sure that the lab gained a great new member).

I would like to thank very, very, very much Margareta. Thanks for sharing your expertise, for guiding me in the complicated world of immunology, for all the time you have spent for moving this work forward. Thanks for the extensive time you have dedicated to me and more importantly thanks for the friendship that we built up during the past months.

I have to thank Daniele. It is your "fault" that I ended up in a crazy, little, empty lab. I followed your advice and I have never regretted it.

From the first moment I came to Heidelberg I felt very welcome and part of a big family. I have to thank for the great moments we spent together over these years: Antonio, Sara, Silvia, Ivan, Gian Mario, Edoardo, Daniele and Francesco.

Mattia for the dinners, coffee breaks and inspiring conversations.

Elisa for guiding and inspiring some of this work investing lots of time, thanks.

Jamie and Joschka for being great friends not just colleagues from next-door.

Ed for the countless coffees and inspiring scientific chats.

Last but not least I can't forget to thank the two "ragazzi". Romolo and Lorenzo you adopted me when Elena was not here. Always present to spend some leisure time together either with a home-made pasta or in the pub. The countless number of beers we had together, ushered us in the funniest and undoubtedly most fun moments of the last years.

PUBLICATIONS AND AWARDS

Parts of this thesis have been presented in:

Conferences and workshop presentations:

ATV Retreat, Oral Presentation, 2015 Kloster Schöntal, Germany

AACR, Poster Presentation, 2015 Philadelphia, USA

ATV Retreat, Oral Presentation, 2016 Kloster Schöntal, Germany

AACR, Poster Presentation, 2016 New Orleans, USA

ASGCT, Oral Presentation, 2017 Washington, USA

PhD Retreat, Oral Presentation, 2017 Weil der Stadt, Germany

Science Spark for Start-Up Symposium, Oral Presentation, 2017 Heidelberg Germany

Research Lounge, Oral Presentation, 2017, DKFZ Heidelberg.

Within the thesis, results from the following publications were included:

Non integrative DNA vector for the genetic modification of cells, September 2017
Patent application number EP17191829.5

Bunse et al. "Suppression of antitumor T cell immunity by the oncometabolite R-2-hydroxyglutarate", *Nature, in revision*

Viarisio et al. "Beta HPV38 oncoproteins act with a hit and run mechanism in UV-induced skin carcinogenesis" *Plos Pathogens* , *in revision*

Bozza, et al. "S/MAR Nano Vectors – a novel genetic tool for the genetic engineering of Patient Derived Pancreatic Cancer Cells", *in preparation*

Awards

Science Spark for Start-Up-Dragon Den Competition, Prize Winner (Best Idea)

CONTENTS

LIST OF FIGURES

List of Figures

LIST OF TABLES

ABBREVIATIONS

Table 1. List of abbreviation

DNA	Deoxyribonucleic acid
RNA	Ribonucleic acid
ssRNA	Single strand RNA
HIV	Human immunodeficiency virus
AIDS	Acquired immunodeficiency syndrome
CRISPR	Cluster regulatory interspaced short palindromic repeats
CAS9	Crispr associated protein 9
GOI	Gene of interest
POI	Protein of interest
RT	Reverse transcription
SB	Sleeping Beauty
AT	Adenine Thymidine
gRNA	guide RNA
mRNA	messenger RNA
S/MAR	Scaffold / Matrix attachment region
Ori	Origin of replication
bp	Base pair
lac	lactose
trp	tryptophan
RBS	Ribosome binding site
AUG	Adenine Uracil Guanine
SV40	Simian vacuolating virus 40
cDNA	complementary DNA
BKV	Human polyomavirus 1
BPV	Bovine papilloma virus
EBV	Epstein bar virus
CG	Cytosine Guanine
TLR	Tall like receptor
pDNA	plasmid DNA
RNAi	interfering RNA

Abbreviations

ARS	Autonomous replicating sequence
CHO	Chinese hamster ovarian
AAT	α1 anti trypsin
SAF-1	Scaffold attachment factor 1
HPLC	High Performance Liquid Chromatography
Kan	Kanamycin
Puro	Puromycin
UCOE	Ubiquitous Chromatin Opening Element
HS4	Chicken β-globin Insulator
FMDV	Foot and mouth disease virus
IRES	Internal Ribosome entry sequence
FACS	Fluorophore activated cell sorting
HGD	Homogentisate 1-2-dioxygenase
HGA	Homogentisic acid
BLI	Bio luminescence imaging
SC	Single cell
MP	Mix population
NTX	Nature technology corporation
NP	Nano plasmid
FDA	Federal drug administration
GSEA	Gene set enrichment analysis
EMT	Epithelial to mesenchymal
ALL	Acute Lymphocytic Leukemia
FDR	False discovery rate
PBMC	Peripheral Blood Mononucleated Cells

1 Introduction

1.1 Current methods used for the Genetic Engineering of dividing cells

The possibility of persistently modifying mammalian cells represents an attractive field for molecular and cell biologists. A system that provides efficient, persistent and stable expression, or suppression, of genes in mammalian cells may represent a useful tool for a variety of studies, such as gene regulation, disease modelling, drug testing and gene supplementation for therapeutic correction. In the past decade many techniques were developed with the aim of introducing foreign DNA into mammalian cells. Among several different approaches, the three most popular and efficient ways to introduce genetic material into cells are represented by the Lentiviral Gene Delivery system (Figure 1, a), the Sleeping Beauty Transposon (Figure 1, b) and the CRISPR/CAS9 method (Figure 1, c).

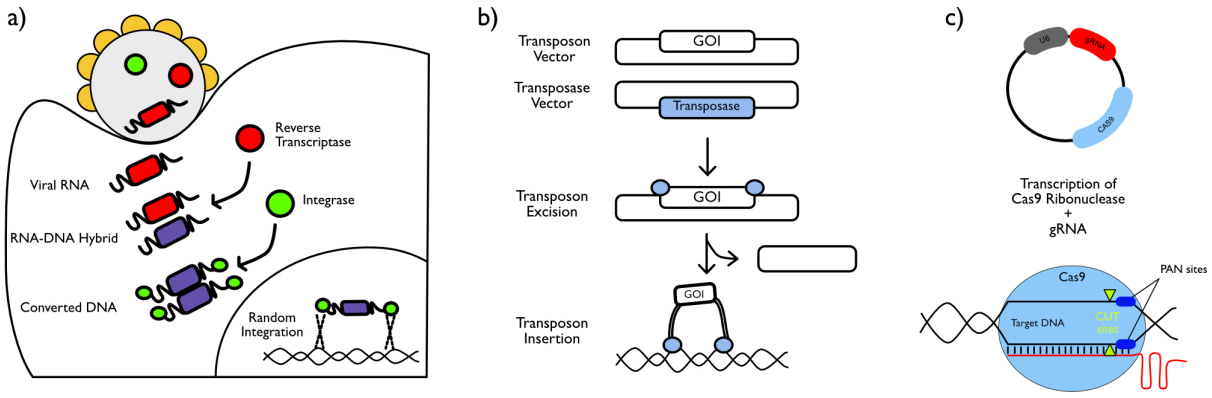


Figure 1. **Current systems used for gene expression and genome editing of mammalian cells.**
(a) In the Lentiviral Gene Delivery system the gene of interest (GOI) is introduced into the genome of a lentivirus and its translocation into the nucleus of the target cells is mediated by the virus itself. (b) Two plasmids DNA are necessary to deliver DNA through the Sleeping Beauty (SB) transposon technique. The first plasmid contains the expression cassette of the GOI and the second generate the enzyme transpose that is responsible for the integration of the GOI into the genome of the target cell. (c) In the CRISPR/CAS9 system, the guide RNA (gRNA) recruits the ribonuclease Cas9 on its homologous region on the cell genome. This complex formation mediates the modification of the cellular DNA at the recognised genomic locus.

1.1.1 The lentiviral system for the genetic modification of mammalian cells

In the lentiviral gene delivery system the Gene of Interest (GOI) is introduced into a viral genome and the virus mediates the delivery of the genetic material into the nucleus of the target cell (Figure 1, a). This technique is based on attenuated Lentivirus and in research these viruses are used primarily as a tool to introduce a gene product in cells or animal models. Lentivirus belong to the family of the *Retroviridae*, group VI, single-stranded RNA that require reverse transcription (ssRNA-RT). They are known to cause chronic and deadly diseases in humans characterised by long incubation periods. The most well studied lentivirus is the Human Immunodeficiency Virus (HIV). It represents the infectious agent that causes the severe human disease known as Acquired Immunodeficiency Syndrome (AIDS) (Sepkowitz 2001).

The genome of this class of virus comprises a ssRNA that, once transferred into the cytoplasm of the target cells is recognised by the cellular transcription machinery and translated into proteins. Complexed with the viral genome, a viral particle carries an essential protein: an RNA dependent DNA-Polymerase. This protein converts the viral RNA genome into DNA (Figure 1, a). Once produced, the DNA is translocated into the nucleus where viral proteins mediate its integration into the cellular genome. The viral genome is incorporated into the cellular genome and its transcription (together with the transcription of the GOI) and translation exploits cellular machinery.

1.1.2 Sleeping Beauty Transposons: an integrating non viral system for the engineering of mammalian cells

The Sleeping Beauty Transposon (SB) is a system composed of two expression units. One mediates the expression of the Sleeping Beauty Transposase, the other carries the GOI. The transposition is

defined "as a precise process in which a DNA fragment is excised from one DNA molecule and moved to another site in the same or different DNA molecule or genome" (Plasterk 1993). Once the two expression systems reach the nucleus of the target cells the insertion site can be anywhere in the nucleus of the cell. The enzyme Transposase mediates the insertion of the transposon in AT-rich sequences. Like for viral DNA, once integrated, in the SB system the transcription and translation of the GOI exploits cellular machinery. Arguably, the most exciting potential application of SBs is for human gene therapy. This system was used to generate *ex vivo* genetically engineered T cells and tested in clinical trials for patients at risk of death from advanced malignancies (Hackett, Largaespada et al. 2010).

1.1.3 CRISPR/Cas9 genome editing system

The third genome editing system is based on the Cluster Regulatory Interspaced Short Palindromic Repeats (CRISPR). It was discovered in bacteria and where it plays a key role in their defence system (Horvath and Barrangou 2010). In research it forms the basis of the genome editing technology known as CRISPR/Cas9. By delivering the Cas9 nuclease complexed with a synthetic guide RNA (gRNA) into a cell (Figure 1, c), the cell's genome can be cut at a desired location, allowing existing genes to be removed and/or new ones added. Often a plasmid containing the main components of the system is transfected into a cell, however hard to transfect cells such as stem cells or neurones require more efficient delivery systems. For these cells the components of CRISPR/Cas9 can be packaged into lenti-, adeno- or adeno-associated virus. A clinical trial based on this system started in 2016 in China. The aim of the study is to edit T cells of patient carrying a stage five non-small cell lung cancer. The Programmed cell death protein 1 (PDCD1) gene is knocked out by CRISPR in the T cells of the patient and the cells are expanded *ex vivo* before their infusion into the patient ("*PD-1 knockout engineered t cells for metastatic non-small cell lung cancer.* www.clinicaltrials .gov Retrieved 2017-08-07).

1.2 Limitations of the current systems used for persistently modifying dividing cells

Despite the efficiency of these systems in generating persistently modified mammalian genomes, they also present several limitations and drawbacks. The random integration of the viral genome can generate genotoxic effects such as *cis* activation or suppression of onco- and tumour suppressor genes which may lead to tumour formation, as well as *trans* activation of endogenous genes due to their interaction with viral proteins. Viral DNA is more prone to be recognised by the cellular RNA-DNA sensing system (Atianand and Fitzgerald 2013). It was reported that nuclear and cytosolic proteins can bind "non-self" DNA or RNA and trigger the innate immunity of cells which leads to the epigenetic silencing of the foreign DNA, induced cell death and the activation of the adaptive immunity when the DNA is delivered into an immunocompetent organism (Tao, Zhou et al. 2016). Although considered a "safer" gene integration system, in the Sleeping Beauty Transposon system, the enzyme Transposase drives the integration of the expression cassette into regions of the genome that are rich in A-T content (Hackett, Aronovich et al. 2011). Integration can occur anywhere and it has been calculated that a mammalian genome has approximately 200 million TA sites. AT rich DNA regions of the nucleus are not associated with coding sequences but they normally map as regulatory elements and often represent the binding site for structural proteins. The efficiency of the CRISPR/CAS9 in the editing of single base pairs (Mali, Yang et al. 2013) is well documented. It represents a useful tool for generating persistent knockdowns as well as for repairing the functionality of genes that present point mutations. Despite its efficiency in modifying single nucleotides, the system becomes particularly inefficient when it comes to the re-introduction of lost fragments of DNA. The chance of having two recombination events at the same genomic locus is very low. The guide RNA (gRNA), then, can sometimes bind alternative complementary regions of the genome, generating unexpected modification of those loci as well. The off target

effects, represent a major problem in the application of this system especially when it comes to the modification of primary human cells with the possibility of using them for clinical application.

1.3 An ideal, non viral, non integrative system for the safe, persistent and stable modification of dividing cells

To overcome the limitation of these methods an ideal system suitable for the genetic engineering of mammalian cells would be represented by a non-integrative DNA Vector. Once delivered to the nucleus of a target cell, the ideal vector should be able to express a gene of interest stably and persistently for the lifetime of the cell. The ideal vector system should overcome all the genotoxic problems induced by the integration of foreign DNA and/or off-targeting. Unlike some viruses, DNA Vectors shouldn't damage or kill the host cells after they have replicated. They should remain episomal and they shouldn't integrate into the cellular genome reducing the possibility of *cis*-transactivation or suppression of onco- and tumour- suppressor genes. Also, as it doesn't have to undergo a packaging process, the vector system would have potentially an unlimited capacity and to be functional it would not need the co-delivery of additional biological macromolecules.

1.4 Plasmid DNA: where do they come from?

In 1952, Joshua Lederberg defined a plasmid as "*any bacterial genetic material that exists in an extrachromosomal state for at least part of its replication cycle"* (Lederberg 1952). To distinguish a plasmid from a virus, the definition was later narrowed to "*genetic elements that exist exclusively or predominantly outside of the chromosome and can replicate autonomously"* (Verma and Somia 1997). Plasmids contain sequences that can recruit a particular set of proteins which initiate and mediated their independent replication. In nature, plasmids provide a mechanism for gene transfer within a population of microbes and typically they provide a selective advantage under a given environmental

state. They may carry genes that provides resistance to naturally occurring antibiotics, or allow the organism to utilise particular organic compounds that would be advantageous when nutrients are scarce (Hayes 2003).

1.5 Plasmids as a vector system for expression and transfer of genetic information

Artificially constructed plasmids are used as vectors in genetic engineering. These plasmids are commonly used to clone and amplify particular genes. The vector itself is generally a DNA sequence that consists of a gene of interest (GOI) which normally is represented by a transgene and a larger sequence that serves as the "backbone" of the vector. The backbone normally contains the Origin of Replication (Ori) and a selective marker, normally, an antibiotic resistance, that allows the selection of the desired molecule in bacteria. The purpose of a vector which transfers genetic information to another cell is typically to express the GOI, or insert, in the target cell. The manipulation of DNA vectors is normally conducted in the bacterial strain *E.Coli*. Depending on their application DNA plasmids can be divided into different categories:

- expression vectors
- transcription vectors
- shuttle vectors

Expression vectors produce protein through the transcription of the vector's insert followed by translation of the produced mRNA. Different features characterise vectors designed for the expression of a GOI in prokaryotic or eukaryotic cells. Normally, plasmid DNA designed for the expression in bacteria utilise a strong promoter like the *lac (lactose)* or the *trp (tryptophan)* operon and a *Ribosome Binding Site (RBS)* including a *Shine-Delgarno sequence* (a translation initiation site 8 base pairs upstream the AUG start codon) that follows the promoter and sustain the efficient

translation of the protein of interest (POI). A Eukaryotic expression vector, instead, requires different sequences. The promoter has to have a eukaryotic origin because it must be recognised by the eukaryotic cell transcription machinery. At the 3' end of the transcribed pre-mRNA is required the presence of a Poly-adenylation tail that stabilises the transcribed RNA. The Kozak consensus sequence plays a major role in the initiation of the translation process mediating the assembl of the ribosome on the mature RNA. The introduction of this sequence in an expression vector results in a more efficient and stable protein production.

The Kozak sequence is identified by the notation gccRccAUGG which is derived from a wide variety of eukaryotic species (Kozak 1987). Some other vectors are designed for transcription only and they represent a useful tool for the *in vitro* production of mRNA. They lack the sequences necessary for poly-adenylation and termination and therefore cannot be used for protein production. RNA is normally generated in bacterial cells and therefore in these vectors the transcription is driven by a strong bacterial promoter which can be easily recognise by the endogenous bacterial RNA-Polymerase.

A shuttle vector is a construct that can be propagated in two different host species. The main advantage of these vectors is that they can be manipulated in *E.Coli* and then used in a system which is more difficult to handle (*e.g.* yeast or mammalian cells). A shuttle vector that is going to be used for the expression of genetic material in mammalian cells needs to have some specific features. Normally it has two compartments (Figure 2). The bacterial backbone, contains the Ori and the antibiotic resistance driven by a bacterial promoter, allows the modification, propagation and expansion in bacterial cells. Separate from this compartment there is the eukaryotic expression cassette, that presents all the features necessary for the transcription and translation in eukaryotic cells previously described for an expression vector. These vectors are the basis for the generation of a class of plasmids extensively used for the expression of genetic material in cells from different species. The bacterial backbone is normally conserved, but according to the organism or species,

the expression cassette can be optimised for the transgene expression. Different organisms and species and even tissues from the same organism use different codon reading and promoters. Shuttle vectors have the big advantage of being flexible, they can be propagated easily in bacteria and then modified *ad hoc* for expression in eukaryotic cells.

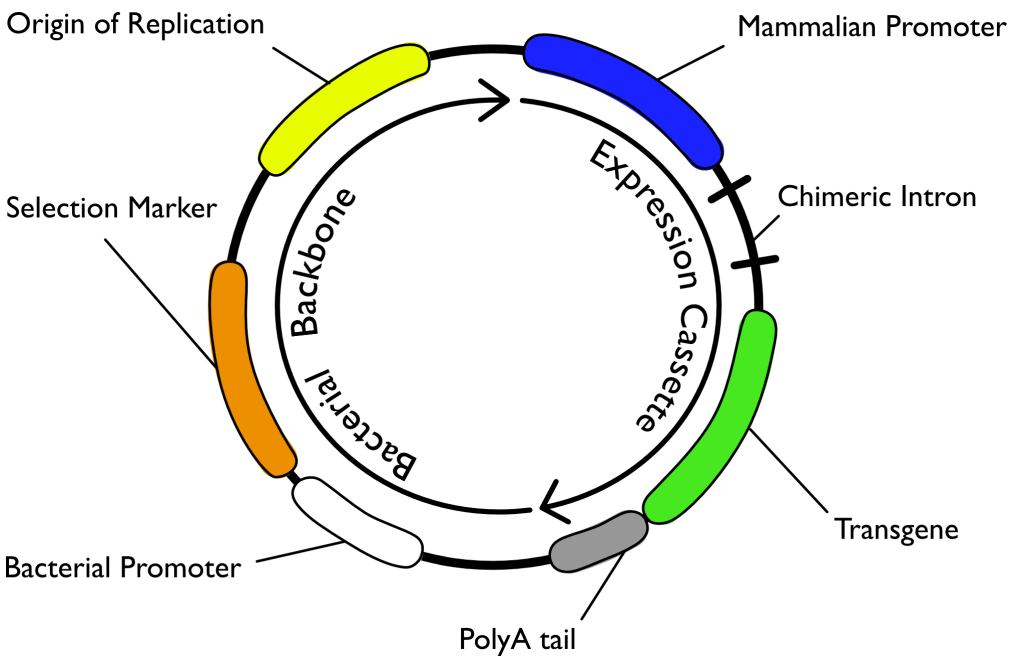


Figure 2. **Schematic representation of a shuttle vector.**
The bacterial backbone contains the bacterial origin of replication and the selection marker. It allows the manipulation and amplification in bacterial cells. The expression cassette which comprises a promoter, a chimeric intron sequence, the transgene of interest and a poly-adenylation signal allows the expression of a gene of interest in eukaryotic cells.

1.6 DNA plasmids for gene expression in mammalian cells

The development of eukaryotic expression vectors has provided a direct and convenient way of introducing novel genetic information into mammalian cells. The over-expression of recombinant genes has found widespread use in the production of therapeutically important proteins. For example, the treatment of the genetic disease haemophilia A requires repeated infusions of the

coagulation factor VIII (FVIII). In the 1970s and 1980s, the protein used for in the treatments was isolated and concentrated from the plasma of healthy donors. This method of isolation, although highly efficacious, carried a significant risk of the transmission of pathogens. The development of recombinant DNA technologies and the generation of vectors able to sustain the expression of transgenes in mammalian cells revolutionized the way of producing this essential coagulation factor. The protein is now produced and purified from mammalian cells in culture (Chen, Fang et al. 1999). The generated FVIII has a higher safety profile as well as a lower cost of production.

The efficiency of gene transfers as well as the improvement in the vector expression cassette found applications in all fields of cellular biology and this tool became essential for studies of gene regulation and protein-protein interaction. When transferred into mammalian cells in culture DNA plasmids are typically able to sustain transgene expression for only a limited period of time, normally two to three days, before they are slowly diluted and lost from the cell population.

If the aim of a study is to investigate the function of a gene over time, then stable and persistent expression becomes necessary. This is typically achieved only by taking advantage of one of the three systems previously described (Figure 1). However, long term gene expression can also be achieved, upon delivery of plasmid DNA utilising particular protocols. DNA plasmids cannot be maintained episomal in the nucleus of eukaryotic cells without the treatment of the transfected cells shortly after DNA delivery with high doses of antibiotics, which force the vector to be randomly integrated into their genome. The maintenance of selection pressure for some days helps also in the selection of those cells in which the integration successfully took place in a region of the genome that allows the expression of the integrated vector. The side effects induced by the integration of DNA Vectors are equivalent to those previously described for the lentiviral and the SB system.

1.6.1 Plasmid DNA (pDNA) vector mediated transgene expression

Typically, pDNA transgene expression in mammalian cells is transient. It reaches its maximum expression 24 - 48 h after administration of the DNA into the target cells. It drops slowly and consistently over subsequent days and rarely persists beyond one week. Following gene delivery there are several reasons for the loss of expression. It may be due to a co-occurrence of epigenetic events and immune responses. pDNA itself can elicit immunostimulatory responses (Mansur, Smith et al. 2014). One of the characteristics which distinguishes bacterial from mammalian DNA is the methylation of the CpG motifs. Normally these DNA sequences are methylated in eukaryotic genomic DNA and appear unmodified in DNA which has a prokaryotic origin (Li, Hursting et al. 2003).

The conspicuous presence of unmethylated CpGs is detected in mammalian cells by a family of protein receptors known as Toll-like receptors (TLRs), in particular the TLR-9 (Hemmi, Takeuchi et al. 2000). This class of protein is strongly expressed in tissues involved in immune function, such as the spleen, peripheral blood leukocytes and tissue exposed to the external environment but they are expressed constitutively in almost the cell types. It was shown that in lungs TLR-9 activate transcription factors such as AP-1 (O'Neill, Golenbock et al. 2013), NF-KB (Kawai and Akira 2007) and interferons (IFN) (Uematsu, Sato et al. 2005), generating an innate immune response against the pDNA backbone which leads to silencing of pDNA-encoded transgene or elimination of the transfected cells (Liu, Liang et al. 2017). The expression of transgenic proteins as well as the presence of foreign DNA can trigger also an immune response resulting in suppression or elimination of the expressing vectors (Sarukhan, Garcia et al. 1998).

Kay et al. (Chen, He et al. 2003) demonstrated that the bacterial backbone is mainly responsible for the silencing of the vector. It is suggested that inactive chromatin structures can spread from methylated regions or can generate small interfering RNAs (RNAi) which can lead to transcriptional inhibition. The gene silencing process is a fundamental mechanism that normally takes place during

the embryonic development. After fertilisation the blastocyst undergoes a rapid genome-wide de-methylation followed by gradual methylation after the implant in the womb. This *de novo* methylation occurs at discrete CpG sites and can spread throughout the chromatin. An active promoter or genomic elements called boundaries or insulators stop this procedure of spreading compact chromatin preventing the silencing of endogenous genes. This phenomenon is not limited to the early phase of embryo development or to integrated transgene. It has been reported that methyl transferase also acts on episomally maintained expression vectors such as EBV-vectors (Hsieh, Lemas et al. 1999)*)* and they are mostly active in the region of the bacterial backbone.

1.6.2 DNA Viruses based on episomally maintained DNA plasmids

To circumvent problems inherent in genotoxicity, the genetic engineering of episomal (extrachromosomal) eukaryotic vectors offers an attractive alternative.
The idea arose from the studies into episomally maintained DNA Viruses (Wade-Martins, Saeki et al. 2003). These viruses are able to hook their genome onto the nuclear matrix of target cells and they are maintained episomally. Like all the other types of viruses, they are infectious agents that take advantage of the host cell transcription and translation machineries for the expression of proteins. The cell replication machinery is recruited then for the synthesis of new viral genomes. Viral proteins are able to build bridges between the viral DNA sequences and the host genome preventing its integration. This interaction represents the basis for the mechanism of latency.

1.6.2.1 Polyomavirus simian virus (SV40) based vectors

The first virus-derived vector for gene therapy was based on genetic elements of the Polyomavirus simian virus 40 (SV 40). Vectors derived from this virus are reported to be effective in delivering constitutively expressed cDNA that encodes protein products as well as untranslated RNA

products to many different cell types (Kalos and Fournier 1995, Ohgane, Yagi et al. 2008). Despite the high efficiency in transgene delivery the usage of this vector system is limited to a transient set-up. This virus can replicate several times within the cells at every cell cycle and as a consequence of the over expression of viral genes results in cell death. For stable expression experiments, episomal vectors were designed on the basis of viral elements of BK virus (BKV), bovine papilloma virus (BPV-1) and Epstein-Barr virus (EBV). Each of these vectors contains a viral origin of DNA replication and viral early genes which activates the viral origin and thus allows the episome to reside in the transfected host cell line as a non-integrated entity.

1.6.2.2 Epstein-Barr virus based episomal vector

The most progress towards the development of an efficient episomal vector have come from work on EBV vectors. Such vectors have been used to deliver a dystrophin expressing plasmid into mice resulting in a significant enhancement in number of muscle fibres expressing the recombinant protein for muscular dystrophy (Tsukamoto, Wells et al. 1999). Although promising, the use of this vector system presented several limitations. In cultured human cells and in the absence of selection medium, episomal EBV vectors have a prolonged but not indefinite retention. The major drawback of the system might be represented by the necessity of the persistent expression of a viral *trans*-acting factor which on its own my lead to transformation of the cells. In order to be retained episomally these vectors, indeed, require the co-expression of the protein EBNA1. To consolidate the safely profiles of this vector system it was shown in a hepatic background that, neither high levels of EBNA1 nor multiple copies of the episome interfere with the expression of liver-specific proteins (Lutfalla, Armbruster et al. 1989). However, *in vitro* experiments suggested that EBNA1 protein may not be as harmless as previously thought. It can bind, at least *in vitro,* to RNA and may therefore be capable of influencing expression at post-transcriptional level (Snudden, Hearing et al. 1994). *Sixbey and Pagano,* (Sixbey and Pagano 1985) showed some functional similarities between

the EBV oriP and the enhancing region near the proto-oncogene *c-myc*. They have demonstrated that under certain circumstances this sequence can also interact with EBNA1 resulting in the down-regulation of the endogenous gene. *In vivo* experiments also create doubts about the safety of EBNA1. The expression of this protein predispose B cells to lymphoma in transgenic mice. The tumour cells are remarkably similar to those induced by transgenic *c-myc* over-expression (Wilson, Bell et al. 1996).

1.6.3 Non viral episomal vectors

The first attempt to generate non-viral episomal vectors date back to the early 80s, when autonomous replicating sequences (ARS) were described in yeast (Stinchcomb, Thomas et al. 1980). They were generated to study and better understand the replication control in higher eukaryotes but they also became immediately popular as a vector platform to genetically modify cells and organisms. This new class of vector reduced the safely problems of viral integrative vectors since they could replicate as an autonomous units and not require any exogenous trans-acting protein. They also avoid problems linked to insertional mutagenesis and do not lead to transformation of the cells caused by virally encoded proteins.

ARS are short (ca 125bp) AT-rich DNA sequences with a highly conserved 11 bp core sequence that were isolated from the yeast genome.

However, transfection of such vectors into mammalian cells never resulted in episomally maintained plasmids and in most cases this constructs were either lost from the cells or integrated into the host genome under selective conditions (Mesner, Hamlin et al. 2003).

Subsequent sequence analyses of various mapped mammalian origins of replication revealed no sequence homologies but rather a number of structural characteristics, such as AT-rich regions, bend DNA and the presence of what is called S/MAR (scaffold matrix attaching region) sequences.

1.7 Scaffold Matrix Attachment Region

Scaffold/matrix attaching regions (S/MAR) are defined as genomic DNA sequences that anchor chromatin to the nuclear matrix during interphase (Mirkovitch, Mirault et al. 1984, Cockerill and Garrard 1986). These sequences size range from 0.3 to 5 kb and possess a high AT content. The binding of the S/MARs to the nuclear matrix form looped domains that contribute structurally to the packaging of chromatin and functionally to the regulation of gene expression. Is still unclear what the real influence of the S/MARs to gene expression is. Their capability to generate curved DNA structures and their nucleotide composition which may destabilise or aid in the unwinding of the DNA duplex differentiate these sequences from standard promoters, enhancers and coding sequences. S/MARs are evolutionarily conserved and they have been identified in the introns (Kalos and Fournier 1995, Ohgane, Yagi et al. 2008) of several large genes, at the borders between two transcriptional units and in close proximity to enhancers. S/MARs are believed to increase transgene expression by facilitating the access of enhancers and transcription factors to gene regulatory regions to drive transcription, as well as providing mitotic stability by attaching the DNA to the nuclear matrix of the segregation of DNA into progeny cells (Jenke, Stehle et al. 2004). The formation of loops induced by the presence of these sequences may insulate genes from inactive chromatin positional effects through the recruitment of chromatin remodelling proteins which maintain the chromatin of the gene of interest in a transcriptionally active state, preventing it from silencing (Jenke, Fetzer et al. 2002) (Bode, Schlake et al. 1995).

1.8 Plasmid Episomal (pEPI): the first non viral autonomous replicon

The huIFNβ-S/MAR based vector pEPI, developed in 1999, was the first non-viral plasmid vector to exploit only mammalian sequences for its replication, maintenance and transgene expression in

mammalian cells (Piechaczek, Fetzer et al. 1999). An S/MAR sequence derived from the human interferon β gene cluster was inserted in the plasmid right after the transgene. This system is dependent on the presence of an active gene upstream of the S/MAR element. The pEPI vector is retained at a low copy number, around five to ten copies per cell when transfected into Chinese Hamster Ovarian (CHO) cells and it is able to provide long term gene expression for at least 4 weeks. Episomal replication of this vector was not restricted to CHO cells but was also observed in several tested cell lines. The efficiency of its replication, its mitotical stability and its segregation into daughter cells was also demonstrated. The vector was shown to be present in all the cell lines at an average copy number of fewer then 10 per cell.

However, despite the success of the pEPI vector *in vitro*, long-term gene expression in mice was not sustained following delivery of the vector to the liver or the lungs and has never been shown.

1.9 The maintenance and replication of pEPI *in vitro*

Although the vector pEPI carries all *cis*-acting sequences required to support episomal replication and maintenance, only a small percentage (less than 5%) of transfected cells are stably established by the episome. This implies that in addition to the DNA sequence, epigenetic factors probably play a crucial role in the establishment of an autonomous replicon, a phenomenon not restricted to non-viral replicons but has also reported for several viruses (Haan, Aiyar et al. 2001). Although many of the DNA vectors reach the nucleus of target cells the establishment of the replicon is due to a stochastic procedure in which the majority of the DNA molecules are subsequently lost within a few days. In stably established cells, the replicon was found only in less condensed chromatin regions and frequently associated with nuclear region involved in RNA processing. (Hagedorn, Gogol-Doring et al. 2017)

Jenke et al. (Jenke, Stehle et al. 2004) demonstrated that the pEPI vector binds to the nuclear matrix

and specifically to the prominent matrix protein SAF-A (scaffold attachment factor A).

SAF-A/hnRNP-U is a prominent protein of the nuclear matrix that binds specifically to S/MARs from yeast to human through the highly conserved 'SAF-box' (Kipp, Gohring et al. 2000). The presence of the S/MAR sequence mediates the association of the vector with the nuclear matrix and the chromosomes scaffold and this interaction correlates with the replication of the vector as an episome. A truncated version of the plasmid, lacking the S/MAR, is prone to integrate into the host-cell DNA (Baiker, Maercker et al. 2000).

S/MARs are involved in cohesion and separation of the chromatids (Mesner, Hamlin et al. 2003) and a vector carrying such sequences can therefore take advantage of the cellular replication machinery and undergo replication and segregation only during the S phase of the cell cycle. Once established the vector seems to be non-dynamic throughout the cell and is associated with histone modification typical of active chromatin, such as histone 3 methylation at Lys-4 (H3K4me1me3). This modification is typical of active chromatin enriched on genes replicating during the S phase. The accumulation of this modification on the S/MAR sequence highlights the relevance of this module for vector function. Only those vector molecules that reach a nuclear compartment favourable for transcription and replication are able to be retained as episomes. The heterogeneity in transgene expression within an established population demonstrates that transgene expression is dependent upon the nuclear localisation of the vector and consequently the chromatin structure that it adopts. Transcription running into or through the S/MAR appears to be necessary for the maintenance of the plasmid creating a chromatin structure that is accessible for replication enzymes. Indeed, silencing of the expression cassette linked to the S/MAR was shown to result in the loss of the episome from the cells (Hagedorn, Lipps et al. 2010).

1.9.1 A tissue specific non viral vector provides persistent transgene expression *in vivo*

The replacement of the original CMV promoter with tissue-specific ones leads to persistent transgene expression for at least 6 moths from a single administration (Argyros, Wong et al. 2008) in the liver of the treated mice. *Argyros et al.* demonstrated that swapping the original CMV promoter from the pEPI vector for a human liver specific one (α1-antytrypsin (AAT)) was sufficient to drive long-term episomal DNA expression *in vivo*. The removal of the S/MAR element resulted in the silencing of the vector within 1 week. The extensive methylation of CpG islands of a CMV promoter was reported to be responsible for the silencing of adenoviral vectors (Brooks, Harkins et al. 2004). In the presence of a S/MAR the CpG islands of the AAT promoter remains unmethylated, whereas the presence of the same sequence was unable to protect a corresponding CMV promoter from such transcriptional silencing. The decline in the transgene expression in the first few days upon DNA delivery may be due to a combination of events. The elimination of the transfected cells damaged during the procedure of infusion process, degradation of unstable DNA, as well as immune responses against the foreign DNA. Although the vector shows a high efficiency in terms of transgene expression after delivery, neither the CMV promoter nor the liver specific vector replicated following liver regeneration after a 70% partial hepatectomy.

1.10 Limitations and drawbacks of the pEPI system

The episomal S/MAR vector (pEPI) first described by *Lipps at al.* has been extensively studied and developed over time. Over the years the efficiency of the pEPI S/MAR vector for the genetic modification of cells has been described (Piechaczek, Fetzer et al. 1999, Argyros, Wong et al. 2008, Wong, Argyros et al. 2011, Hagedorn, Antoniou et al. 2013). The vector was shown to be efficient when used to deliver transgenes in cells in culture as well as in the liver of mice, sustaining the

expression of genes of interest for several months. The prototype vector pEPI depends on a transcription unit starting from constitutively expressed *Cytomegalovirus* immediate early promoter and directed into a 2000 bp long MAR derived from the human β-interferon gene. It contains two mammalian transcription units, one necessary for the vector propagation in bacteria which was also optimised for conferring the antibiotic resistance in cells and the other carrying the elements for the expression of the GOI. The S/MAR sequence is placed between the two transcript units (Figure 3).

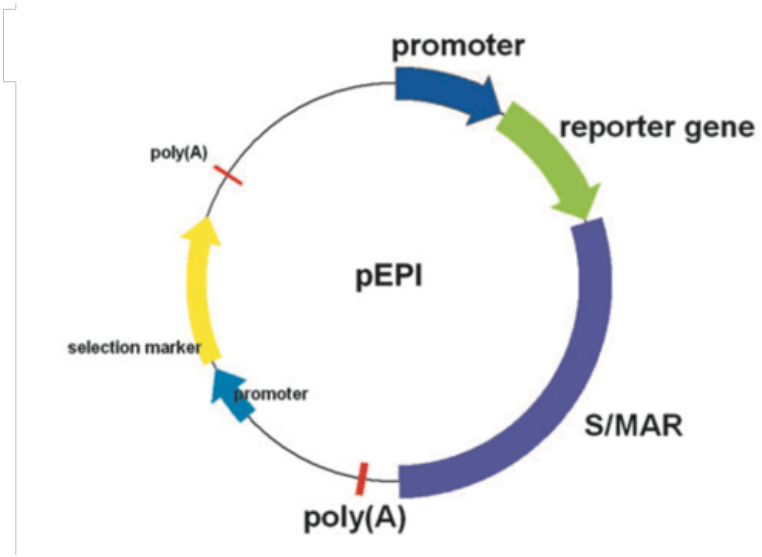


Figure 3. **Schematic representation of the pEPI DNA vector system.** pEPI consists of two mammalian transcription units oriented in a clockwise direction. The first, a CMV promoter drives the transgene transcription unit and it is oriented into the S/MAR. The second, the Simian Virus 40 Ori/Promoter (SV40-P) drives the neomycin transcription unit and it has been used for bacterial and mammalian selection. Figure from *Piechaczek et al.* (6)

pEPI replicates episomally at a copy number of approximately 5-10 molecule per cell in all mammalian cell tested, it is mitotically stable and provides long-term expression of transgenes and shRNAs. The vector has been demonstrated to be replicated once per cell cycle due to its association with early replication loci. The establishment of stable pEPI episomes in transfected cells is very inefficient; only 0.5 - 5 % of transiently transfected cells develop into stable clones. It is assumed that the vector establishment in the nucleus of the target cells is a stochastic event and strongly depends on the nuclear compartment that it reaches after transfection. In 2013 *Hagendorn*

et al. showed that the introduction of genomic *cis*-acting sequences into pEPI improved the establishment efficiency. They demonstrated that the ubiquitous chromatin opening element (UCOE) enhanced the transgene expression and the establishment efficiency presumably via additional interactions with the nuclear matrix. It was also reported that another genomic sequence, called cHS4 is able to enhance the number of established clones but not the expression of the transgene (Figure 4).

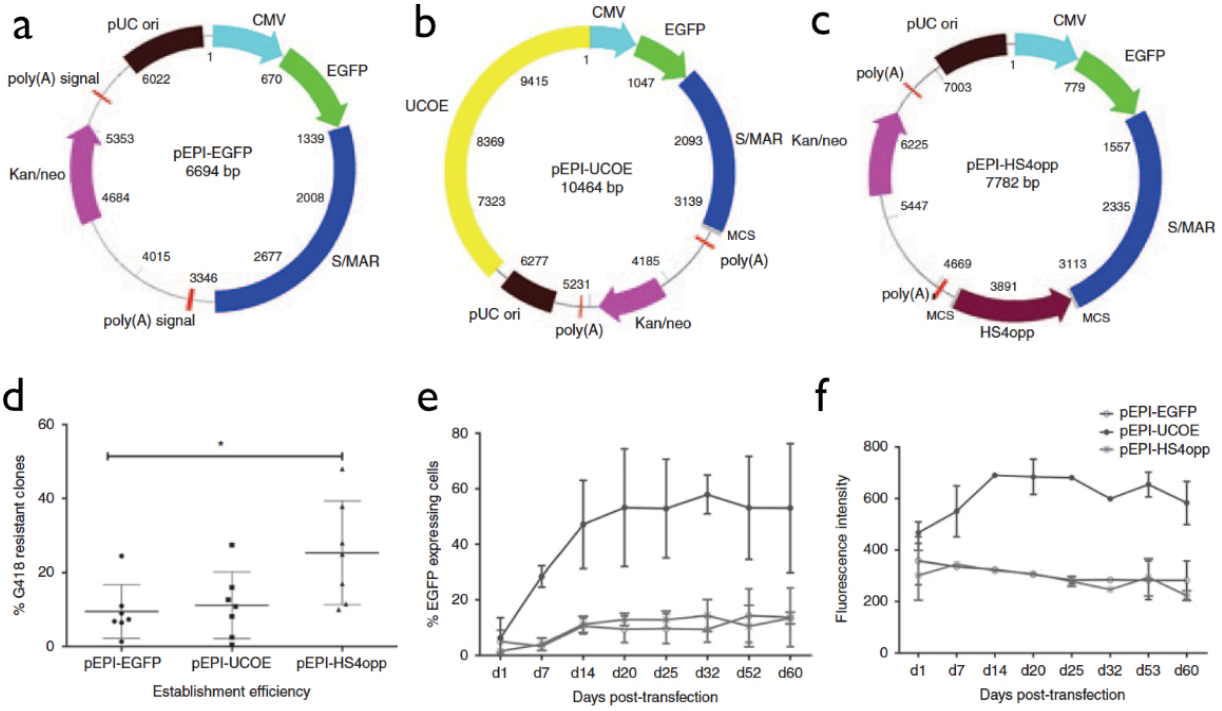


Figure 4. **CHO cell line establishment with different versions of pEPI.**
(a) In the basic pEPI vector, the CMV promoter drives the expression of the reporter gene GFP. This expression cassette sustains the expression of the transgene in the target cells and it also responsible for the functionality of the S/MAR which is placed right after the GFP coding sequence. The bacterial backbone containing the Kanamycin/Neomycin resistance and the pUC origin of replication complete the vector. (b) The introduction in the basic pEPI system of the cis-activating genomic sequence UCOE before the expression cassette and (c) the chicken insulator (HS4) after the expression cassette prevent the vector from chromatin induced silencing. (d) Colony forming assays show that the pEPI-HS4 vector generates a significant higher number of resistant cells when compared to the other two variants. However, the number of GFP expressing (e) cells and the intensity of the transgene expression (f) improve only in the pEPI-UCOE version of the plasmid. enhances the number of established cells (a) and the transgene expression (b-c). Figure adapted from *Hagendorn et al*. (35)

Although the introduction of the cHS4 insulator results in the generation of more resistant cells (Figure 5, d) only ~ 20% of those test positive for transgene expression (Figure 5, e). The cHS4 sequence is therefore improving the expression of the cassette that provides the resistance to the antibiotic but it has no effect on the transgene. The presence of the vector in the cells is confirmed by their acquired resistance to the selection but the mammalian expression cassette most likely still undergoes a process of chromatin induced silencing since the number of positive cells range between 5% and 20%. The UCOE sequence, instead, doesn't affect the total number of established cells and for those which became established the number of GFP expressing cells ranges from 40% to 80%. Although this vector improved significantly the number of transgene expressing cells, the big variation within replicates and the proportion of negative cells in the populations underlines the major problem related to this vector system. Independently from the plasmid, the obtained populations present a large proportion of non-expressing cells even with selection with antibiotics. In these populations the vector could be lost over time or it might undergo a process of epigenetic induced silencing. The copy number assay revealed that 5 to 10 copy of the plasmid per cell were present in the populations. Since only a proportion of the cells retained the transgene expression it seems that the vectors are retained and replicated in the cells but in a large proportion they became inactive. This process can be the result of epigenetic silencing. The vectors are present and can be detected with the copy number assay but they don't contribute in the overall GOI expression. Though these prototype vectors represent a suitable tool for simple studies as cells labelling with reporter gene (e.g. Luciferase or GFP) for more sophisticated work such as gene-rescue it presents several limitations.

1.11 The Next generation DNA Vector

Variable expression in different cell lines, gene silencing *in vivo* and its low establishment rate represent the major limitations of the prototype pEPI vector. Modifications have been made with

the aim of increasing the transgene expression as well as the establishment rate. Some of the features of the vector are crucial for its episomal maintenance. The deletion of the S/MAR sequence results in its random integration under mammalian selection and deletion of the promoter or the insertion of a termination signal between transgene and S/MAR also result in the loss of the episome (Rupprecht, Hagedorn et al. 2010). *Argyros et al.* developed the vector pLuca swapping the original CMV promoter for a range of mammalian derived ones with the aim of generating a platform suitable for gene therapy. Moreover, to avoid innate immune reactions a reduction of the CpG motifs has proven to improve the vector establishment *in vitro* as well as the transgene expression *in vivo* (Haase, Magnusson et al. 2013). As previously mentioned *Key et al.* demonstrated that the bacterial backbone and more specifically, the gene that provides the resistance to the antibiotic can generate small interfering RNAs which can lead to transcriptional inhibition (Chen, He et al. 2003).

To overcome the problems induced by the presence of the bacterial backbone, *Bigger et al. (Bigger, Tolmachov et al. 2001)* developed a bacterial strain which exploits the properties of the *cre recombinase* enzyme to generate what they called DNA mini-circles a class of vector that lacks bacterial elements.

1.11.1 The S/MAR minicircle technology

Minicircles are DNA plasmids in which the bacterial backbone is removed and they are constructed in bacteria with the use of recombinases (Bigger, Tolmachov et al. 2001, Vaysse, Gregory et al. 2006). The introduction of two Flp sites, one before the promoter and one after the mammalian expression cassette allow, upon *Cre recombinase* induction, an intramolecular recombination that generates two circles: one containing the minicircle vector comprising exclusively the promoter-transgene cassette and one containing the bacterial backbone called miniplasmid.

This technology was used also the generate S/MAR-minicircles (Broll, Oumard et al. 2010)

(Argyros, Wong et al. 2011). Interestingly the prototype pEPI as well as the minicircles establish within a certain number of plasmid copies per cell, which suggest a stringent copy number control in the recipient cells. Although the number of vector molecules per cell ranges between 5 and 10, the expression levels of the minicircle vectors are higher when compared to normal plasmids. S/MAR minicircles establish *in vitro* in the absence of selection pressure and are maintained as episomes *in vivo* when delivered to the liver of mice. (Argyros, Wong et al. 2011) providing a longer and higher expression of the transgene.

This once again leads to the suggestion that epigenetic features, chromatin structure, nuclear localisation and last but not least DNA composition strongly influence the regulation of transcription from a non-viral episome.

1.12 Application of non viral, non integrative DNA vectors

Gene Therapy represents an obvious field of application for non-integrative DNA Vectors.

The delivery of corrective genetic material into cells is meant to cure or at least alleviate the symptoms of the disease. Although gene therapy seems an easy approach, in practise, considerable problems have emerged. In principle it consists in the identification of an appropriate DNA sequence and cell types an in the development of suitable tools for getting enough of the DNA into the target cells. The lack of efficient delivery systems, the lack of sustained expression and the host immune reactions remain still a big challenge to overcome for gene therapy to become a routine practise. (Verma and Somia 1997). Optimal vectors for gene therapy require (a) high and stable level of expression of the gene of interest, (b) high transfection efficiency, (c) no integration into the cell genome, (d) no transformation features that may induce secondary diseases. Many of the systems currently available utilise viral vectors derived from retrovirus, lentivirus or adenovirus.

A less immunogenic, episomally maintained DNA vector should represent a valuable alternative particularly where the current necessity of personalised approaches (e.g. autologous therapy) are

becoming more important. The treatment of severe diseases by the *ex-vivo* modification of your own cells is opening an astounding field of therapy as well as immunotherapy is becoming a standard front line approach. Autologous cell therapy (ACT) is a novel therapeutic intervention that uses an individual's cells, which are cultured and expanded outside the body and reintroduced into the donor. The advantages of such practise include the minimisation of risks from immunological reaction and disease transmission. Adoptive immunotherapy of malignancies involves the passive transfer of anti-tumour reactive cells into a host in order to mediate tumour regression. These cells are engineered *in vitro* by adding T Cell Receptors (TCR) or Chimeric Antigen Receptors (CAR). Currently, like in most of the gene therapy approaches, integrative system like viruses (Kershaw, Teng et al. 2005) or SBs (Deniger, Yu et al. 2015) are used for the genetic modification of these cells. Despite the efficiency the so modified cells present may present the side effects previously described. In was reported (Hacein-Bey-Abina, Von Kalle et al. 2003) that a promising clinical with T cells modified with a lentivirus may leads to complicate side effects due to the unexpected premalignant cells proliferation induced by the random integration of the viral genome.

Argyros, Wong et al. (Argyros, Wong et al. 2008) demonstrated that an episomally maintained plasmid (pDNA) expression system harbouring a Scaffold/Matrix Attachment Region (S/MAR) can provide persistent and robust transgene expression in human cancer cells *ex vivo, in vitro* and *in vivo* and it can be further improved when the bacterial backbone is removed (Argyros, Wong et al. 2011). Thus this class of DNA vector can be further exploit for gene therapy approaches as well as for their capability of genetically engineer human primary cells such as lymphocytes.

1.13 Aim of the project

The principal aim of this project is to refine and enhance the DNA vector system described earlier from *Piechaczek et al.* (Piechaczek, Fetzer et al. 1999) and *Hagedorn et al.* (Hagedorn, Antoniou et al. 2013) to develop a range of next generation DNA vectors which provide several advances over the prototype vectors. The low efficacy of pEPI in establishing dividing cells as well as the current proprietary restriction for the use of minicircles for research inspired this work.

The use of minicircle was not only limited form commercial licenses but also the modification and the preparation of the vectors is particularly long, tedious and expensive. The intracellular and intramolecular recombination generates not only a supercoiled form of the plasmid but also several concatemers which can be excluded from the DNA preparation only with a preparative High Performance Liquid Preparation (HPLC).

Thus, the idea in this PhD project is to generate a cheap, ubiquitous and efficient non integrative DNA vector platform that can be used for the persistent, stable and safe genetic modification of mammalian cells.

2 MATERIAL AND METHODS

2.1 Materials

The materials used in these experiments are listed below, along with the details of the suppliers from which they were purchased.

2.1.1 General Chemicals and Reagents

Table 2. General chemicals and reagents

Agarose	Sigma Life Science	Sodium dodecyl sulfate (SDS)	Serva
LB-Broth	Invitrogen	Protease inhibitor cocktail tablets	Roche
Ethanol	Sigma-Aldrich	Tris	Sigma-Aldrich
Isopropanol	Sigma-Aldrich	EDTA	Acros
Formaldehyde	Sigma-Aldrich	Sodium Chloride	Sigma-Aldrich
Kanamycin	Roche	Hydrochloric Acid	Sigma-Aldrich
Luciferin	BIOMOL	Hydroxyl Acid	Sigma-Aldrich
Molecular DNA Marker	Thermo Fisher	Polysorbate 20	AppliChem
Paraformaldehyde	Sigma-Aldrich	Boric Acid	Merck
SOC Medium	Invitrogen	Potassium Chloride	AppliChem
Magnesium Chloride	Merck	Calcium Chloride	Merck
Methanol	Sigma-Aldrich	Crystal Violet	AlfaAesar
Phenol:Choloform	Sigma-Aldrich	Agar	Roth
SOC Medium	Clonetech	Milk powder	Roth
PeqGree DNA/RNA Dye	PeqLab		

2.1.2 Enzymes and molecular biology reagents

2.1.2.1 Restriction Enzymes

Table 3. Restriction enzymes

*Hind*III	Thermo Fischer	*Hind*III	Thermo Fischer
*EcoR*I	Thermo Fischer	*Pst*I	Thermo Fischer
*Xho*I	Thermo Fischer	*Bam*HI	Thermo Fischer
*Mlu*I	Thermo Fischer	*EcoR*V	Thermo Fischer
*BsrG*I	Thermo Fischer	*Nde*I	Thermo Fischer
*Nhe*I	Thermo Fischer	*Age*I	Thermo Fischer
*EcoR*V	New England Biolab	*Hind*III	New England Biolab
*BamH*I	New England Biolab	*BstZ*17I	New England Biolab
*Spe*I	Thermo Fischer	*Sma*I	New England Biolab

2.1.2.2 Other enzymes for DNA manipulation

Table 4. Other enzymes for DNA manipulation

Klenow large fragment DNA Polymerase I	Invitrogen
T4 DNA ligase	Invitrogen
In Fusion HD Cloning	Clonetech
Plasmid-Safe-ATP-Dependent DNase	Epicentre

2.1.2.3 PCR reagents

Table 5. PCR reagents

2'desoxynucleotide 5'triphosphate mix (dNTPs)	Invitrogen
Taq DNA Polymerase Hot Start	Qiagen
10X Loading buffer	Thermo Fisher
CloneAmp HiFi PCR Premix	Clonetech

2.1.3 Tissue culture reagents

Table 6. Tissue culture reagents

DMEM	Sigma
RPMI	Sigma
IMDM	Sigma
Genetical (G418) sulphate	Roth
Puromycin	Panreac AppliChem
Fetal Calf Serum	Gibco
OptiMEM serum-free medium	Life-Technology
Penicillin/Streptomycin (x100)	Sigma-Aldricht
L-Glutammate	Gibco
Trypsin EDTA (x1)	Sigma-Aldricht

2.1.4 Transfection reagents

Table 7. Transfection reagents

jetPEI DNA Transfection Reagent	Polyplus
jetPrime DNA Transfection Reagent	Polyplus
Capan-1 Transfection Reagent	Altogen
Amaxa Human T Cell Nucleofection Kit	Lonza
Amaxa P3 Primary Cell Nucleofection Kit	Lonza
Amaxa SE Nucleofection Kit	Lonza
Amaxa AD1/AD2 Nucleofection Kit	Lonza
Amaxa 4D Nucleofector Optimisation Kit	Lonza

2.1.5 Bacterial Media

Table 8. Bacterial Media

LB Broth		1% Sodium Chloride
		1% Tryptone
		0.5% Yeast Extract
		ph adjusted to 7.5 with NaCl
LB Agar	LB Medium supplemented with	1.5% agar
SOC Medium		2% (v/v) Tryptone
		0.5% (v/v) Yeast Extract
		10 mM Glucose
		10 mM NaCl
		2.5 Potassium Chloride

2.1.6 Antibiotics

Stock solutions of antibiotics were prepared at the working concentrations below and stored at -20° C

Table 9. Antibiotics stocks concentration

Kanamycin	50 mg/ml	Sigma-Aldrich
Puromycin	1 mg/ml	Panreac AppliChem
Ampicillin	50 mg/ml	Panreac AppliChem

2.1.7 Southern blot analysis reagents

Table 10. Southern blot analysis reagents

Depurination Solution	250 mM HCl
Denaturation Buffer	500 mM NaOH
Neutralisation Buffer	1.5 M NaCl

	0.5 M Tris
	adjust to pH 7
20X SSC	3M NaCl
	300 mM Sodium Tricitrate
	adjust to pH 7
1x Church Buffer	250 mM Sodium Phosphate
	1 mM EDTA
	1% BSA
	7% SDS
	adjust to pH 7

2.1.8 Bacterial strain

Table 11. Bacterial strain

E.Coli DH10B	Invitrogen	Stbl3	Invitrogen
E.Coli DH5a	Invitrogen	Stellar Super competent Cells	Clonetech

2.1.8.1 Commercial kits

Table 12. Commercial kits

QIAprep Spin Miniprep Kit	Qiagen
QIAquick Gel Extraction Kit	Qiagen
Endofree Plasmid Maxi	Qiagen
Quiquick Spin PCR Purification Kit	Qiagen
RNeasy Mini Kit	Qiagen
DNeasy Blood and Tissue Kit	Qiagen

BCA protein assay	Piercenet

2.1.9 Histology and Immunohistochemistry reagents

Table 13. Histology and Immunohistochemistry reagents

ABC Complex/HRP	DakoCytomation
Avidin/Biotin Blocking Kit	Vector Laboratories Inc
Biotinylated goat anti-rabbit immunoglobulins	Dako
DAB (3,3'-dianminobenzidine)	Vector Laboratories Inc
Horse Serum	Dako
Biotinylated horse anti mouse anti-rabbit immunoglobulins	Vector Laboratories Inc
Biotinylated horse anti mouse	Vector Laboratories Inc
Hydrogen peroxide	Sigma-Aldrich
PBS (non tissue culture)	Sigma-Aldrich
PAP (hydrophobic barrier pen)	Vector Laboratories Inc
Rabbit polyclonal Luciferase antibody	Santa Cruz Biotechnology
Haematoxylin stain	BDH
Histo-clear	National Diagnostics (USA)

2.1.10 Cell lines and primary cells

Table 14. Cell lines and primary cells

Hek293T	ATCC
Capan-1	ATCC
BxPC-3	ATCC
Mouse Lungs Fibroblast	Isolated from BALB mice

Panc-1	ATCC
MiaPaCa-2	ATCC
PC-3	ATCC
Paco-2	Patient Derived Pancreatic Cancer Cells
PBMC	Isolated from healthy donors
CD8+	Isolated from healthy donors

2.1.11 Equipment

Table 15. Equipment

FusionSL Vilber Lourmat	PeqLab	Duomax 1030	Heidolph
Centrifuge 5424R	Eppendorf	MiniVac Power	PeqLab
Centrifuge 5430R	Eppendorf	Cell Sure Lock	Invitrogene
PersonalHyb	Stratagene	Thermomixer Compact	Eppendorf
Thermomixer comfort	Effendorf	MicroPulser	BioRad
RM5 (Roller)	CAT	MR Hei-Tec	Heidolph
KB 2400-2N	Kern	PeqStar PCR machine	PeqLab
Mini PROTEAN Tetra Cell	BioRad	Light Cycler 96	Roche
PCR Workstation	Labcaire	QIAcube	Qiagen
PowerPac Basic	BioRad	PowerEase	Invitrogene

40-0708 DNA Electrophoresis	PeqLab	Horizontal Gel Electrophoresis system	Horizon
NUARE incubator	Tecnomara	Safe 2020	Thermo Fisher
CK40 Microscope	Olympus	Amaxa Nucleofector II	Lonza
Amaxa 4D-Nucleofector	Lonza	Centrifuge 5810	Eppendorf
N90 UV-Transilluminator	Konrad Benda	Sorvall RC6+	Thermo Fisher
Keyence BZ-9000	Keyence	Sorvall RC5+	Thermo Fisher
Certomat SII	Brown Biotech	GenPure Pro	Thermo Fisher
pH Meter 538	WTW	Precellys Control Device	PeqLab
MS2 Minishaker	IKA	NanoDrop 2000c	Thermo Fisher
MultiskanEx	Thermo Fisher		

2.2 Methods

2.2.1 Molecular biology techniques

2.2.1.1 Agarose gel electrophoresis

Electrophoresis was performed on 1% agarose gels with 3 µl PeqGreen DNA/RNA Dye pro 50 ml of agarose solution. Before loading, the DNA samples were mixed with the 10X Loading Buffer and loaded into wells with an appropriate molecular weight marker. Electrophoresis was usually carried out at 75-100 V in a *40-0708 DNA Electrophoresis chamber* PeqLab with a PowerPac Basic BioRad power supply. Visualisation of DNA fragments was achieved by viewing on the UV-Transillumiator. Images were capture using the FusionSL Vilber Lourmat gel documentation software package (PeqLab).

2.2.1.2 Purification of DNA using Qiagen column

Up to 10 µg DNA can be purified with the PCR purification kit from Qiagen. This method uses a column with a silica matrix membrane that absorbs DNA at high salt conditions while unwanted contaminants pass thorough. This was used to purify double stranded DNA fragments from PCR, other enzymatic reactions and when was necessary to purify linearised plasmids. 5 volumes of binding buffer were added to 1 volume of DNA sample and mixed. The sample was then applied to a spin column and centrifuged at high speed for 1 min. The columns were washed with a ethanol based solution and the DNA equated in 30-50 µl TE Buffer.

2.2.1.3 Purification of DNA from agarose gel

DNA fragments of interest were excised from the gel under UV illumination using a scalpel blade

and purified using a QIAquick Gel Extraction Kit from Qiagen. The column membrane can be washed under high salt conditions to remove agarose and the ethanol to remove salt and other contaminants. Briefly, gel slices were incubated at 50° C in 3 gel volumes of the solubilisation buffer for 10 min until the agarose had dissolved. One gel volume of isopropanol was added and the mixture applied to the spin column and allowed to bind to the matrix. The columns were washed with wash buffer then the DNA eluted in 30-50 µl TE buffer.

2.2.1.4 Digestion with restriction enzymes

Restriction digests were carried out according to the supplier's instructions. DNA samples were digested with the appropriate enzyme in a compatible buffer and distilled water. Usually up to 1 µg of DNA were digested with 1-5 µl of restriction enzymes (10 U/µg) with appropriate 10x enzyme reaction buffer in a 20 µl (for the Thermo Fisher enzymes) or 50 µl (for the New England Biolab's) reaction volume, made up with distilled water. The incubation period and the temperature was usually performed at 37° C for 10 min (Thermo Fisher) or 1-2 h (New England Biolab). Double digests were either performed in one step in the enzyme performed optimally in a similar buffer or in two sequential steps with an intervening buffer exchange step between the digests. The desire DNA fragments were purified either with through PCR Purification Kit from Qiagen or thorough the gel extraction protocol previously described and elute in 30-50 µl of TE buffer.

2.2.1.5 DNA modifying enzymes

Fill-in overhanging restriction endonuclease termini using Klenow

When it was not possible to generate compatible sticky ends in both vectors and insert fragments, terminal overhangs were blunted by treatment with the Klenow fragment of DNA Polymerase I from *E.Coli*. The Klenow fragment exhibits 3' to 5' exonuclease activity and 5' to 3' polymerase

activity, but lacks the 5' to 3' exonuclease activity. During restriction digestion, 0.5U Klenow/µg DNA was added to the reaction mix along with a small quantity (<0.5 mM) of dNTPs and incubate at 37° C for at least 1h.

2.2.1.6 Ligation of digested fragments

T4 DNA Ligase was used for the blunt end ligations. The relative amounts of digested vector to insert were adjusted so that the molar ratio of insert vector varied between 3:1 and 20:1 The fragments were usually incubated with 2 µl T4 DNA Ligase (1U/µl) in the provided 5x T4 DNA ligase buffer at 14° C overnight. After which, usually 1-2 µl of the ligation mixture was transformed by heat shock into Stellar Supercompetent cells (Clonetech)

2.2.1.7 In-Fusion Cloning

When possible the insertion of one or more DNA fragments into an acceptor vector was made through the In-Fusion cloning system provided by Clonetech. Briefly, the insert was cloned into the plasmid via homologous recombination instead of ligation of two compatible ends. The acceptor vector was prepared for the cloning by restriction digestion and the it was purified either with the PCR Purification kit from Qiagen when linearized or with the QIAquick Gel extraction kit when a double digests was required. The insert was prepared through PCR and the primers were designed in order to present 15 bp homologous to the target sequence. The insert was also purified with the PCR Purifiation kit from Qiagen. Normally, 100 ng of the linearised vector and 50 ng of the amplicon were mixed together with the In-Fusion Cloning mix from Clonetech following the manufacture protocol in a total volume of 10 µl. The reaction was incubated for 15 min at 50° C, chilled into ice and 2.5 µl were then transferred into 50 µl of Stellar Supercompetent Cells. After 30 min incubation in ice, the cells were transformed through heat shock.

2.2.2 Bacterial cell methods

2.2.2.1 Growth and storage of *Escherichia Coli*

Liquid culture of *E.Coli* DH10B, *Stbl3*, Stellar Supercompetent, and *E.Coli* DH5a were grown into sterile Luria-Bertani (LB) broth with the appropriate antibiotic at 37° C in a shaking incubator at 200 rpm. Culture volumes were dependent on the nature of the experiment and varied between 5 ml for a plasmid extraction by Miniprep kit to 250 ml for production of larger amount of pDNA by Maxiprep. To obtain separate bacteria colonies derived from a single cell, cultures were spread onto solid agar plates and incubated at 37° C overnight. For short term storage bacterial cultures were stored at 4° C. Long term storage of bacteria were stored at -80° C following addition of glycerol to the final concentration of 20% (v/v) glycerol.

2.2.2.2 Transformation of bacterial cells

2.2.2.2.1 Transformation by heat shock

Commercially available chemically competent *E.Coli* DH5a and Stellar Supercompetent Cells were aliquoted into 50 µl volumes in chilled 1.5 ml tubes containing the desire amount of pDNA to transform, 1-100 ng or 2.5 µl of the In-Fusion reaction. The cells were incubated on ice for 30 min, subjected to a 45 second heat shock at 42° C, then allowed to recover on ice for 2 min. 450-950 µl of SOC medium was added and the cultures shaken at 200 rpm for 60 min. After this time 10 µl and 100 µl undiluted cells were spread onto LB agar plates containing the appropriate antibiotic for selection of transformants.

2.2.2.2.2 Transformation of DH10B *E.Coli* with electroporation

In preparation for transformation, BIORAD Gene Pulser (1mm) cuvettes were chilled on ice and DH10B bacteria cell aliquots thawed on ice. For the plasmid rescue up to 2 µg of DNA was added to bacteria aliquots and the pipetted into chilled cuvettes. The cuvettes were electroporated in a BIORAD Gene Pulser using the standard protocol for bacteria. Immediately afterwards the cells were incubated at 37° C for 60 min in a shaking incubator set at 200 rpm. The cells were plated as described before.

2.2.3 Isolation of pDNA

2.2.3.1 Small scale isolation of pDNA by Miniprep

For small scale plasmid isolation, 5 ml cultures of *E.Coli* containing the plasmid of interest were grown in LB with the appropriate antibiotic selection and plasmid DNA was isolated using the Miniprep spin column kit from Qiagen. Bacterial were pelleted by centrifugation at 15000 rpm for 1 min in a Eppendorf 54324 R Centrifuge, re-suspended in a RNase A-containing resuspension buffer then lysate in the provided lysis buffer. The solution was neutralised and protein precipitated by adding neutralisation buffer, debris was spun out by centrifugation ad 15000 rpm fro 10 min and resulting supernatant was transferred to a spin columns and allowed to bind the column matrix. Following two wash steps, the pDNA was eluted in 50 µl elution buffer. The protocol results in isolation of approximately 10 µg DNA.

2.2.3.2 Large scale isolation of pDNA by Maxiprep

For production of larger quantities of DNA, the EndoFree Plasmid Maxiprep kit from Qiagen was used according to the manufacture's instructors. The principle is the same as the Miniprep kit, but the volumes are scaled up. Briefly, 5 ml started cultures were used to inoculate 250 ml fresh LB

medium in 1 l flask containing the appropriate antibiotic. The cells were pelleted for 15 min at 6000 rpm in a Sorvall RC6+ centrifuge and resuspended by vortexing in resuspension buffer containing RNase A. The cells were then lysed with the addition of lysis buffer at room temperature for 5 min. The lysed cells were neutralised with neutralisation buffer and the lysate was incubated in endotoxin-removal buffer prior to binding to the column. The lysate was passed over a column with an ion-exchange resin, which binds DNA at a pH of 7 and ionic strength of 750 mM NaCl. Following elution, DNA was precipitated with 0.7 volumes isopropanol, pelleted by centrifugation at 15000 g for 30 min in a Sorvall RC6+ centrifuge, then washed with 70% ethanol before resuspended in an endotoxin-free elution buffer. The theoretical yield is approximately 1-1.5 mg of endotoxin free plasmid DNA.

2.2.4 Determination of DNA purity and concentration

To measure the concentration and purity of recovered pDNA, a NanoDrop 2000c spectrometer was used (Thermo Fisher). This spectrophotometer requires only 1µl on undiluted samples for assessment of concentration of double stranded DNA. To measure DNA, the instrument automatically detects the high concentration and uses 0.2 mm pathlength to calculate the absorbance at OD 260. An OD260 of 1 corresponds to 50 µg/ml for double stranded DNA. The absorbance ration of A260/A280 provides an estimate for the purity of the nucleic acid where acceptable levels of DNA purity have ratios of 1.7 to 1.9. A higher ration is due RNA contaminations, whereas a smaller ratio suggests protein impurities. Elution buffers used in each protocol were used as a controls for OD measurements.

2.2.5 Mammalian cell culture methods

2.2.5.1 Growth and maintenance of mammalian cells

The mammalian cancer cell lines were maintained in the respected recommended medium according to ATTC protocols. Normally Dulbecco's modified Eagle's medium (DMEM), Roswell Park Memorial Institute (RMPI), Iscove's Modified Dulbecco's Medium (IMDM) containing Glutamine and supplemented with 10% fatal calf serum (FCS), penicillin (5U/ml) and streptomycin (50mg/ml) were used to culture the cells in monolayers in sterile Petri Dishes in a NUARE incubator (Tecnomar) and maintained at 37° C in a humid 5% CO2/95% air atmosphere. When the cells were 80/90% confluent, growth medium was aspirated and the monolayer washed once with DPBS. Trypsin-EDTA solution was added and the dish incubated at 37° C until the cells began to detach from the flask and from each other. The dish was gently tapped if necessary, to further dislodge cells. Trypsin activity was neutralised by adding an equal volume of growth medium and the cells pipetted up and down until a homogenous single cell suspension was achieved. An aliquot of this was transferred to a new dish containing fresh growth medium and the dish replaced in the incubator. The cells were passaged twice weekly at a ration 1:5 - 1:10.

2.2.5.2 Growth and maintenance of patient derived human pancreatic cancer cells

The human patient derived pancreatic cancer cells (Paco2) were kindly provided by Dr. Martin Sprick, HiStem GmbH (Heidelberg). The cells were cultured in 75 cm2 PRIMARIA tissue culture flasks (Corning) in monolayers until 80/90% confluence. The growing medium as a well-defined cocktail of cytokines and growth factors was also provided from Dr. Sprick. Once a week the medium was aspirated from the flask and the cells washed with DPBS. Accutase solution (Promo Cell) was added and the flask incubated for 15 min at 37° C until the cells were detached and

dislodged. An equal volume of CO2 independent IMDM medium containing 1% BSA was added to neutralise the accutase and an aliquot of the cells was transferred, at a ration 1:3, into a new flask.

2.2.5.3 Long term storage in liquid nitrogen

For long term storage, confluent monolayers were trypsinised and resuspended in culturing medium containing 20% FCS and 10% DMSO. 1 ml volumes were aliquoted into cryotubes and frozen slowly to -80 C for 24 h. After this time the vials were stored in liquid nitrogen until further required. To bring cells up from the liquid nitrogen, the cells were thawed rapidly by placing them into a water bath at 37° C. An aliquot was the diluted in fresh medium and 10% FCS, cells pelleted at 1000 rpm for 5 min, resuspended in medium with 10% FCS and seeded in a new culture dish. The media was aspirated away and replaced the next day to ensure removal of any remaining traces of DMSO.

2.2.6 Transfection

2.2.6.1 Transfection using JetPEI DNA transfection reagent (Polyplus)

For the transfection of on well of a 12 well plate, 80000-200000 cells were seeded one day prior the transfection in 1-2 ml of growing medium. 2 µg of pDNA were diluted into 50 µl of 150 mM NaCl solution and 4 µl of jetPEI were diluted as well in 50 µl of 150 nM NaCl. The jetPEI solution was added to the DNA solution, vortexed and briefly span. After 15 min incubation of the combined solution at room temperature, it was added drop-wise to the cells in 1 ml of serum-containing medium.

2.2.6.2 Transfection with jetPRIME DNA siRNA Transfection reagent

For optimal DNA transfection of one well of a 12 well plate 80000-150000 cells were seeded one day prior the transfection in 1 ml of growing medium. 1µg of pDNA and 3 µl of jetPRIME reagent were diluted into 100 µl of jetPRIME buffer for complex formation. The solution was vortexed for 10 second and span down briefly. After 15 min incubation at room temperature is was transferred drop-wise to the cells in 1 ml serum-containing medium.

2.2.6.3 Nucleofection

The nucleofection solutions and the correspondent buffers were purchased from Lonza and the transfection was carried in either the Amaxa Nucleofector II or the Amaxa 4D-Nucleofector device (Lonza) following Lonza manufacture users guidelines. Briefly, 1-5x106 cells pro transfection were isolated with Trypstin-EDTA or accutase treatment and centrifuged at 200xg for 10 min at room temperature. The supernatant was discarded and the cells re-suspended carefully in 100 µl room temperature Nucleofector Solution pro sample. 1-5 µg od pDNA were added to the solution and the tube was gently flanked to generate an homogenous mixture. The cell/DNA suspension was then transferred into nucleofection cuvettes and it was then placed into the Nucleofector Cuvette Holder of the correspondent device. The transfection was achieved applying the recommended pulse. After the pulse 500 µl of pre-equilibrated culture media was added to the cuvette and the cells resuspended prior to transfer into a new well of a 12 well plate containing 1 ml of pre-warmed growing medium

2.2.7 Plasmid rescue

Usually 500 - 1000 ng of DNA extracted from cells were used for transformation into DH10B *E.Coli.* Bacteria were transformed by electroporation in a BIORAD Gene Pulser using the standard

pule for bacteria. Transformed colonies were selected on agar plates containing the appropriate antibiotic. DNA was isolated from individual resistant clones, subjected to restriction analysis with the appropriate enzymes, and analysed by gel electrophoresis on 0.8% agarose gels.

2.2.8 Gene expression analysis following pDNA administration

2.2.8.1 Fluorescent Microscopy

Fluorescent cells of interest were visualised under the GFP or RFP channel and photographed using a *Keyance* fluorescent microscope at 4x-20x magnification.The images were transferred to a computer and where necessary the brightness, contrast and colour levels were adjusted using *Affinity Designer* and *Affinity Photo* software. When necessary, fluorescent and bright field images were merged using *Keyence software,* a program that performs the task automatically

2.2.8.2 *In vitro* Bio-Imaging

Cells were washed in PBS, and their medium replaced by DMEM serum-free containing 150 μg/ml D-luciferin. After 5 min incubation in the cell culture incubator they were imaged for bioluminescence with the FusionSL (PeqLab). BLI was performed in the dark and the auto function was used to define the binning level. The acquisition time were ranged from 1 to 10 min depending from the cell lines.

2.2.8.3 FACS

Flow cytometry analysis were performed on a LSR Fortessa (BD Bioscience). When cells were cultured in a two dimensions system they were first detached as previously described, washed for two times with PBS before being re-suspended in 200-500 μl of PBS. For the analysis of the GFP

expression, shortly before the analysis of the samples the fluorescent marker Propidium Iodide for the discrimination of live/death cells was added to a final concentration of 0.01 mg/ml. Cells were incubated for 1 min and then placed in the FACS. The acquiring of the data was performed through the DIVA software and the data were analyzed with the software FlowJo.

2.2.9 Drug resistance selection

Stably transfected mammalian cells were placed in selective medium 48 h after transfection by addition of Geneticin (G418 sulphate) at a concentration of 1 mg/ml or Puromycin at a concentration of 0.5 µg/ml in growing medium. The cells were placed under selection pressure for 2-3 weeks with media refreshed every 3-4 days. After selection, colonies or mixed populations of cells containing the *neo* or the *puro* drug resistance gene were checked for expression, and placed in fresh normal medium or long-term gene expression studies.

2.2.10 Determination of protein concentration

The protein concentration of each sample was determinate using the BCA Protein Assay kit from Pierce. One volume of cell lysate was diluted in one volume lysis buffer and 10µl of each diluted sample was added to the wells of a flat-bottomed 96-well plate. When possible samples were analysed in triplicate. A batch of protein asset reagent was prepared by mixing Reagent A and B in a ration of 49:1, and 200 µl of this reagent was added to each well. Plates were incubated for 30 min at 37° C in the dark. The absorbance of each sample at 570 nm as measured using a *MultiskanEx* (Thermo Fisher) and the data exported into *Microsoft Excel.* A dilution series of albumin standards prepared in duplicate was assayed alongside the samples and used to generate a standard curve, to which the absorbance values of the unknown samples could be compared and their protein concentrations determined.

2.2.11 Western blot analysis

Approximately 5x105 cells or 10 mg of tissue were lysed in 50 - 500 µl 1 x lysis buffer containing 2% SDS, 10 mM Tris pH 7.5 and 0.1 mg/ml Protease inhibitor (Roche) and centrifuged at 20000 g for 30 min at 4° C. The protein concentration of the cell lysate was determined using the BCA protein assay and equal amounts (20 µg) were separated on a 4-12% SDS-PAGE gel with a molecular weight marker (Thermo Fisher) Following separation, total protein was transferred to a PVDF membrane (Millipore) using a iBlot2 (Invitrogen). The membrane was blocked with 5% non fat milk, washed with PBS-Tween 2- and probed with primary antibodies at the appropriate dilution. For detection, secondary antibodies conjugated to HRP was incubated with the membrane at the appropriate dilution for 1h at room temperature and immune complexes visualised by enhanced chemiluminescent detection system by FusionSL Vilber Lourmat system.
Smad4 Western analysis was performed using primary mouse monoclonal Smad4 antibody (Santa Cruz) at a 1:200 dilution and a secondary anti-mouse HRP-conjugated antibody at a 1:1000 dilution (Jackson Lab).

2.2.12 Preparation of total DNA

For preparation of genomic DNA from cells and tissue, DNeasy Blood and Tissue kit from Qiagen was used according to the manufacturer's instructions. 10 mg of tissue was lysed with kit provided lyse solution in tube containing ceramic beads in strong agitation with the Precellys Control Device (PeqLab). Genomic DNA was exacted by proteinase K treatment of the cells followed by ethanol binding of DNA. The cells were pelleted for 10 min at 300 g and re-suspended in 200 µl of PBS. They were lysed by adding 20 µl of proteinase K solution followed by 200 µl of lysis solution. The mixture was then vortexed and then incubated for 10 min at 70 C to achieve complete lysis. The lysate was prepared for binding by adding 200 µl of Ethanol and vortexing. The lysate was then

loaded onto a column with a silica membrane that binds DNA after treatment with ethanol and centrifuged for 1 min at 6500 g. The column was then washed twice and the DNA eluted in 100 µl volume.

2.2.13 Southern blot analysis

For DNA analysis total DNA was extracted using the DNA Blood&Tissue Extraction (Qiagen) and quantified using a NanoDrop 2000c (Thermo Fisher) spectrophotometer. For southern analysis total DNA (10-15 µg) were digested overnight with an appropriate restriction enzyme(s), mixed with 10x Loading Dye and separated slowly on 0.8% agarose gel at 20 mV overnight. The gel was immersed in 0.25 M HCl for 10 min, incubated twice for 15 min in depurinisation buffer followed by a 15 min incubation in neutralisation buffer. To transfer the DNA to nylon membrane by capillary action, the apparatus shown in Figure (3).

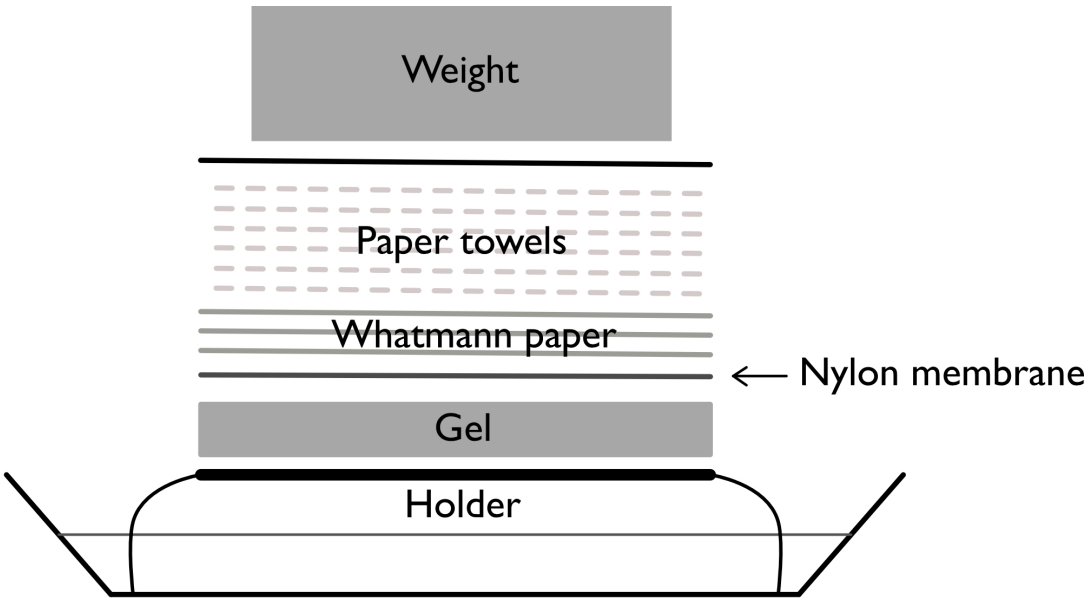


Figure 5. **Schematic representation of the Southern Blot apparatus**

The gel was supported on a layer of Whatman 3MM paper with a tank containing 20x SSC nucleic acid transfer buffer. A Hybond XL nylon membrane from Amersham Bioscience was soaked with buffer and placed on top of the gel, taking care to remove any bubbles. Once the paper towel were positioned, a weight was balanced on top, and the apparatus was left overnight to allow transfer to be completed. The following day the apparatus was disassembled and the nylon membrane was exposed to UV-radiation for one min to permanently cross-link the DNA to the membrane.
The GFP or the Luciferase gene was used to generate DNA fragments that were labelled with 32P (Prime-It II Random Primer Labelling kit, Agilent Technologies) and used as a probe. The hybridisation was performed in Church buffer at 65° C for 16 h.

2.2.14 Animal procedures

Animal work was carried out with the assistance of Corinna Klein, and with the support of the DKFZ animal staff. All surgical procedures were performed using sterile techniques.

2.2.15 Determination of copy number

Relative amounts of plasmid DNA in mammalian cells were calculated by real-time PCR (LightCycler 96, Roche). *Universal ProbeLibrary System Assay Design* software from Roche was used to design the oligonucleotide primers (Sigma) and the probe (Roche) for luciferase or gfp to determine amounts of S/MAR plasmid, and primers and probes specific for the human GAPDH gene to enable normalisation between the samples thorough calculating the number of cells used as the input.

Table 16. qPCR primers for copy number assay and Roche probes number

GAPDH For gctgcattcgccctctta	10
GAPDH Rev gaggctcctccagaatatgtga	10
GFP For cgacggcggctactacag	5
GFP Rev gtggatggcgctcttgaa	5

Amplification reaction (11 µl) contained 50 ng genomic DNA, 5.5 µl Roche Mix, 100 mM primers and 0.1 µl probe. Serial dilutions of the plasmids containing appropriate sequences to produce a standard amplification curve for quantification an all samples were tested in triplicate.

Step	Temperature - Time
1.Preincubation	95° C - 600 sec
2.Denaturation	95° C - 10 sec
3.Annealing - Extension	60° C - 30 sec

Repeat step 2-3 for 40 cycles.

2.2.16 Polymerase Chain Reaction (PCR) for the Amplification of DNA fragments

The amplification of a desired DNA fragments was carried through PCR in a PeqStar (PeqLab) thermocycler. For the design of the oligonucleotide primers, it was used the software *SnapGene*. Normally in a final volume of 25 µl, the amplification reaction contained 1-10 ng of template DNA, 10 µM primers and 12.5 µl HiFi PCR Premix (Clonetech).

Step	Temperature - Time
1. Denaturation	95° C - 2 min
2. Denaturation	98° C - 10 sec
3. Annealing	Primer Tm – 5° C - 10 sec
4. Extension	72° C - 5 sec/Kb
5. Final Extension	72° C - 10 min
6. Storage	8° C

Repeat step 2 to 4 for 30 cycles

The desired fragments were processed afterword with the PCR Purification Kit (Qiagen) for removal of residual contaminants.

3 Results

3.1 Establishing a stable genetically modified cell line with S/MAR DNA vectors

Cells can be persistently labelled with S/MAR DNA Vectors and this can be defined as an established population when the cells that it comprises harbour at least one copy of the DNA vector and actively express any transgene it carries. The establishment process is represented schematically in Figure 6. Upon DNA delivery, the cells are treated with an antibiotic which allows the selection of those that have the vector and eliminates any that are un-transfected. The duration of the selection pressure is cell and antibiotic-dependent and it can range from 2-3 up to 14 days, with regular medium change.

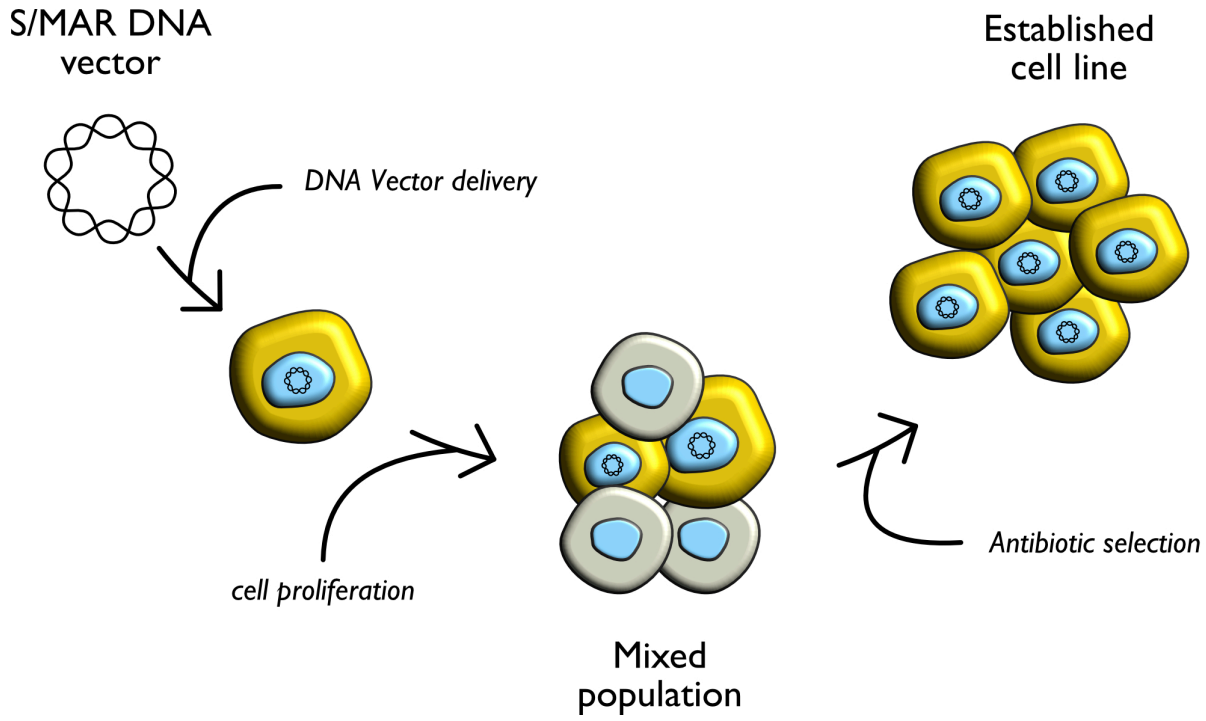


Figure 6. **Schematic representation of the workflow for establishing cells with an S/MAR DNA vector.** DNA vectors containing a S/MAR sequence are delivered to target cells in a process called transfection. Upon delivery, cells are grown for a few days and then treated with antibiotics that allow the selection of those which have the vector and that induce the cell death of those that are not transfected. The resulting population is composed by cells which presents at least one copy of the vector and that express the gene of interest.

3.2 Illustrating the limitations of the prototype pEPI-UCOE DNA vector system

Hek293T cells were established following the process described in Figure 3 with the current state of the art vector published from *Hagedorn et al.* (Hagedorn, Antoniou et al. 2013). Upon transfection with pEPI-UCOE the cells were selected with the antibiotic Neomycin (G418) for 14 days, as described in the publication. In this vector system the CMV promoter drives the expression of the reporter gene GPF. Its expression was used to monitor the cells. Surprisingly, the establishment procedure generated a HEK293T cell population in which the majority of the cells were not positive for the transgene expression; the majority of the cells modified with pEPI-UCOE were negative (Figure 7) but resistant to antibiotic pressure.

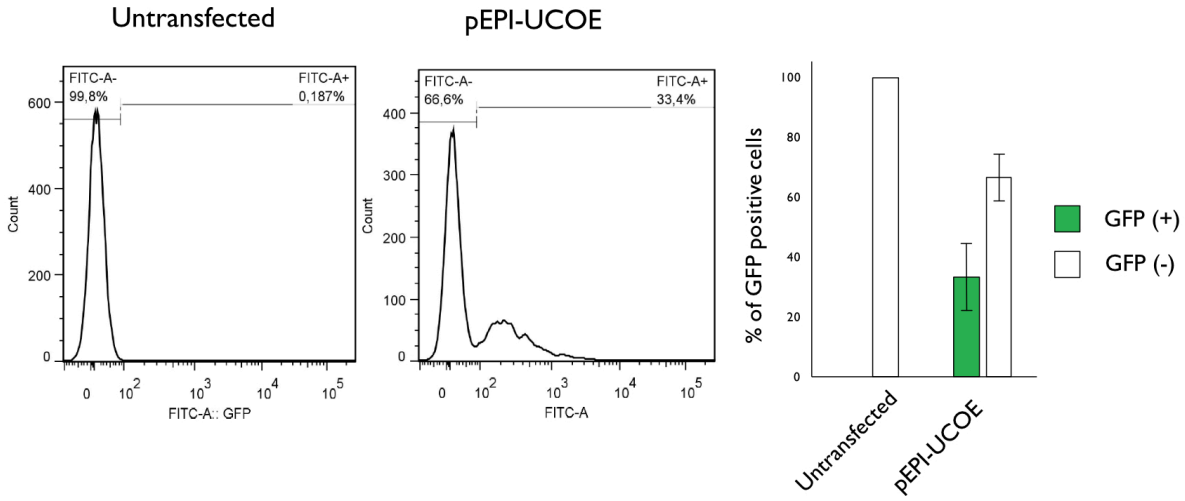


Figure 7. **Analysis of Hek293T cell line establishment with the prototype S/MAR based DNA vector pEPI-UCOE.**
FACS analysis on Hek293T cells established with the vector pEPI-UCOE reveal that after 14 days of constant selection with the antibiotic Neomycin only a small proportion of the population is positive for the transgene expression (FACS histograms). The percentage of positive cells is summarized in the right panel as the average of three independent experiments and the error bars show the standard deviation. The establishment rate of the plasmid pEPI-UCOE is inefficient and most of the cells in the population don't express the transgene.

To investigate the reason for the low establishment efficiency, a plasmid rescue assay was performed. The scope of this analysis was to characterise the vector's molecular status in these cells. To determine whether the vector was present but not active or, if it was lost during the cells' replication, total DNA from established cells was isolated and used to transform electro-competent bacteria. The assay is based on the principle that only circular non-integrated DNA that carries a bacterial antibiotic resistance marker is able to transform prokaryotic cells and sustain their growth under antibiotic pressure. A representative number of colonies were grown, plasmid DNA was extracted and it was subjected to restriction analysis (Figure 8). The restriction patterns, generated from the digestion of plasmids extracted from the bacterial colonies, were compared to the those created with pEPI-UCOE. Surprisingly, the rescued vectors looked smaller and also not completely digested. These findings lead to conclusion that vector rearrangements occurred in the established HEK293T. This could also be the main cause for the low percentage of GFP expressing cells. However, although the cells were not positive for the transgene expression, the vector was still kept in its episomal form and it was not integrated. The plasmids were able to sustain the bacterial growth in solid and liquid medium, which meant that they carried the kanamycin resistance gene. The plasmid rescue assay and the FACS analysis supported the hypothesis of rearrangements. In 33.3% of the cells the vectors were able to sustain the expression of the GFP. However, in the remaining 66.6%, the cassette responsible for the expression of the transgene was missing but in the episomes, the gene responsible for the antibiotic resistance was still present and they were able to transform bacteria. In the pEPI-UCOE vector platform the mammalian and bacterial selection are generated from the same expression cassette.

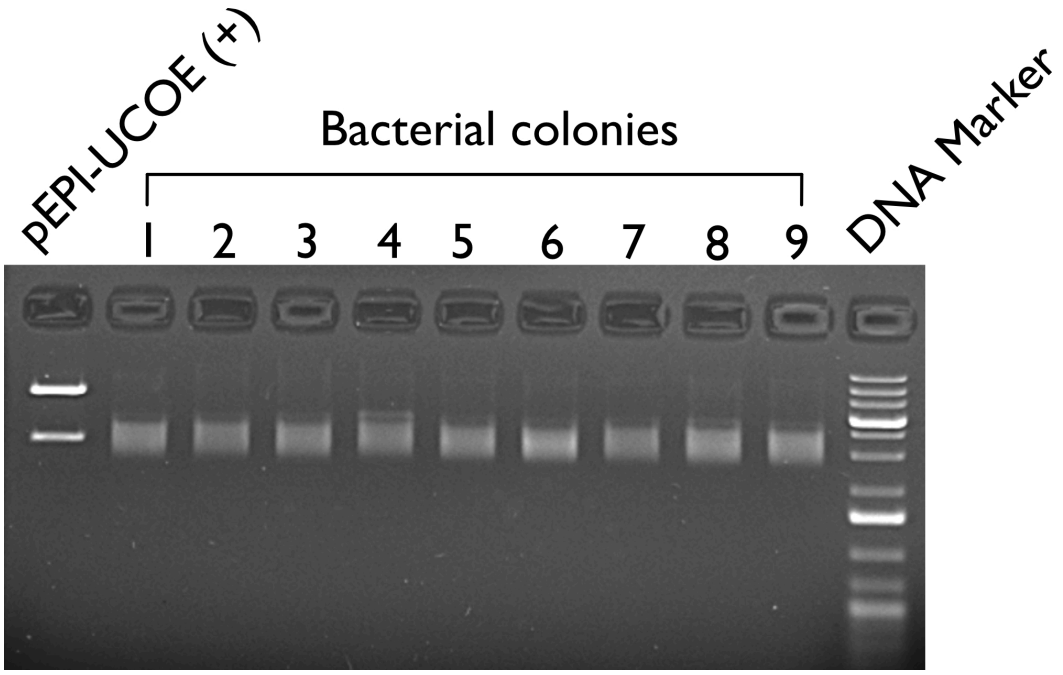
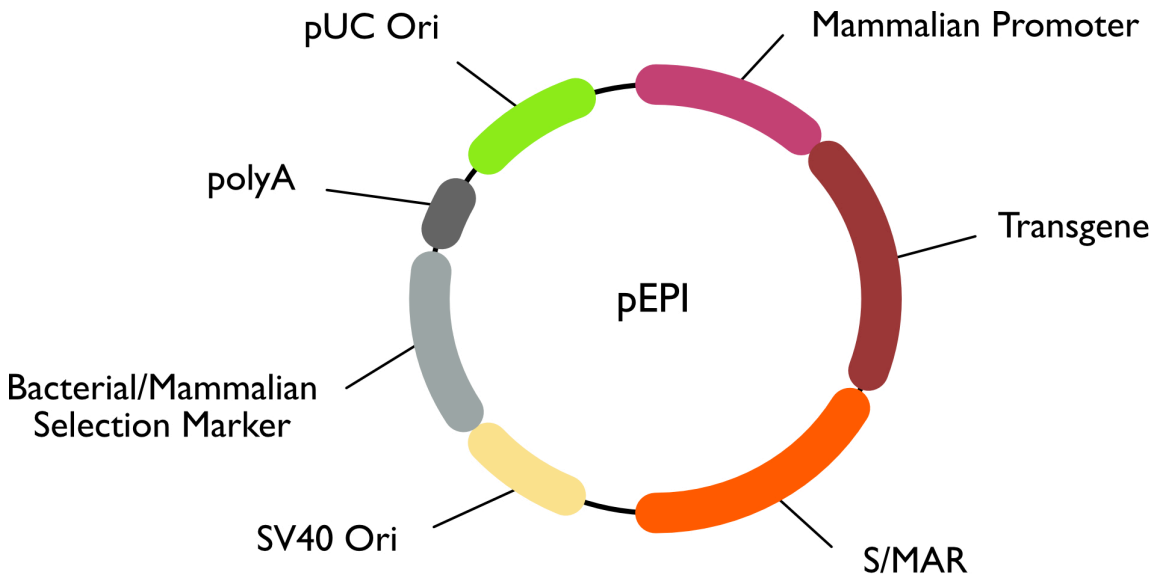


Figure 8. **Plasmid rescue and restriction analysis following bacterial transformation of total DNA derived from Hek293T cells established with pEPI-UCOE.**
9 colonies derived from the transformation of bacteria with total DNA extracted from HEK293T cells established with the vector pEPI-UCOE were grown and plasmids were isolated with the MiniPrep Kit (Qiagen). The DNA obtained from the bacterial colonies and DNA from the Maxi preparation used to transfect the cells were digested with the enzyme *BamH*I. The DNA fragments were resolved on an 1% agarose gel and the restriction patterns compared. The analysis showed that none of the DNA extracted from the bacterial cells had a restriction pattern that is similar to the one generated with the control vector

To test the molecular integrity and functionality of the rescued plasmids, Hek293T cells were re-transfected but no transgene expression was detected, thus confirming the defective nature of the rescued vectors. (data not shown).

3.3 Development of the next generation pS/MARt DNA Vector

The low efficacy in establishing cells and the stability issues that arose using pEPI-UCOE, ushered in the development of a new S/MAR vector platform. Minimally sized DNA vectors such as minicircles have previously been demonstrated to be more efficient than bacterial plasmids, but due to commercial restrictions it was decided to develop a vector system which could be generated and evaluated in the DNA Vector Lab without restriction. The originally described pEPI plasmid vector was extensively studied over the past decade and it was known that, in order to keep the plasmid episomal the S/MAR motif had to be part of a transcriptionally active cassette. In this vector the functionality of the S/MAR was coupled with the transgene expression but it was totally independent from the resistance marker. In a rather simple but significant modification of the plasmid the activity of the S/MAR was decoupled from the expression of a gene of interest (GOI) and linked directly to the expression of a drug resistance gene (Figure 9). Together with this innovation Puromycin (Puro) was chosen instead of Neomycin (Neo/G418) primarily because the selection procedure is more efficient and the non-established cells could be cleared more quickly. The bacterial backbone of the original vector was also swapped for a more modern minimally sized one. This new vector platform was called pS/MARt. The link between the mammalian selection marker and the transgene should also avoid the instability issue noticed in pEPI, reducing therefore the number of resistant but non transgene expressing cells.



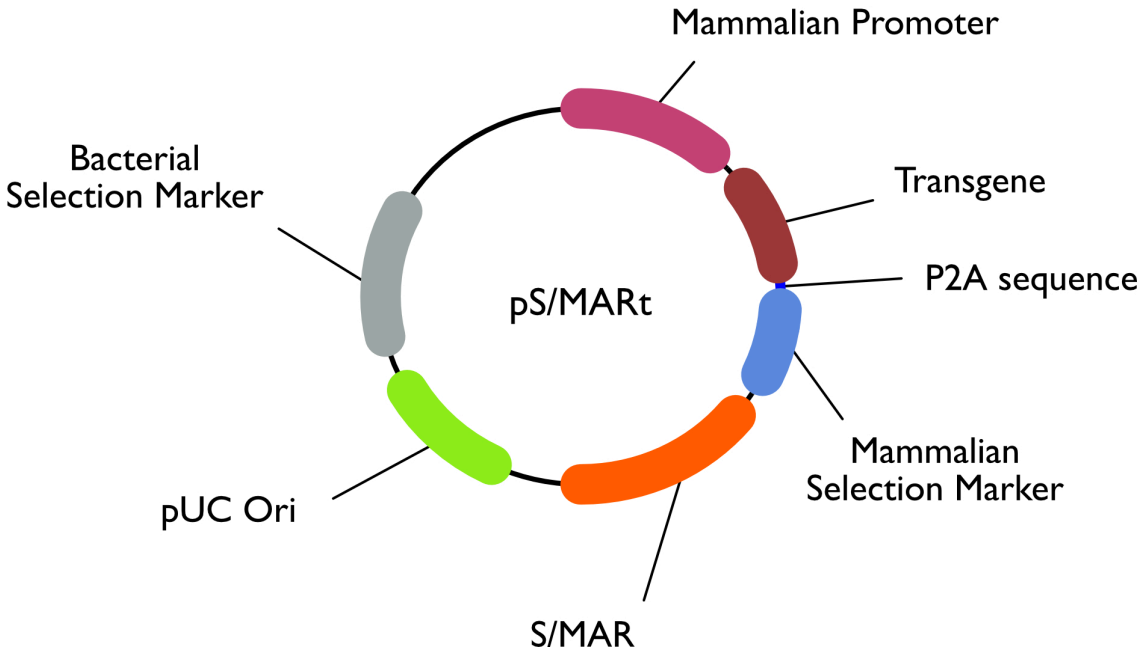


Figure 9. **Cartoon representation of pEPI and the next generation pS/MARt DNA Vectors.** The vector system pEPI present a large bacterial backbone which contains the hybrid bacterial/mammalian selection marker Kan/Neo that allows the propagation in bacteria as well as the section in mammalian cells. A pUC Ori is necessary for the replication in bacteria and the SV40 Ori is present for the replication in mammalian cells. A second expression cassette is present in the vector. The CMV promoter drives the expression of a GOI and this transcription activity is responsible for the functionality of the S/MAR sequence. In the pS/MARt vector system, the expression of the mammalian selection marker is linked to the expression of the GOI and actively also sustain the functionality of the S/MAR. The mammalian selection gene Puro is linked molecularly to the GOI and the bacterial backbone is minimised. It presents only the pUC Ori and the Kan resistance gene for the selection and propagation in bacterial cells.

To test the functionality of the new S/MAR vector three different variants were created: one in which the Puro expression was linked via a P2A sequence to the reporter gene GFP, the second, where a IRES sequence linked the Puro and the GFP and a third in which, a fusion protein was generated. The P2A sequence creates a direct molecular link between the first and the second gene. This sequence was described as a "self-cleaving" small peptide and it was first identified by *Ryan et al.* (Ryan, King et al. 1991) in the foot and mouth disease virus (FMDV). The length of the P2A peptide is 18-22 amino acids. The designation "2A" refers to a specific region of the picornavirus polyproteins. During the translation of the polyprotein mRNA, the ribosomes skip of the glycyl-prolyl peptide bond at the C terminus of the 2A peptide leads the cleavage of the peptide and form two proteins with a ration of 1: 1. The link represented from an Internal Ribosome Entry Sequence (IRES) do not generate such protein equimolar ration. The IRES sequences are distinct regions of RNA molecules that are able to recruit the eukaryotic ribosome (Pelletier and Sonenberg 1989) and induce the translation of the protein in a cap-independent manner. The binding of the ribosomes to the IRES is rather random and normally the expression of the second gene is ~20% of the first. All the vectors were tested in HEK293T and compared to pEPI-UCOE. FACS analysis revealed that cells established with the version of pS/MARt carrying the P2A linker (Figure 10, c) and the IRES (Figure 10, d) generated ~97% of GFP expressing cells and an insignificant proportion of negative cells (Figure 10, e), whereas the positive cells established with pEPI numbered only 33% (Figure 10, e). The version of pS/MARt that was generated to produce the fusion protein between the GPF and Puromycin did not function as expected and was able to establish only 9% of the cells and it was not considered for further experiments. It was hypothesised that the peptide generated by the fusion between the Puro and the GFP was capable of providing antibiotic resistance but it was not able to form a structurally active GFP tetramer. The cell populations established with pS/MARt produced a higher (Figure 10, g) and more homogenous

expression profile than the one generated with the pEPI system whose GFP expression varied widely with a significant contamination of non- and poorly expressing cells.

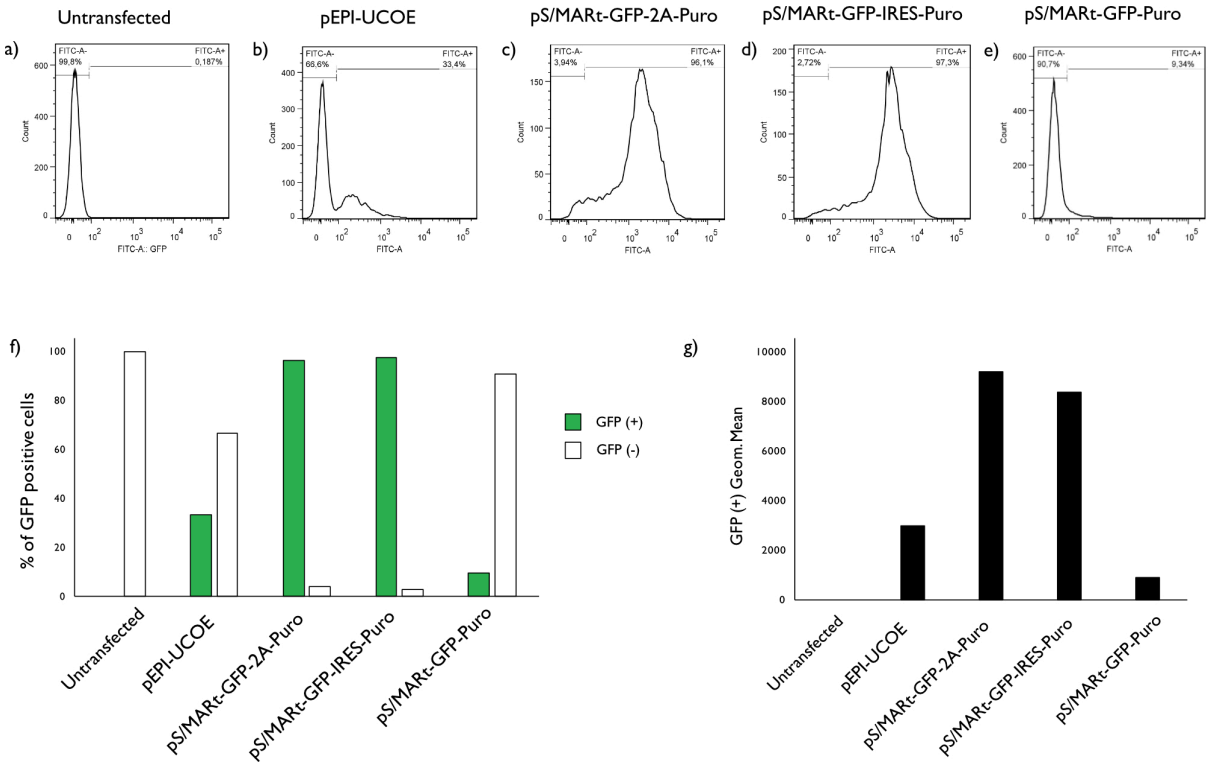


Figure 10. **HEK293T populations established with pEPI and different versions of pS/MARt**. FACS analysis show that pS/MARt established a higher number of cells (c-d) when compared to the (a) untreated control and pEPI (b). Not all the version of pS/MARt are efficient in the process. The version that carries the fusion protein GFP-Puro doesn't establish as well as the other two. (f) The green bars represent the percentage of cells that express the reporter gene GFP and are considered established. The white bars represent the non-expressing and therefore non established cells. (g) The geometric mean of the expression of the different established populations demonstrates that pS/MARt-GFP-2A-Puro and pS/MARt-GFP-IRES-Puro are better not only in terms of establishment rate but also in terms of transgene expression within the populations.

3.4 Improving cell establishment using pS/MARt DNA Vectors

Based on the work carried out by Zhang *et al*. and Ted H.K. Kwaks *et al.* (Kwaks, Barnett et al. 2003) two new variants of the plasmid were generated in order to test two different insulator sequences. In independent papers the authors demonstrated that the presence of insulator elements (UCOE and Element40) improve cell establishment as well as transgene expression when used in viral systems. The insulators were cloned before the mammalian promoter generating the vectors: pS/MARt-Ele40-GFP-2A-Puro and pS/MARt-UCOE-GFP-2A-Puro. Our idea was to protect the mammalian expression cassette from any possible influence induced by the bacterial backbone. A colony forming assay was performed to compare the different versions of pS/MARt to pEPI (Figure 11).

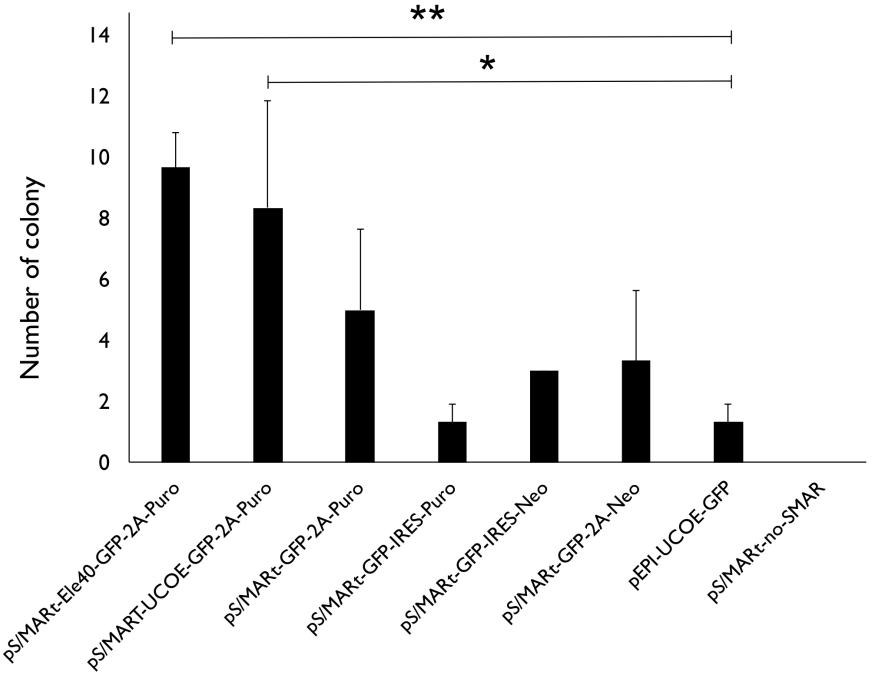


Figure 11. **Colony forming assay.**
For each construct 100 positively transfected cells were plated into a 10cm tissue culture dish after FACS sorting and cultured in the presence of the appropriate antibiotic (0.5 ug/ml Puro or 0.75 mg/ml Neomycin). After 4 weeks the developing colonies were fixed, stained and quantified. The introduction of genomic elements in the pS/MARt vector generated a significantly higher number of colonies compared to the pEPI vector prototype (t test has been used to generate the statistical comparison between the vectors ** $p < 0.005$, * $p < 0.05$)

In order to determine whether the establishment procedure was influenced or not by the antibiotic selection, corresponding Neo/G418 versions of the plasmids were also produced.
The introduction of the genomic insulator sequences before the mammalian expression cassette improved the efficiency of cell establishment. The colony forming assay also revealed that the P2A link between the reporter and the selection marker was more efficient than the IRES sequence. The vector based on the P2A sequence which creates a direct link at the transcriptional and translational level between transgene and selection marker, was selected for further studies. Since the insulator sequence significantly improved the number of the resistant colonies, the version of pS/MARt that comprised the anti-repressive Element 40 was elected for future analysis. We discontinued development of the UCOE version although it was reasonably efficient due to previous work using this genomic sequence and due to restrictions in its use (Patent number: WO 2002024930 A2).

3.5 Vector development: from pEPI to pSMARter

The aim of this thesis was to develop a new vector system that could be widely applied in a range of research applications without restriction. We also wanted to produce a vector that was easy to manipulate and simple, cheap and quick to produce. The continued development and refinement of this vector technology continued for the duration of this project. The application of the DNA vector system in increasingly intriguing and difficult cells and projects challenged the vector's capabilities and as the complexity of the experiments increased the development of novel, more modern versions of the vector was required to overcome the apparent limitations of the previous version. For obvious time limitations not all the experiments could be performed with the most up to date version of the plasmid. This section will provide an overview of the DNA vector

development that commenced with derivatives of the original pEPI vector in June 2014 and ended with the latest generation of the pS/MARt(er) DNA vector in August 2017 (Figure 12). In the following sections the application of the DNA vector systems and their advantages and disadvantages will be described. The ultimate goal of this project was to generate a platform which might be used for the efficient genetic modification of primary human material for personalized medicine and gene therapy.

The pEPI vector system, as previously described, with its intrinsic instability couldn't be considered a suitable tool for any of these "dream" applications. The first version of pS/MARt was created removing from pEPI a fundamental feature (f1 Ori) for the episomal maintenance and replication in mammalian cells and the Neomycin resistance gene. The removal of these features allowed us to use a minimally sized and more modern bacterial backbone where the minimal SV40 promoter draw the expression of the resistance gene Kanamycin for the selection in prokaryotic cells and the pUC Ori allowed a high copy maintenance and replication. In the pS/MARt vector the selection in eukaryotic cells was carried out through Puromycin (Puro) resistance which provides a quicker and more effective selection pressure. The expression of the Puro was directly linked to the expression of the GOI and the S/MAR functionality. A genomic insulator was also introduced in pS/MARt.1 to shield the eukaryotic expression cassette from the bacterial backbone. This vector was shown to be more efficient in generating established cells
(Figure 11).

3.6 Development of minimally sized Nano-S/MARt DNA Vectors

Kay et al. (Chen, He et al. 2003) demonstrated that the *de novo* methylation at discrete CpG sites in plasmids bacterial backbone is the main responsible for the silencing of the vector. It is suggested

that inactive chromatin structure can spread from methylated regions or can generate small interfering RNAs (RNAi) which can lead to transcriptional inhibition. The minicircle technology (Bigger, Tolmachov et al. 2001, Vaysse, Gregory et al. 2006) overcame these problems generating vectors that lacks the bacterial components.

To further improve the DNA vector system that could potentially be used for clinical application we developed a range of Nano-S/MARt (NP-S/MARt) vectors in collaboration with Nature Technology™ (NTX). This new class of minimally sized vectors which can be produces without extraneous bacterial sequences is based on pS/MARt.1 but presents a profound change in the bacterial backbone. The technology developed by NTX produces nano-plasmids which are very similar to minicircles. The manipulation and preparation of NP-S/MARt is discussed in detail in a dedicated paragraph below. The new nano-vector technology has all the advantages previously discussed for minicircle and many more. The first and potentially most important for future clinical application, is that the composition of the vector is approved by the Federal Drug Administration (FDA) for use in humans. The NP technology also introduced the capability to work with more delicate and difficult cells, like human patient derived and/or human primary material.

The removal of the bacterial backbone was followed by another step of vector development with the aim of producing an expression cassette that could mimic a human endogenous gene. In a dedicated paragraph below, it will be described in detail the production of the Nano-S/MAR-splice vector, in which the transgene presented an intro-exon like structure with the purpose of stabilizing the transcript.

3.7 Replacement of the S/MAR motif

Ultimately, in the last step of development, the S/MAR element, the most fundamental and important keystone feature of this class of vectors was replaced. The S/MAR sequence isolated

from the human β-globin gene in 1999 and since then considered an irreplaceable feature of this vector system was replaced with a novel, smaller and more efficient one, generating the vector pS/MARter. The vector pS/MARter, although still in plasmid form and not yet in a nano-format, by any measure outperforms every other vector previously described including the nano-vector derivatives. Thus, it can be considered the ultimate step in this current process of DNA vector development.

As new vectors were developed over the year newer constructs were always compared to previous versions in order to determine whether the new modifications were providing any benefit and producing any desired effect.

In a final experiment all the vectors generated during this project were tested and compared for a range of functionality including: their efficacy in generating stable cells, their transgene expression in established populations, the molecular integrity of the plasmids, the relative RNA levels and the number of copy plasmids necessary to establish cells. HEK293T cells were transfected with the plasmid displayed in Figure 12.

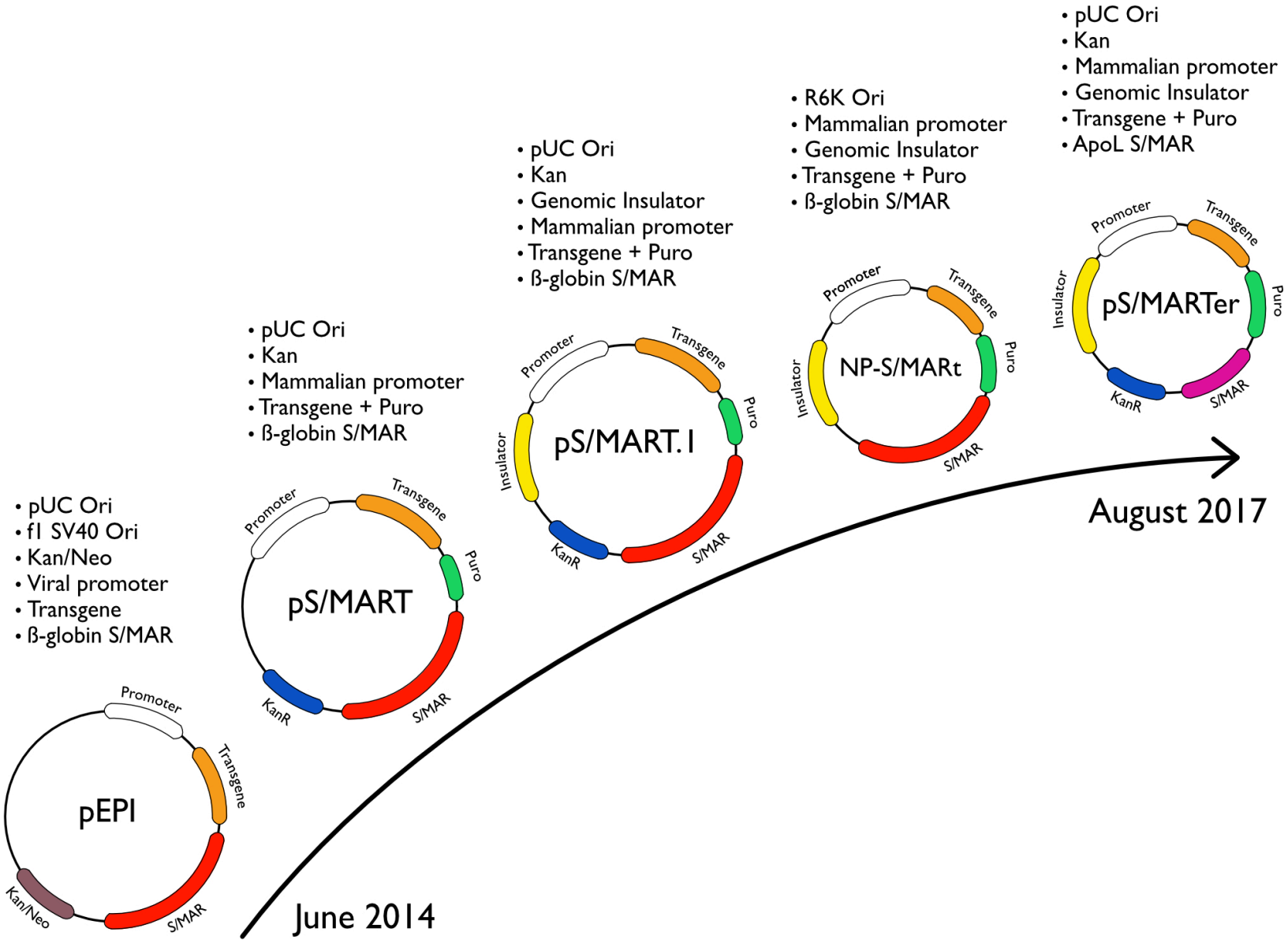


Figure 12. **Schematic representation of the vector development over time.**

Each of the different forms of the vector comprised the same mammalian cassette with a typically used CMV promoter driving the expression of the reporter gene GFP. The translation of the GFP gene was coupled via a P2A sequence to the eukaryotic selection marker Puromycin. The S/MAR sequence was placed after this transcription unit. The genomic insulators UCOE and anti-repressive element 40 (Ele40) were used as boundaries between the bacterial backbone and the expression cassette.

The pS/MARt series of plasmids, Nano-S/MARt and pS/MARter were compared to previous best-in-class pEPI-UCOE. Cells were monitored weekly for GFP expression via FACS and 38 days post transfection the RNA and DNA were extracted and analyzed. The establishment rate was also evaluated by colony forming assay (Figure 13).

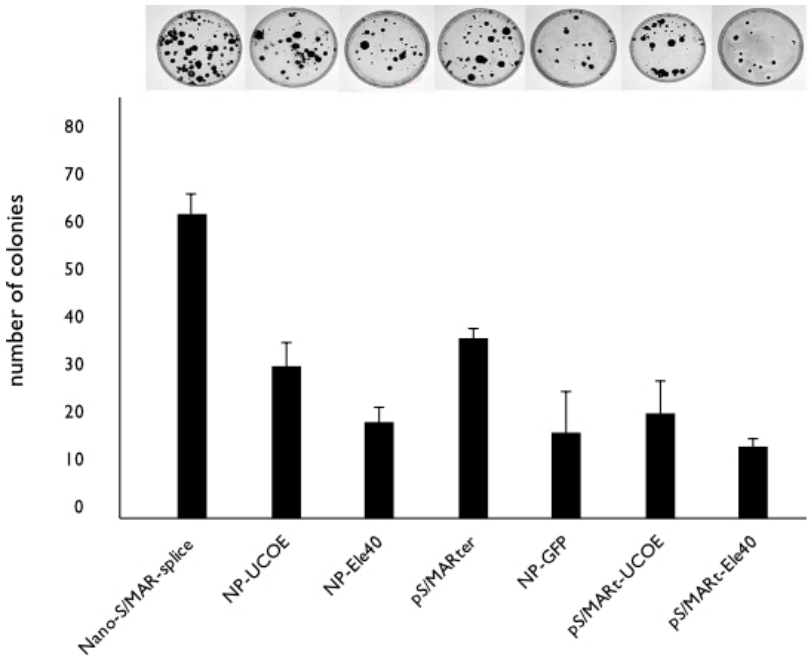


Figure 13. **A comparison in colony forming efficiency between all the DNA vectors generated in this study.**
The graph reports the number of colonies obtained with the different version of the vector. Upon DNA delivery, cells positive for the expression of the reporter gene GFP were isolated via FACS sorting (FACS Aria II) and 100 cells were plated into a 6 cm cell culture dish. The cells were then cultured for 4 weeks in presence of 0.5 μg/ml Puromycin. After 4 weeks the developing colonies were fixed with PFA and stained with Crystal Violet. The results are expressed as the average of 3 independent experiments. A representative picture of the experiments outcome is displayed above the graph. The Nano-S/MAR splice vector generates the highest number of colonies. The plasmid pS/MARter although carries a bacterial backbone performed better than the other Nano vectors and pS/MARts.

24 hours after transfection GFP expressing cells were isolated via FACS sorting (FACS Aria II) and 100 cells were plated into a 6 cm dish. For each vector, three replicates were produced. The cells were cultured for 38 days in presence of 0.5 μg/ml Puro when transfected with pS/MARt based plasmids and 0.75 mg/ml Neo for those transfected with pEPI. The resulting colonies were fixed with PFA, stained with Crystal Violet and counted to address the vectors establishment efficacy. The number of colonies and pictures representative of the experiment outcome is reported.
The Nano-S/MAR-splice vector generated the highest number of colonies followed by its non spliced version NP-UCOE and NP-Ele40. pS/MARter, harboring the ApoL MAR instead of the β-interferon one, had an establishment rate that was similar to the Nano plasmids although it sill presented a normal bacterial backbone. pS/MARt-UCOE and pS/MARt-Ele40 presented a lower efficacy. The result showed that the β-Interferon MAR was less efficient that the new ApolipoproteinB MAR non nano-vectors. The splicing of the S/MAR resulted in more colonies probably because the removal of an unstable sequence from the pre-mRNA improved the message stability. Cells established with this vector are more prone to tolerate the selection pressure.
The number of colonies established from pEPI is undoubtedly higher than any of the S/MARt vectors (Figure 14).

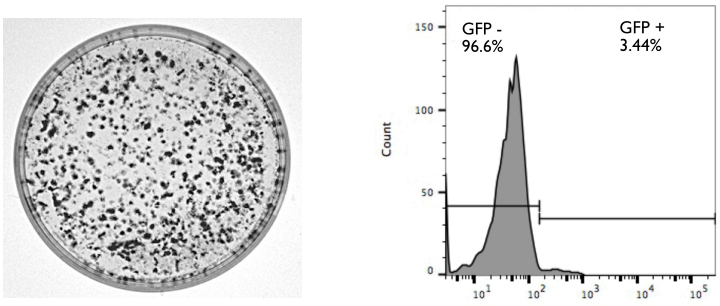


Figure 14. **Established colonies with pEPI-UCOE and FACS analysis.**
The established cells with the vector pEPI-UCOE were fixed in PFA and stained with Crystal Violet. One of the plate replicate was used to assess to number of GFP expressing cells in the established colonies. FACS analysis on the right shows that only a small proportion of the population is positive for the transgene expression.

Although this system looks superficially efficient, FACS analysis on the population reveled that only a small percentage of the cells were positive for the expression of the transgene GFP. A great number of colonies were formed but only few, if any, were real established ones. The molecular integrity of the new plasmid-generations was assessed with Southern Blot analysis and a plasmid rescue assay was performed on those vectors that had a bacterial backbone.

The genomic DNA of the established HEK293T populations was extracted and digested with the restriction enzymes *Bam*HI. This enzyme acts as a single cutter for all vector species and its cleavage site (GGATCC) is recurrent in the human genome. The genomic DNA was digested, the DNA fragments were separated on an agarose gel and analyzed via Southern Blot. In parallel the DNA maxi-preparation used to transfect the cells were digested with *Bam*HI and used in the Southern Blot to control the plasmids size. The Southern Blot (Figure 15) showed that all the vectors isolated from mammalian genomes had the same size of their respective controls.

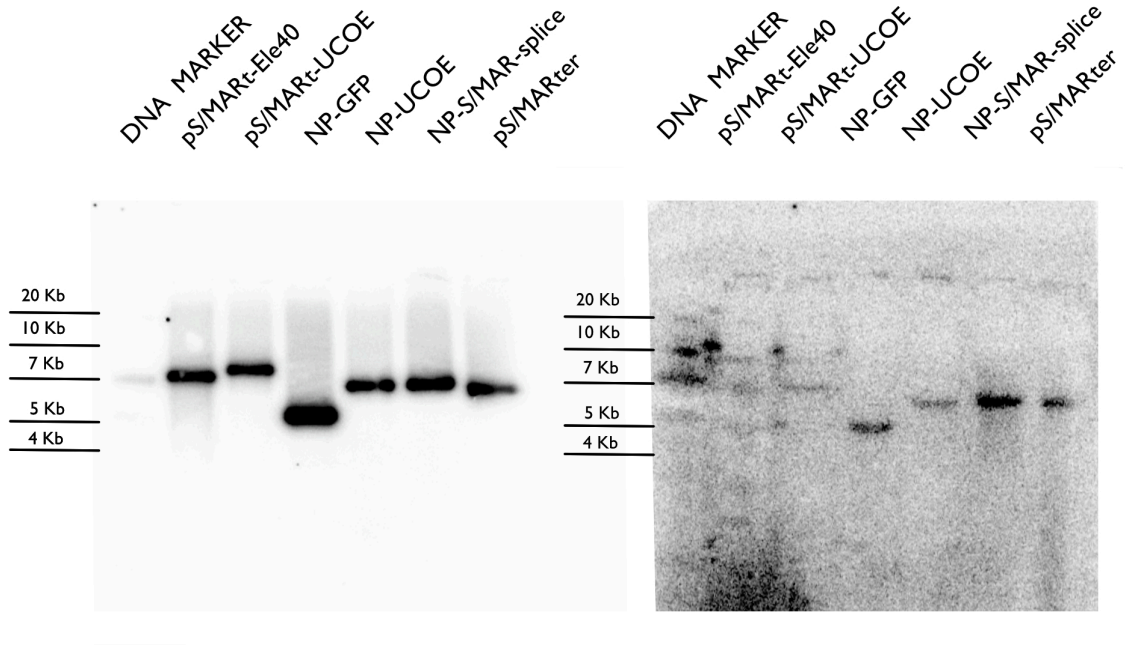


Figure 15. **Southern Blot analysis of pS/MARt vectors, Nano-S/MARt and pS/MARter.** Total DNA from Hek293T established with different versions of the S/MAR vector was extracted (Blood&Tissue DNA Kit, Qiagen) and subjected to digestion with the restriction enzyme Bam*HI* (NEB) for 12 h at 37°C. The DNA fragments were resolved on a 0.8% agarose gel and transferred on a nylon membrane. Simultaneously, plasmid DNA from the maxi preparations used to transfect the cells before establishment were treated with the same approach. The reporter gene GFP was used to generate the radioactive probe for testing the controls (left panel) and the vectors in the cell population (right panel). All the plasmids have the same size when compared to the correspondent reference vector which demonstrate their episomal maintenance.

The absence of smears and/or additional bands demonstrated also that the majority of the SMAR plasmids were kept episomally. This analysis showed that these vector didn't integrate and didn't undergo clear rearrangements. Although the episomal *status* was clearly confirmed by Southern Blot, only more detailed integration analysis could show whether rare integration events took place or not. The integrity of the plasmids pS/MARt-UCOE, pS/MARt-Ele40 and pS/MARter was investigated via plasmid rescue. Electro competent bacteria were transformed with total DNA extracted from HEK293T established with these vectors. Restriction analysis on pDNA extracted from bacterial colonies revealed that the rescued vectors (Figure 16) had the same size of the correspondent plasmid used to transfect the cells.

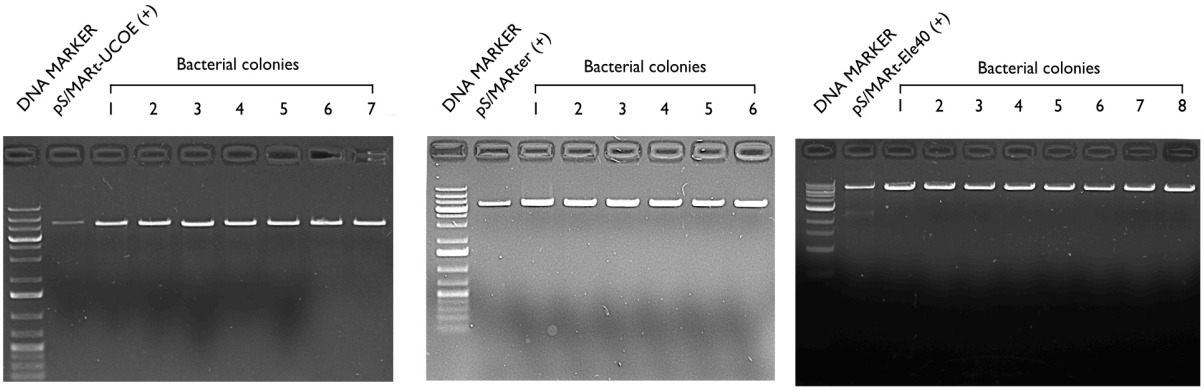


Figure 16. **Plasmid rescue analysis of pS/MARt-UCOE, pS/MARter and pS/MARt-Ele40.** Genomic DNA from HEK293T cells established with pS/MARt-UCOE, pS/MARter and pS/MARt-Ele40 was transformed into *E.Coli* DH10B (Invitrogen). Upon transformation the bacteria were plated on LB-Agar plates with additional Kanamycin and incubated over night at 37 °C. Colonies were picked and grown for 12 h in LB medium and Kan before the plasmid DNA was extracted (Qiagen, Miniprep Kit). For the analysis, the plasmid DNA from the colonies was digested with the enzyme BamHI. As a control also the maxi preparation used to transfect the Hek293T was digested with the same enzyme and run on agarose gel as reference. All the rescued vectors had the same size of the correspondent plasmid used as reference control.

The expression of the reporter gene was monitored weekly for more than one month (Figure 17). The analysis showed that between DNA delivery (day 1) and the end of the selection process (day7) the number of GFP expressing cells dropped dramatically. In this time window, the non-

transfected and the not established cells were killed and only those able to maintain the vector grew. From day 7 to day 14 the cells underwent the expansion phase and at this stage, they were mostly positive for the expression of the transgene. The analysis of the median fluorescence intensity in the different populations, revealed that the Nano-S/MAR-splice vector had the highest expression profile, while cells established with the ApoL MAR presented the lowest.

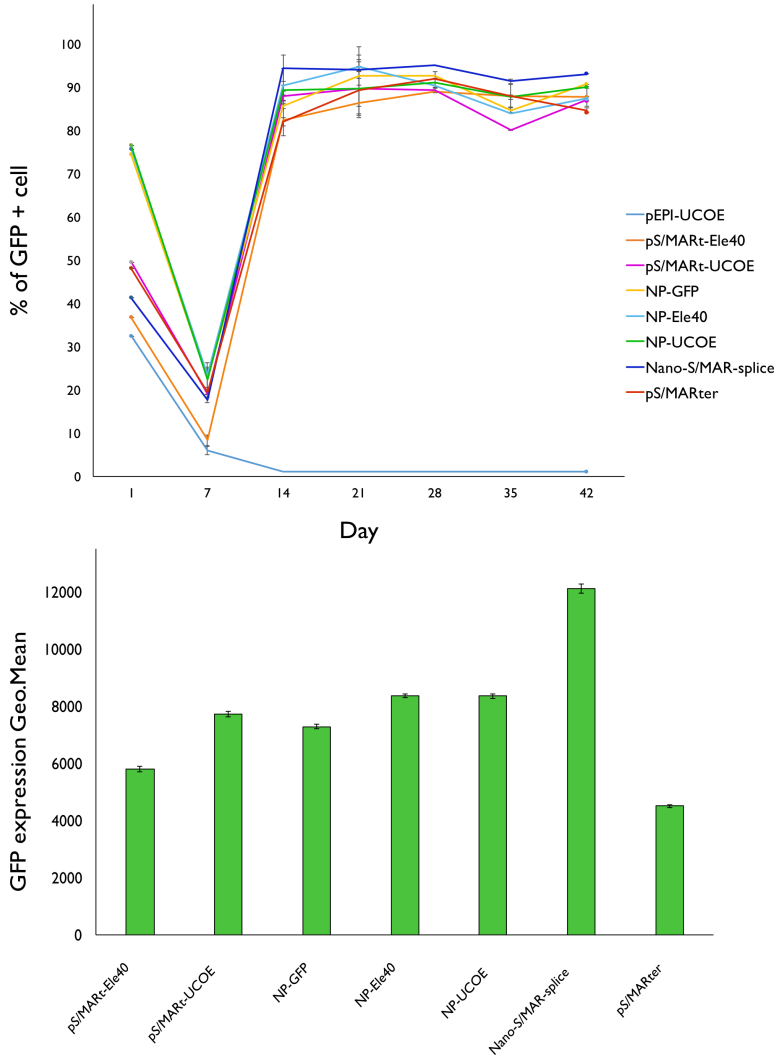


Figure 17. **Reporter gene analysis over time and expression in the different populations.** Hek293T transfected with different version of pS/MARt were followed over time and the number of cells positive for the expression of the reporter gene GFP was addressed weekly (upper panel) via FACS (Lsr, Fortessa). At day 42, the expression within the populations was estimated (lower panel). The values in both representations are the average of the results obtained in 3 independent experiments. All the plasmids except pEPI generate cell populations in which almost all the cells express the reporter gene. The cell population established with the Nano-S/MAR-splice vector presents the highest GFP expression

The expression monitored by FACS was evaluated at the RNA level via quantitative real-time PCR (q-PCR) that confirmed the results previously displayed (Figure 18). When compared to the expression of the endogenous gene GAPDH, the RNA levels of the cells established with the Nano-S/MAR-spliced vector are the highest, followed by the Nano vector series and the three different versions of the episome that carry a bacterial backbone.

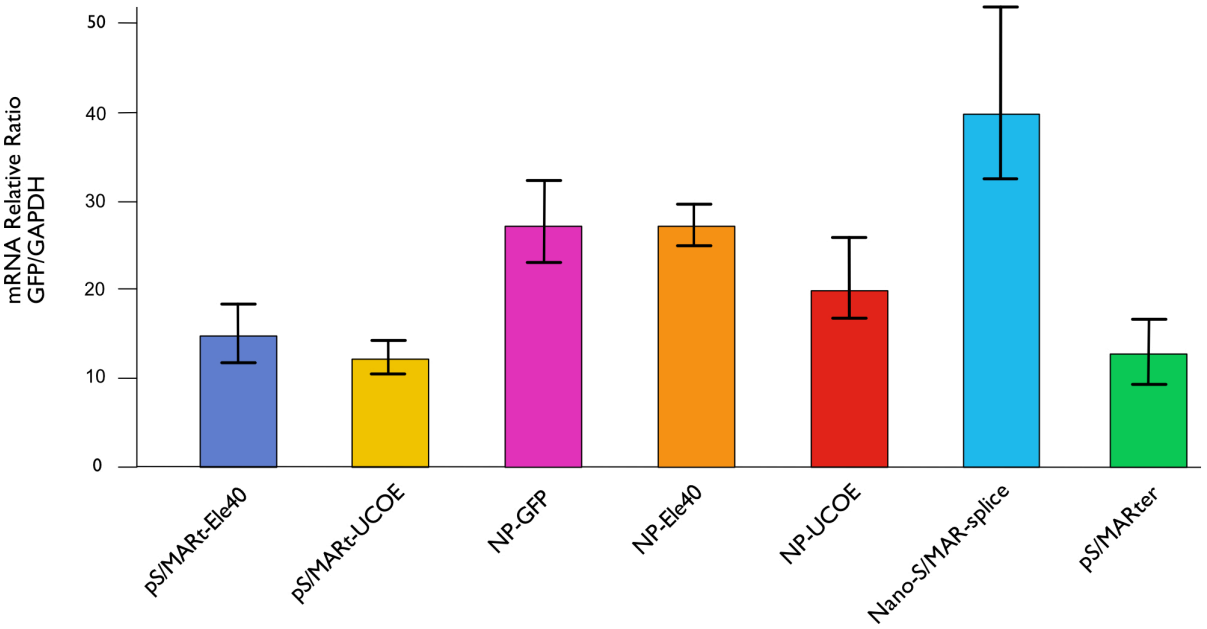


Figure 18. **Relative RNA expression values in HEK293T established with different versions of the vector.**
Total RNA was extracted from established Hek293T populations (RNAeasy Kit, Qiagen) and via quantitative Real-Time PCR (qPCR) the GFP RNA relative levels were calculated using the endogenous gene GAPDH as reference. The Nano-S/MAR-splice vector sustained the highest transgene expression. The other nano plasmids display expression levels that are higher than the vectors with a normal bacterial backbone.

3.8 Application of the new vector system

3.9 Generation and *in vitro* validation of Luciferase labelled Isogenic pancreatic cancer cell lines

To validate the efficiency of the new plasmid in modifying cancer cells, the reporter gene GFP was swapped for Luciferase (luc). The primary advantage of utilising Luciferase over GFP is the possibility to visualise expressing cells through Bioluminescent imaging (BLI). This technique allows non-invasive sequential imaging of *in vivo* transgene expression. In these experiments it was utilised to monitor tumour development and early metastasis reducing the number of animals used per experiment. As part of a collaboration within a European Consortium a panel of different pancreatic cancer cell lines were labelled using our vector system and used as a tool for *in vitro* and *in vivo* screening of drug compounds. The labelling was performed with the non-integrative S/MAR vector system over other technologies because most of the cell lines were refractive to viral transduction. The pancreatic cancer cell line Capan-1 was used to compare the establishment efficiency of the pS/MARt and pEPI vectors. 12 weeks of repetitive rounds of Neomycin selection were necessary to generate modified Capan-1 cells with pEPI (Figure 19,a), whereas the establishment with pS/MARt.1 took only 5 weeks and a single administration of Puro (Figure 19, b). Human BxPC-3, MiaPaCa-2, Capan-1 and Panc-1 cell lines were also established with the vector pS/MARt.1-Luciferase. The reporter gene expression in the modified lines was evaluated in accordance with the number of cells (Figure 19, c). As demonstrated by a linear relation, all the components of the different populations contributed to the Luciferase expression. These cell lines were created with the aim of generating xenograft models and this validation excluded the possibility of mosaicism in their further applications *in vivo*. For a more thorough evaluation, the luc-labelled cells growth in presence and absence of gemcitabine (a clinically approved pancreatic cancer drug) was compared to the match parental unmodified cells. They showed *in vitro* a

behaviour that was similar to the respective parental controls (Figure 19, d). The growth curves demonstrated that, even though the cells were modified with the vector pS/MARt and they were growing and replicating with the episomes, the plasmids presence as well as the expression of a reporter gene was not influencing their behaviour.

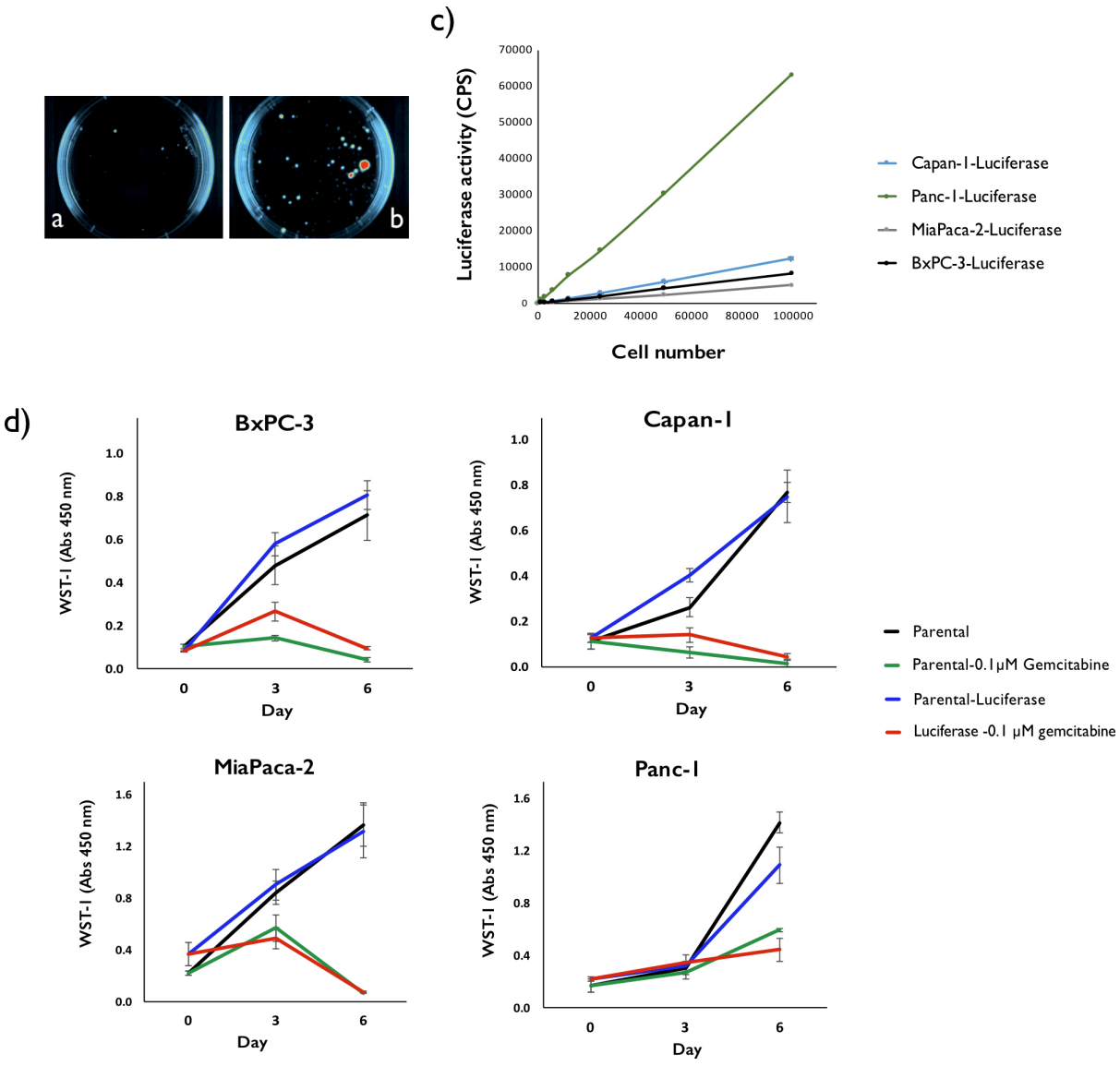


Figure 19. **Luciferase labelled pancreatic cancer cells**.
(a) Capan-1 colonies formed after 12 weeks of selection and several rounds of G418 treatments and (b) colonies formed after 5 weeks of selection with pS/MARt and only a single Puromycin treatment. (c) Luciferase expression was evaluated in comparison to the number of cells. Value represents the mean of six replicates for each cell number. (d) Proliferation was determined by measuring number of viable cells upon treatment with the control compound gemcitabine. Gemcitabine inhibits the proliferation of both, parental and luciferase labelled cells with the same efficacy. The cell validation was provided by Pharmatest Service (Turku, Finland)

3.10 Utilising DNA Vectors to generate Pancreatic Cancer Xenograft Models

The pancreatic cancer cell line BxPC-3 was used to validate in xenograft models pS/MARt luciferase labelled cells. *In vitro,* BxPC3-luc cells showed a behaviour similar to the parental unmodified control line. In order to investigate their applicability for *in vivo* drug tests, their tumorigenic potential in xenograft models was evaluated. BxPC-3-luc and unmodified cells were inoculated orthotopypcally into the pancreas of athymic nude mice and the tumour growth was monitored over time with non-invasive imaging (Figure 20, a).

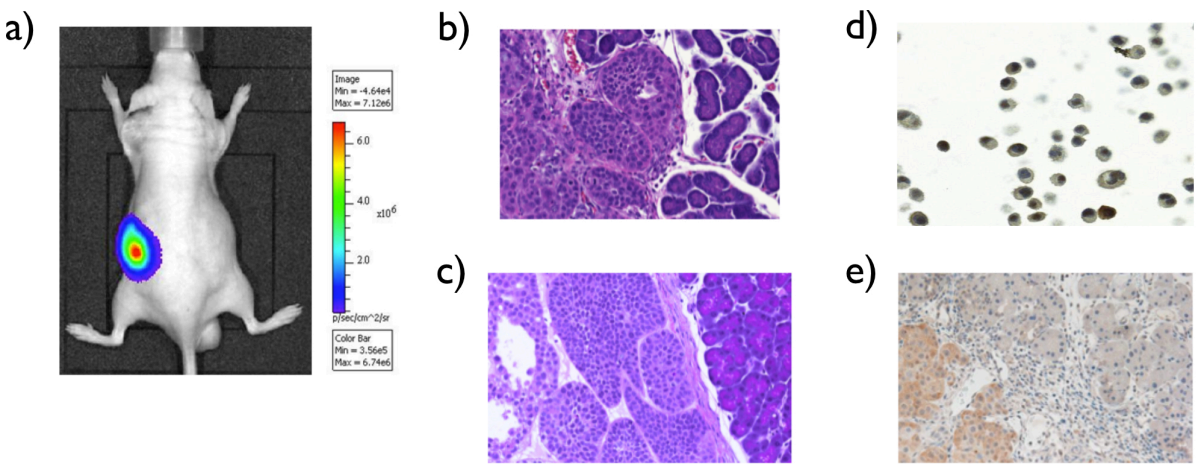


Figure 20. **Xenograft models with pS/MARt-Luciferase labelled BxPC3 and histological analysis.**
(a) Bioluminescent detection of orthotopically injected BxPC-3 in the pancreas of nude mice. Images were taken with an IVIS Lumina 2, 10 min after luciferin injection (3mg/mouse, ip). (b) Histopathological assessment of parental BxPC-3 and (c) BxPC-3 Luciferase. (d) IHC staining for Luciferase before injection and (e) after tumor development. The tumors generate with the luciferase labelled BxPC-3 have the same morphology to those generated from the unmodified control cells. The BxPC-3 luc cells show sustained expression of the reporter gene prior and after orthotopic injection.

The luciferase labelled cells maintained their transgene stable expression also *in vivo*. Tumours derived from the pancreases of mice injected with parental control BxPC-3 and luciferase labelled cells were taken and used for histopathological assessment. Genetically modified BxPC-3 formed tumours that had the same phenotype to those generated from the parental unmodified cells

(Figure 20, b-c). Immuno-histochemical staining for luciferase confirmed that before (Figure 20, d) and after orthotropic injection and tumour establishment (Figure 20, e) all the genetically engineered BxPC-3 were showed sustained expression of the reporter gene.

These results confirmed that the pS/MARt vector platform was able to produce highly reliable genetically tagged cancer cells and when applied in orthotropic xenograft studies they formed a reliable and essential non-invasive imaging platform.

This BxPC-3 luciferase labelled cells were used for testing the pancreatic cancer experimental drug compound VAL401 (ValiRx, UK). Four groups were included in the study: 1) control untreated, 2) gemcitabine, 3) VAL401 (1mg/kg, p.o. daily), 4) VAL401 (2mg/kg, p.o. daily). Orthotopically injected cells were successfully followed for 5 weeks by BLI (Figure 21, a).

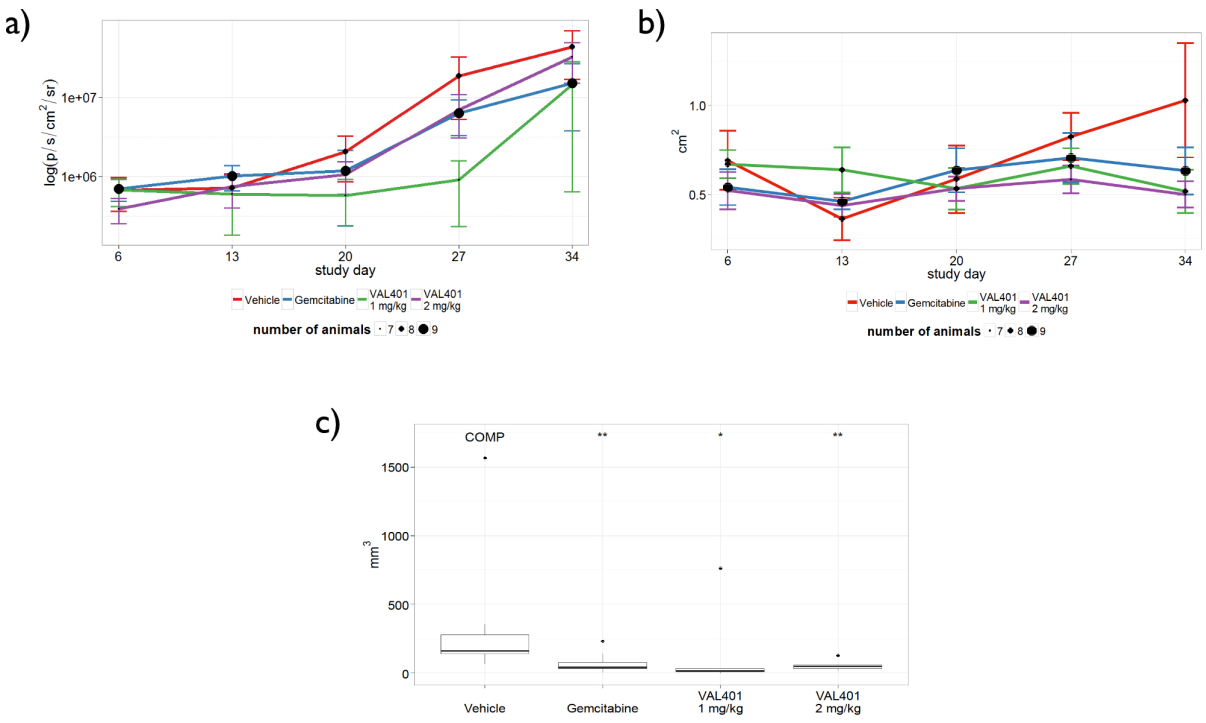


Figure 21. **Testing the efficacy of a new pancreatic cancer treatment.**
(a) Bioluminescent detection of Luciferase upon vehicle treatment, reference compound (gemcitabine) and study compound VAL401 (VAL401 1mg/kg and 2 mg/kg). Images were taken with and IVIS Lumina 2, 10 min after luciferin injection (3mg/mouse, ip). (b) Measure of tumors size based on the area of the detected bioluminescence. (c) Measure of tumors size upon mice sacrifice. The tumor growth was successfully followed *in vivo* for 34 days and the luciferase expressing cells formed a useful tool for the test of a new anti-pancreatic cancer drug.

Indirect measures of the tumours size showed that both doses of the experimental compound reduced the cancers dimension (Figure 21, b). Endpoint analysis on primary pancreatic neoplasms (Figure 21, c) confirmed previous data inferred by BLI. The reference compound gemcitabine as well as the experimental one significantly reduced the tumour size when compared to the vehicle treatment.

I*n vitro* and *in vivo* validations confirmed the reliability of the genetic modified cells obtained with the pS/MARt system. These cancer cell lines showed a proliferation rate that was comparable to the unmodified controls. When applied in orthotropic xenograft studies those cells formed a reliable and essential non-invasive imaging platform that improved substantially the efficacy testing of anticancer drug candidates reducing the number of animals necessary for the experiments.

3.11 Immortalisation and reprogramming into iPSC of mouse primary lung fibroblasts

pS/MARt.1 harbouring with the T-antigen from the SV-40 virus (Figure 22, a) was tested for its efficacy in transforming mouse primary lungs fibroblasts (Figure 22, b). High expression of the transforming agent was enough to induce the fibroblasts immortalisation and proliferation. The resistance marker was therefore swapped for the reporter gene GFP and it this was used to monitor the cell growth and to select the positive one via FACS sorting (Figure 22, c-d). T-antigen expression was evaluated in Western Blot (Figure 22, e) and these modified fibroblasts were used for the PhD project of Alicia Roig-Merino

(DNA Vector Laboratory Research, DKFZ). Reprogramming these cells into pluripotent stem cells (iPSC) demonstrated that pS/MARt was capable of sustaining the expression of a transgene during a process that involved profound epigenetic and phenotypic changes. The reprogramming factors were delivered via lentivirus and the iPSC were analysed for GFP (Figure 22, g) and alkaline

phosphates expression (Figure 22, h). Common markers for pluripotency such as Nanog, Oct4 and SSEA-1 (Figure 22, i) were used to further validated the pluripotency and the success of the reprogramming process. iPSC obtained from immortal fibroblasts would generate unreliable differentiated cells that cannot be used in further studies, however the experiment fulfilled the expectations. For the first time a non- integrative DNA plasmid was capable of sustaining a transgene expression during the reprogramming process. The experiment opened the possibility of exploiting the new technology in the field of stem cell modification and stem cell therapy.

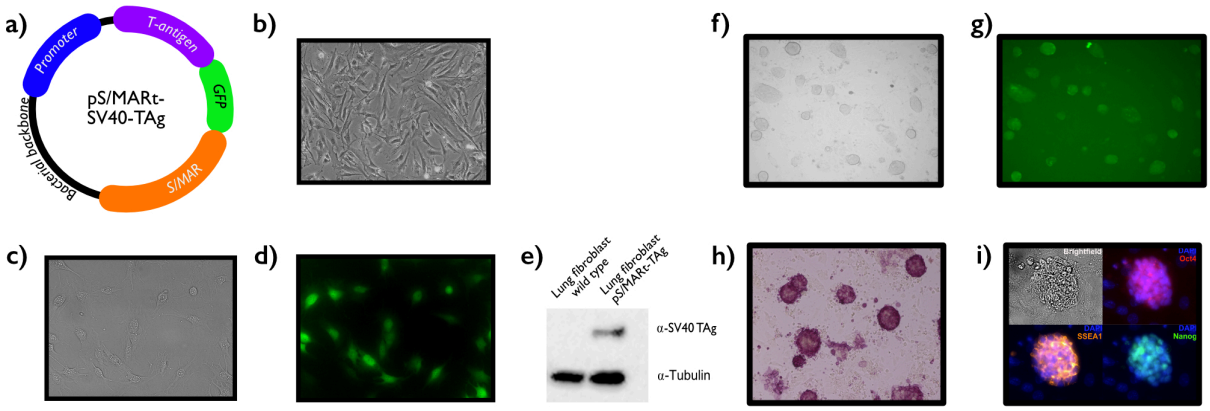


Figure 22. **Immortalisation and reprogramming of primary mouse lung fibroblasts.** (a) pS/MARt vectors containing with the transforming factor T-antigen from the SV-40 virus was used to transfect primary murine lungs fibroblast (b). The presence of the vector in the immortalised fibroblast was confirmed by fluorescent microscopy (c and d) and Western Blot (e). The immortalised fibroblasts were used to generate iPSC (f) which were tested for GFP (g). The alkaline phosphatase test (h) and the immunofluorescent staining for pluripotency markers Oct4, Nanog and SSEA-1(i) was used to confirm the successful reprogramming process.

3.12 Pronuclear injection of pS/MARt vectors

The efficacy of pS/MARt in establishing cells in culture has clearly been demonstrated. Typically, at least one round of selection is necessary during this process in order to eliminate the un-transfected cells as well as those that are not able to retain the vector. The establishment process is a stochastic event and in approximately 3% of rapidly dividing cells the S/MAR vector is maintained in its episomal *status* in absence of selection (data not shown).

In collaboration with Dr. Franciscus Van der Hoeven (DKFZ, Transgenic Service) pS/MARt expressing the reporter gene GFP was injected into a 1-cell stage zygotes of BL6 mice (Figure 23).

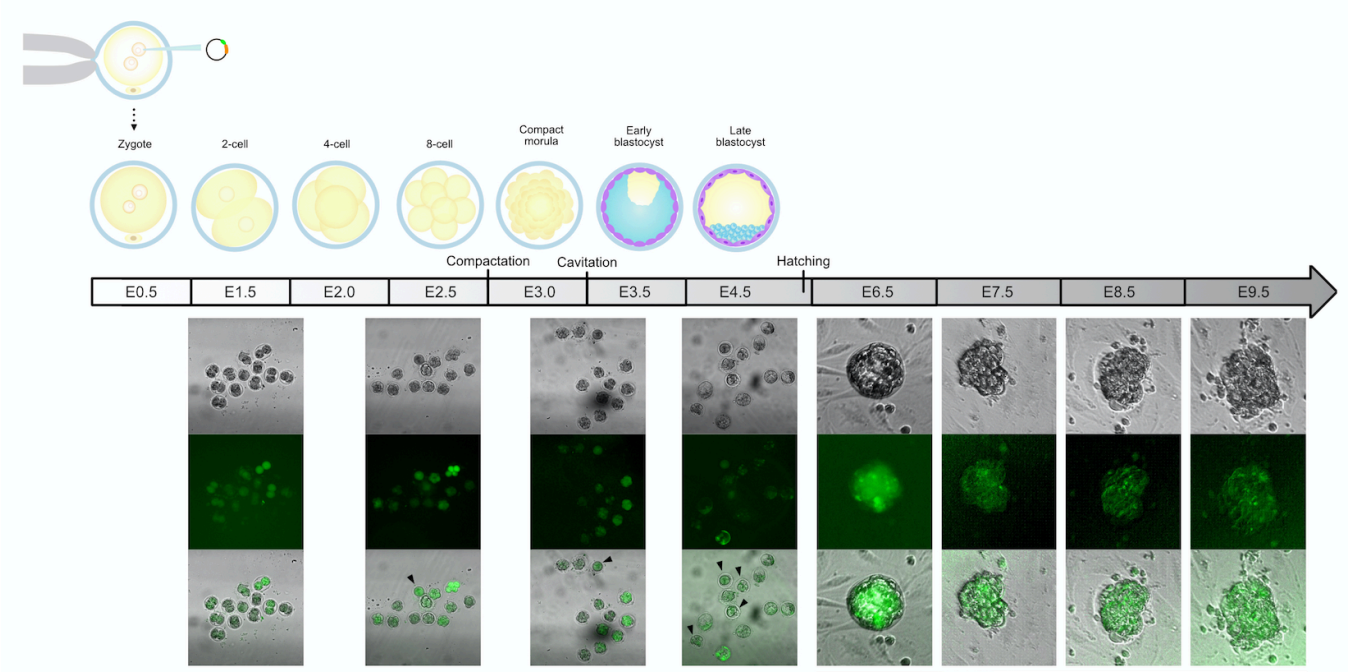


Figure 23. **Pronuclear injection and *in vitro* follow-up of the embryonic development.** 1-2 pl of 1-3 ng/ul of filtered and dialyzed pS/MARt vector was injected in the pronucleus of 1-cell stage embryos. The embryonic development was followed in culture and the GFP expression was monitored via fluorescent microscopy. On day 4.5 the embryos were transferred on gelatin-feeder coated wells and the embryos that hatched in these artificial condition was monitor for transgene expression to death (Embryos follow-up and culture was performed by Alicia Roig Merino (DKFZ, DNA Vector Lab. The figure was provided by Alicia). The pS/MARt vector was able to sustain the expression of the reporter gene GFP *in vitro* until the embryo hatched.

The aim was to determine whether the vector was able to sustain the transgene expression during the embryonic development and remain active and unsilenced. Upon injection in the pronucleus the embryos were transferred into pseudo pregnant mothers and Alicia Roig Merino (DKFZ, DNA Vector Lab) is investigating the GFP expression in the new born transgenic mice via fluorescent microscopy and GFP-PCR on tail biopsies.

In order to evaluate potential toxic effects induced by the plasmid, 12 embryos were kept in cultured and monitored over time. One arrested the development at the 1-cells stage, three stopped dividing and the other 8 underwent compaction and progress to the blastocyst stage. In one case the embryo hatched and attached to gelatinized-feeder coated plated and GFP expression was still detectable.

For the first time this new class of DNA vectors was able to sustain the expression of a transgene during the embryonic development *in vitro* remaining unsilenced and active. The analysis of the animals will definitely prove if the pS/MARt vector system can be also used for the generation of transgenic mice.

3.13 Utilising pS/MARt vector for Liver Gene Therapy

Alkaptonuria is a rare inherited genetic disorder in which the body cannot completely process the amino acids phenylalanine. It is caused by a mutation in the Homogentisate 1-2-dioxygenase (HGD). People who have both of copies of their HGD gene mutated build up an intermediate substance called homogentisic acid (HGA). This accumulation leads to osteoarthritis and the formation of stones. The phenylalanine degradation takes place in the liver and to better understand this disease HGD knockout mice were generated. The liver represents an easy target for the delivery of DNA. The hydrodynamic injection is indeed an efficient delivery process to transfer plasmids to this organ.

Argyros et al. developed a modified version of pEPI for the targeted expression in the liver of mice that carried a tissue specific promoter instead of the viral CMV. Although it guaranteed the luciferase expression for the lifetime of the animal an appreciable decline was observed, while minicircles remained more active in the same period of time. pS/MARt showed *in vitro* higher efficiency in establishing cells than pEPI. It was therefore tested for its capability in expressing the wild type cDNA of the HGD gene *in vivo*. A synthetic minimally sized liver specific promoter (P3) was used to drive the expression of the gene and two different versions of the vector were created and delivered to the liver of HGD knockout mice (experiment performed in collaboration with the group of Dr. Jonathan Jarvis, Institute of Aging and Chronic Disease, University of Liverpool). In one vector the P3 promoter drew the expression of the wild type HGD cDNA (pS/MARt-P3-HGD), in the other, it was responsible for the expression of the reporter gene Luciferase (pS/MARt-P3-Luc). Upon DNA delivery the HGA levels in the plasma of mice treated with pS/MARt-P3-Luc were comparable to the saline treated group. In comparison, in the animals where the HGD gene was restored, the HGA level in the plasma was restored to normal levels (Figure 24).

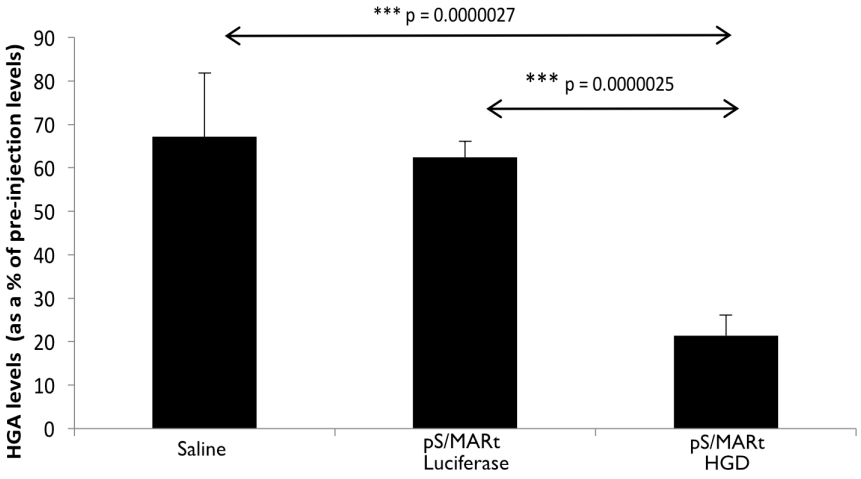


Figure 24. **HGA plasma level upon hydrodynamic delivery of pS/MARt vectors and saline control.**
HGA levels in the plasma of the treated mice were measured upon delivery of the vectors expressing the wild type cDNA of the HGD gene and the Luciferase reporter gene. The HGA plasma levels were reduced below toxic levels from a single administration of pS/MARt expressing the HGD cDNA.

These levels were maintained for several weeks and they were used as proof of the vector's functionality. The expression of the enzyme was inferred from the analysis of the plasma but, in order to determine whether also in this new system the transgene expression declines more detailed analysis such as regular qPCRs and Western Blots are ongoing.

3.14 Generation of Isogenic Pancreatic Cancer Cell Lines and the genetic rescue of the tumor-suppressor SMAD4

A range of pancreatic cancer cell lines labelled with the reporter gene luciferase was tested for cell growth *in vitro* and, in the case of the BxPC-3 also used for evaluating the efficacy of a new pancreatic cancer drug *in vivo*.

The pancreatic cancer cell line Capan-1 was selected to also test the vector system for its capacity to rescue the functionality of a fundamental tumour suppressor gene whose loss is fundamental in many aggressive pancreatic tumours. The interest in these particular cells and this gene arose from a collaboration with Prof. Offringa's group (DKFZ, Molecular oncology and gastrointestinal tumours). Previously, the DKFZ core facility attempted to produce genetically modified Capan-1 with lentiviral vectors but, it proved difficult to obtain stably expressing lines with reports of variable and/or diminished expression. We were challenged to apply pS/MARt in this difficult to modify cell line. The aim was to initially generate Capan-1 cells that expressed the reporter gene luciferase and demonstrate the capability of the vector in modifying these cells. If this proved successful, the idea was to rescue the functionality of the tumour suppressor gene SMAD4 which was knocked out in this line. The mutation of SMAD4 represents one of the key events in the development of aggressive pancreatic cancer. Capan-1 cells represent a reliable model for the study of this disease because they present mutations in four primary pathways: KRAS (G12V), TP53

(A159V), CDKN2A (del), SMAD4 (S343*)(Deer, Gonzalez-Hernandez et al. 2010). The study aimed to investigate the effect of the resurrection of SMAD4 in relation to the TGFβ pathway. pS/MARt-luc and pS/MARt-SMAD4-luc vectors were used to generate stable cell lines following two different approaches. In one, the cells expressing SMAD4 were expanded via single cell clonal expansion. In the other, a mixed population expressing the tumour suppressor was obtained. The expression of SMAD4 was evaluated via Western Blot (Figure 25, a) and its impact on cancer development was tested *in vivo* by injecting $0.5x10^6$ cells orthotopically into NSG mice to generate xenograft models.

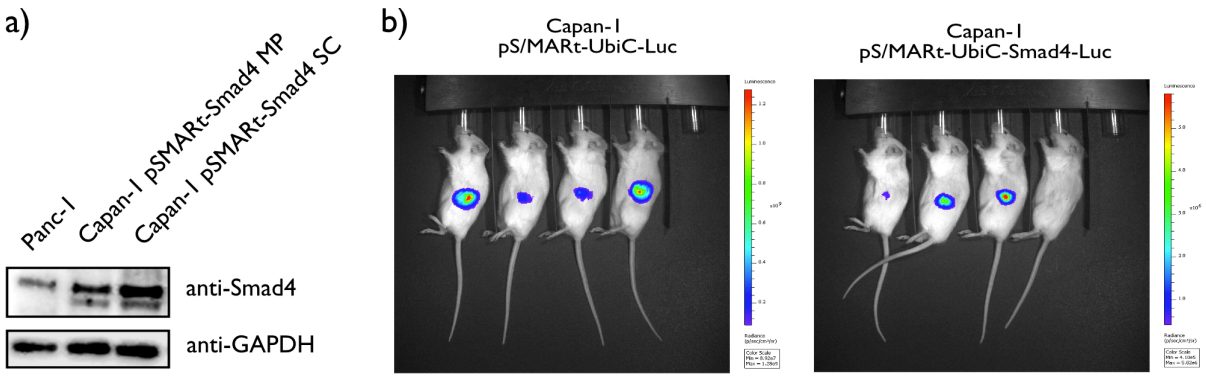


Figure 25. ***In vitro* and *In vivo* analysis of the modified Capan-1 cells**.
(a) The expression of the tumour suppressor SMAD4 in the Capan-1 cell line was evaluated in comparison to the pancreatic cancer cell line Panc-1 which has normal expression of this gene. (MP, mix population. SC, single cell). (b) The cells were injected orthotopically in NSG mice and the tumour growth monitored through BLI. The images were taken 10 min after luciferin injection (3mg/mouse, ip) with an IVIS Lumina 2.

Both vectors used for the modification of the cells carried the reporter gene Luciferase, therefore, four weeks after orthotopic implantation the tumour development was analysed by BLI. The signals generated from animal injected with the Capan-1-Luciferase cells were significantly higher than those produced from Smad4 rescued cells (Figure 25, b) which meant that also the correspondent tumours were bigger.

In accordance with previous observations (Yatusome et al Clinical & Experimental metastasis 22:461-473, 2005) the rescue of the functionality of the tumour suppressor SMAD4 leads to a reduction in tumour growth. All the mice injected with parental or control cells developed a primary tumour in the pancreas and those in which the xenografted cells had restored SMAD4 showed small non-invasive accumulations of transplanted cells. The single cell clonal expansion (SC) cell line as well as the mixed population (MP) generated small tumours but metastasis was not detected. For the first time, taking advantage of the high sensitivity of the BLI technique, metastasis were observed in the liver and in the lungs of mice injected with Capan-1 luciferase labelled cells (Figure 26, a). The histo-pathological analysis revealed that the luciferase modified cells developed tumours (Figure 26, c) phenotypically similar to those formed from the parental unmodified cell line (Figure 26, b), characterized by a highly differentiated ductal structure. Both cell lines expressing SMAD4 formed primary tumours with identical histology (Figure 26, d-e). They appeared less differentiated and showed a higher recruitment of stroma.

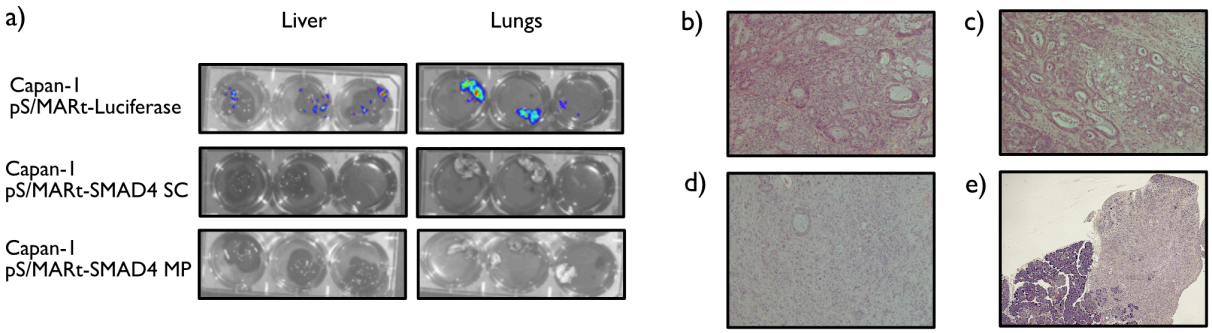


Figure 26. **Detection of Luciferase expression in liver and lungs metastasis and histological analysis of primary tumours.**
(a) Detection of metastasis in the liver and in the lungs of the mice injected with the Luciferase labelled Capan-1 cells. Primary tumour formed from the Capan-1 luciferase (c) present the same morphology and a well-defined ductal differentiation, like those generated from the parental unmodified cells (b). The restoration of Smad4 induce profound change in the phenotype of the tumour (d - e). Both cell lines, SC (d) and MP (e) form an identical primary tumour.

Since the Capan-1-luc and parental cells generated identical primary tumours and metastasis, the differences observed in the tumour masses generate by Smad4 rescued Capan-1 appeared to be directly dependent to the re-introduction of the tumour suppressor gene SMAD4.

Primary tumours from Capan-1-luc and Capan-1-SMAD4 cell lines were compared for: phenotype (Figure 27, A), proliferation via staining of Ki67 (Figure 27, B) and expression of SMAD4 (Figure 27, C and D).

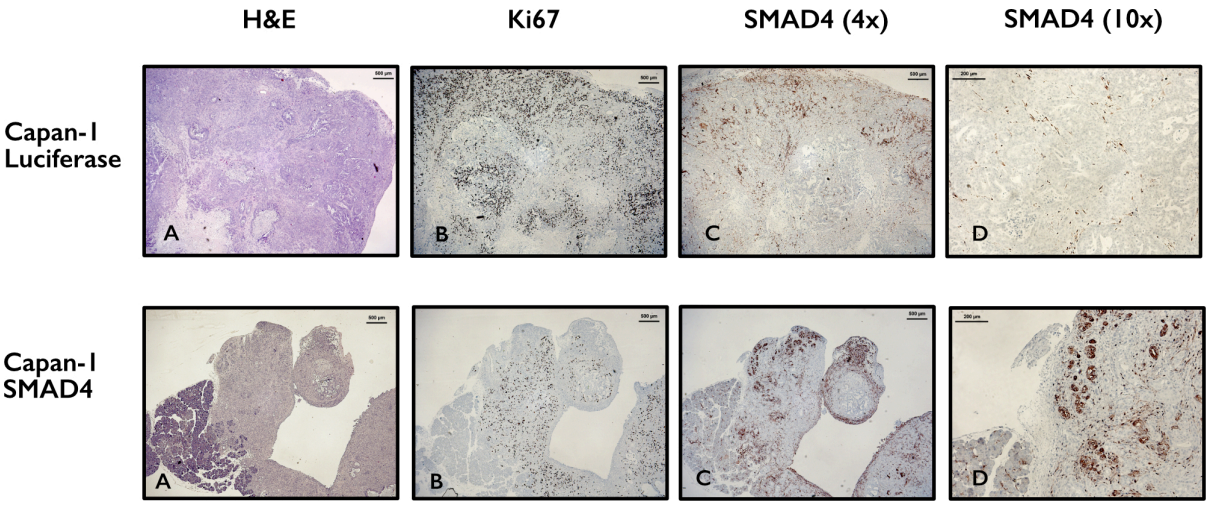


Figure 27. **Immunohistological assessment of Capan-1 luciferase and Capan-1 Smad4.** Capan1-Luciferase (A to D upper panel) and Capan-1-SMAD4 (A to D lower panel) were assessed for morphology (A) with the H&E staining, proliferation (B) via staining for the proliferative marker Ki67 and SMAD4 expression. SMAD4 was evaluated at two different magnifications (C, 4X and D, 10X). The tumours formed from Capan-1 modified with the reporter gene luciferase shows a defined ductal differentiation typical of pancreatic cancers and a high proliferative rate. Instead, the primary tumours formed from cells where Smad4 functionality is restored appear less differentiated and with a significantly lower proliferation potential. All the masses stained positive for Smad4. Capan-1 Smad4 rescued cells strongly stained for this protein, whereas in tumours formed with Capan-1-luciferase cells only infiltrating fibroblasts express the gene.

A significantly reduced number of Ki67 positive cells (Figure 27, B) in tumours formed with Capan-1-Smad4 cells, demonstrated that those masses had a lower proliferation rate. The positive staining for Smad4 (Figure 27, C and D) confirmed the vector activity upon orthotropic injection and tumours development. It demonstrated that the observed effects were induced by the re-acquired functionality of Smad4. Tumours formed from Capan-1 luciferase cells showed also positive staining for Smad4 (upper panels C and D) due likely to cross reactivity of the antibody. A mouse

monoclonal anti-Smad4 antibody was used for the analysis and it bound probably also SMAD4 proteins endogenously expressed from infiltrating fibroblasts. To investigate the effects produced by the vector as well as by SMAD4, genome wide RNA profile analysis on primary tumours was performed. The RNA was extracted from 3 mice per group and analysed on a IlluminaHuman12 chip (Analysis performed at the DKFZ Genomic and Proteomic core facility). The RNA quality control, the microarray hybridisation and the data normalisation were performed by the core facility. The differences in gene expression was analysed with the Partek Genomic Suite software (Thermo Fisher) upon logarithmic transformation of the microarray data. The RNA expression levels were introduced into the non-rooted hierarchical clustering tool of the software and they were also analysed for Principal Component Analysis (PCA) differences (Figure 28).

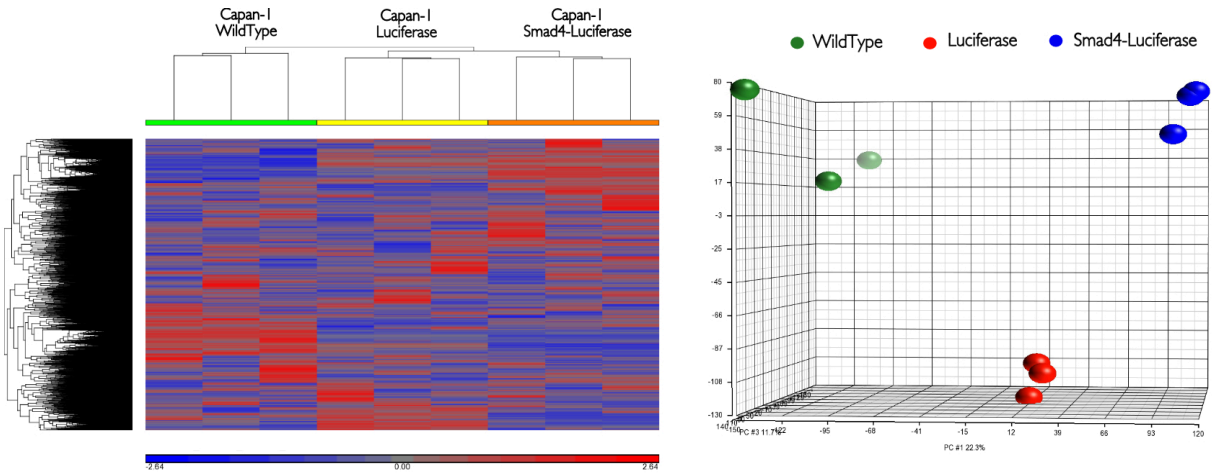


Figure 28. **Non rooted hierarchical clustering and PCA analysis of RNA profiles from Capan-1 derived pancreatic tumors.**
Total RNA from pancreatic tumors was extracted (RNAeasy Kit, Qiagen) and the expression profiles were determined by microarray hybridization with a Human Illumina-12 Chip. Genome wide RNA expression values were analyzed with the Partek Genomic Suite software in a non-rooted hierarchical clustering test (left panel) and a PCA assay (right panel). Pancreatic tumors from three mice per group were considered for the study.

The un rooted hierarchical clustering showed 3 different clusters that correspond to the tumours obtained from the different xenografted Capan-1tumours. The phylogenic tree generated by the software showed two different branches. One for the parental unmodified cells and another one divided into two nodes. The first, clustered closer to the parental cells represent the luciferase-

labelled Capan-1 derived sample whilst in the other, more distant, cluster the SMAD4-luciferase tumour samples were located. The different branches generated in the hierarchical clustering demonstrated that cells which underwent the selection procedure and establishment were more similar compared with the unmodified parental samples. The analysis of the PCA confirmed the differences between cell lines. The three replicates per group clustered together also in this assay which showed that the differences were conserved within the biological replicates.

Differential gene expression analysis was performed and a cut-off of 2 fold up/down-regulation with a False Discovery Rate (FDR) of 0.1 was applied (Figure 29).

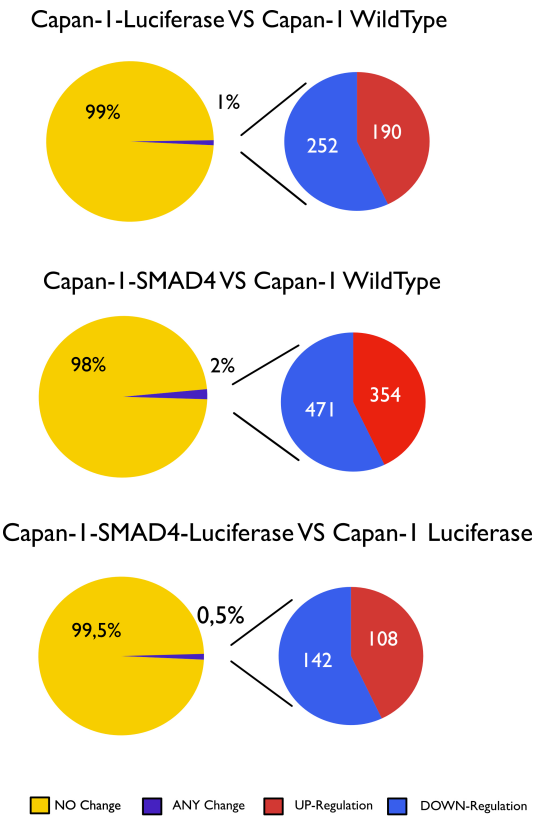


Figure 29. **Genome wide RNA profiles analysis of primary xenografted tumors generated from different Capan-1 cell lines.**
RNA from primary tumors was extracted and used for microarray analysis on a IlluminaHuman.12 chip. The array was performed in the DKFZ Genomic and Proteomic Core Facility and also the normalization across the samples was performed there. For the analysis, primary tumor from 3 mice were analyzed per cell line. The gene expression was evaluated with the Partek Genomic Suite Software (Thermo Fisher) 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 dow-regulated in the modified cell lines when compared to the parental one (first two panels) and between them (bottom panel).

1% of the genes were perturbed when the expression profiles from Capan-1 luciferase tumours were compared to those formed from parental cells. 252 genes appeared downregulated and 190 upregulated. The overall gene expression pattern appeared more perturbed when SMAD4 was re-introduced. 2% of the genes were changed with 354 up- and 471 down-regulated. 0.5% of the genes changed in tumours generating with the cells expressing the double cassette SMAD4-luciferase when compared to the luciferase expressing cells. 108 were up and 142 down regulated.
The comparison between the expression profiles revealed that only a small proportion of the genes was affected by the establishment procedure and the expression of the reporter gene. These genes however, didn't influence the cells behaviour during the tumour development and/or the metastatic process. The differences became more prominent when SMAD4 was re-introduced. The effect of the tumour suppressor was already appreciated by histological analysis and further confirmed at the molecular level. The analysis of differently expressed genes was performed to identify SMAD4 dependent perturbations and the effects induced by the vector and the expression of the inert gene luciferase (Figure 30).

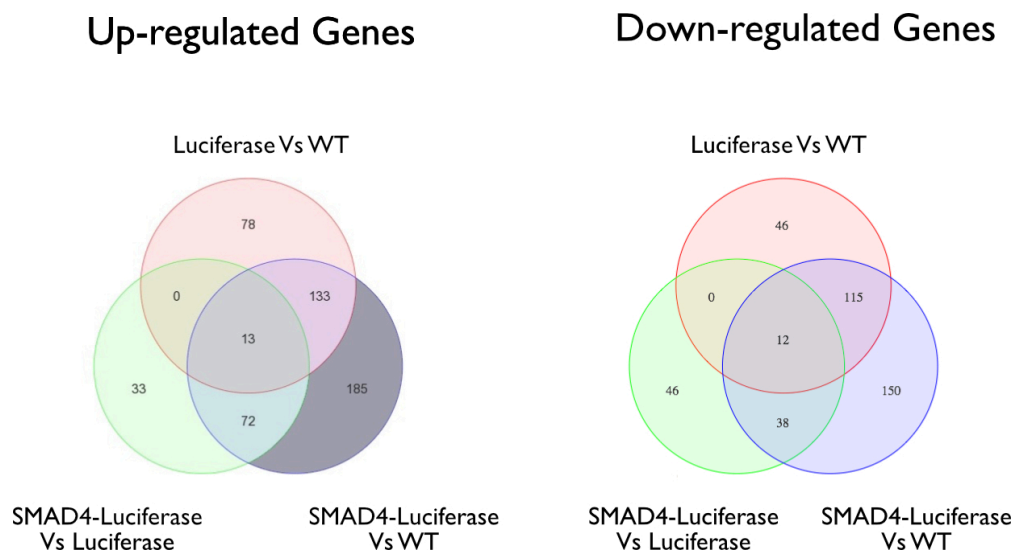


Figure 30. **Differential gene expression clustering.**
Genes that had an upregulation of at least 2 fold with FDR=0.1 are displayed with a Venn diagram in the left panel and genes with at least 2 fold downregulation are shown in the right one. In the left panel are reported the genes which up regulation is Smad4 dependent by intersecting those that presented at least 2 fold up-regulation in the comparison between tumors from the three cell lines. The same approach (right panel) was used to identify the genes which down modulation was Smad4 dependent.

The results are summarised as Venn Diagrams and the transcripts that were up or down modulated by the re-introduction of Smad4 (Table 17 and Table 18) as well as those that changed as a consequence of the selection procedure were isolated (Table 19).

Genes up-regulated in Smad4 positive cells are mostly associated with cell motility whereas the luciferase expressing cells presented up-regulation in transcripts involved in the interferon α and the inflammatory pathway.

The profiles were further investigated via Gene Set Enrichment Analysis (GSEA). This analysis compared the gene expression values generated in the microarray to the gene expression in the most common cancer hallmarks. The strongest enriched hallmark in cells expressing SMAD4 was the Epithelial to Mesenchymal Transition (EMT) as displayed in Figure 31.

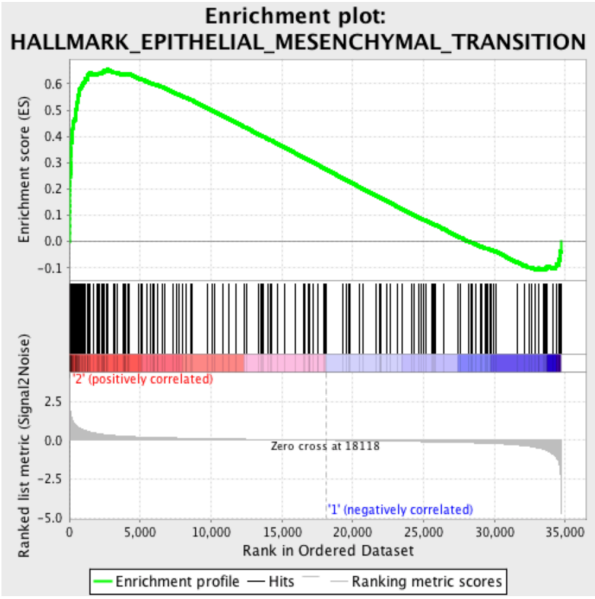


Figure 31. **GSEA of Smad4 modified cells versus Luciferase expressing Capan-1.** The GSEA analysis shows a strong enrichment for the hallmark underling the epithelial to mesenchymal transition in Capan-1 cells where the functionality of the tumor suppressor gene Smad4 was restored.

In accordance with previous publications, the re-introduction of SMAD4 in pancreatic cancer cells that harbour mutations in this gene, induces EMT transition (David, Huang et al. 2016). *David et al.* demonstrated with genetically modified mice that, when pancreatic cancer is induced and the cells harbour Smad4 mutations, the *ex vivo* rescue of this gene functionally induced the EMT transition that leads the apoptosis and cell death. The epithelial to mesenchymal transition is normally responsible for the cells acquired plasticity which triggers the metastatic process. In the presence of functional Smad4 however, upon TGF-β stimulation, the apoptotic pathway is triggered and the cells undergo the programmed cell death.

The enrichment in the EMT hallmark was a clear sign of SMAD4 functionality in the tumour models generated with the pS/MARt vector system. Although active, the genes involved in the EMT transition, upon stimulation with TGF-βI, didn't trigger the apoptotic pathway nor in culture nor *in vivo*. Likely, the Capan-1 culturing system in presence of Fetal Bovine Serum (FBS), which contains TGF-βI, induced a secondary aberration that allowed the modified cells to grow. They also have mutations in other key pathways and the resurrection of a single gene was unlikely to be able to inhibit the cells growth and inducing the cells death. Proteins involved in cell proliferation pathway were found up-regulated which probably compensated the apoptotic signals induced by the activity of SMAD4. Indeed, the cells did present activated hallmarks for the PI3K-Akt as well as high levels of DAPK1. PI3K-Akt is well known to induce cell proliferation and to be an antagonist of the TGF-β pathway. DAPK1 also acts as an inhibitor of the apoptosis. It was demonstrated that it has pro-proliferative properties in presence of mutated p53.

The elucidation of the effect of SMAD4 in pancreatic cancer tumour biology is beyond the expertise of the lab and the aim of this project. The main question, however, was whether the vector was capable of sustaining the expression of a crucial tumour suppressor gene in a representative pancreatic cancer cells or not, and whether the cells modified with the S/MAR technology were a reliable tool for tumour modelling. The result clearly demonstrated that the

genetically engineered cells with the S/MAR technology formed a reliable and useful platform for the study of pancreatic cancer and that these cells could be taken for further studies. The interest of the lab was more on the "vectorology" and to understand better the interaction between the vector and the cell's genome. The array data generated from the Capan-1 cells were then used to investigate which genes were perturbed but not SMAD4 dependent (Table 19). The GSEA analysis showed an enrichment for the hallmarks associated with inflammatory responses (Figure 32).

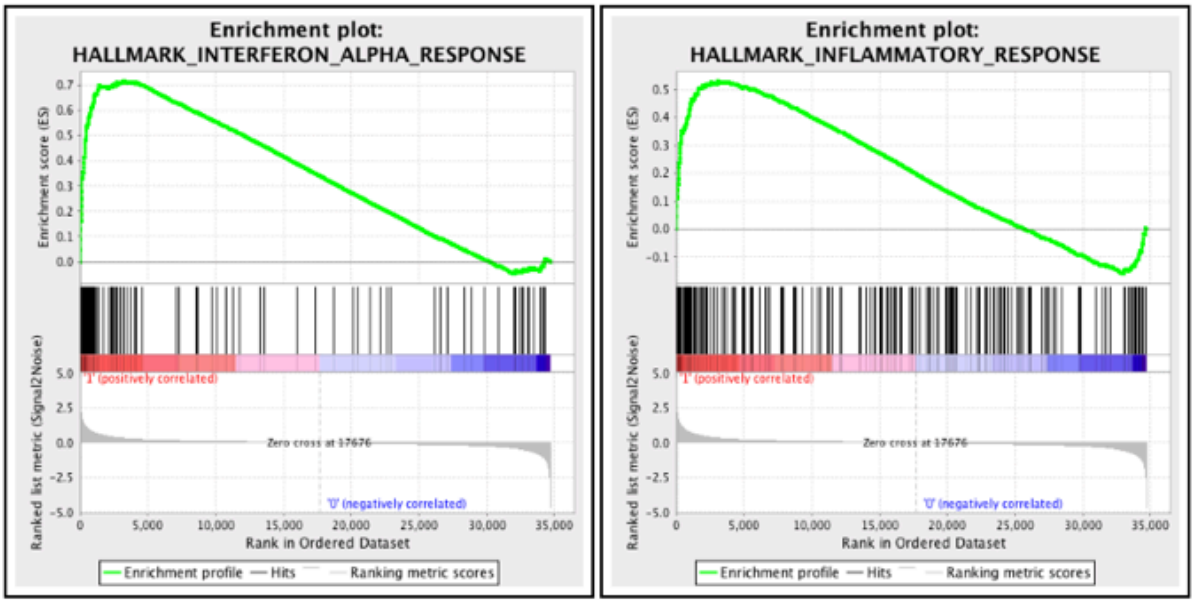


Figure 32. **Enriched hallmarks in Capan-1 luciferase cells compared to Capan-1 wild type.** GSEA analysis of Capan-1 luciferase expressing cells show strong enrichment for the hallmarks underling the interferon α and the inflammatory responses when compared to wild type Capan-1.

The interferon α pathway is associated with the cellular immune response to viral infection. It is part of the innate immunity and it is triggered by cytoplasmatic proteins that can bind infectious agents' genome during the translocation into the nucleus. The fact that these signature appeared enriched in the analysis suggested that also pS/MARt was recognised as a foreign entity and its presence induced an inflammatory state. Foreign DNA is certainly recognised by the cells during its translocation into the nucleus during the transfection process, but, upon establishment it should be

treated as part of the cell genome. The cell inflammatory state even after orthotopic injection suggested two possibilities. In the first, the DNA is recognised during the transfection and the cells maintain a memory of its presence. In second the episomal DNA is sensed by the cellular DNA sensors at every replication when the nuclear membrane is disaggregated for the mitosis. During this process the nucleus is accessible for the cytoplasmic proteins that can bind the DNA and induce an interferon response which is also associated with chronic inflammation. How a cell senses the DNA, which are the proteins responsible for this process and where this process takes place is still unclear. There is evidence that the protein STING and TLR9 are involved in the binding of DNA in the cytoplasm and this binding leads to the triggering of the innate immunity. STING and TLR9 act both in the cell cytoplasm and they would be candidates also for pS/MARt recognition. TLR9 binds unmethylated CpG islands, like those that characterise the pUC Ori or the Kan resistance gene and STING has a high affinity for DNA. It is normally not active under normal conditions, since is unlikley that genomic DNA leaves the nucleus. However, if the recognition takes place in the nucleus, other proteins would be involved. The vector system developed would then, not only a useful tool for generating cancer labelled cells or tumor models, but it may also be a useful platform for studying the pathogen DNA-cell interaction. To replicate, DNA viruses need to access the nucleus and here they have to elude the cellular mechanism of recognition. Our system could be used to further investigate this mechanism and the events that characterise the establishment of a persistent viral infection.

3.15 Generating minimally sized Nano-S/MARt vectors

The efficacy of the new pS/MARt technology for generating genetically modified cell-lines was described previously. Some cells such as primary human cells are typically refractory to transfection with traditional bacterial plasmids that contain a large bacterial backbone which comprise an antibiotic resistance marker, the pUC replication origin and other bacterial sequences which are not required for gene expression in mammalian cells. To overcome this key aspect of toxicity which can be induced by bacterial sequences contained in these bacterial vector systems, we developed minimally sized next generation S/MAR DNA vectors based on Nanoplasmid™ technology in collaboration with Nature Technology Corporation (Luke, Carnes et al. 2014). The idea was to produce a novel, more modern, DNA vector platform suitable for genetically modifying cells with a particular interest for application to patient derived (PDX) pancreatic cancer cells. This novel DNA nano-vector is characterised by a minimally sized bacterial backbone and an antibiotic free RNA-Out selection system (Figure 33). In this class of vector the bacterial backbone is reduced to a small interfering RNA (iRNA) motif and they can be manipulated and expanded in a specially engineered strain of *E.Coli.* The selection of the transformed bacteria is made on LB-Agar, enriched with sucrose. The vector expresses a small RNA that can bind a complementary sequence on the mRNA produced from the bacterial genome, responsible for the production of the enzyme levansucrase. This protein converts the sucrose into a toxic compound that induces the bacterial death. When the plasmid is present in the prokaryotic cell, the iRNA binds its complementary sequence, the levansucrase production is inhibited and the bacterial cells survive and proliferate.

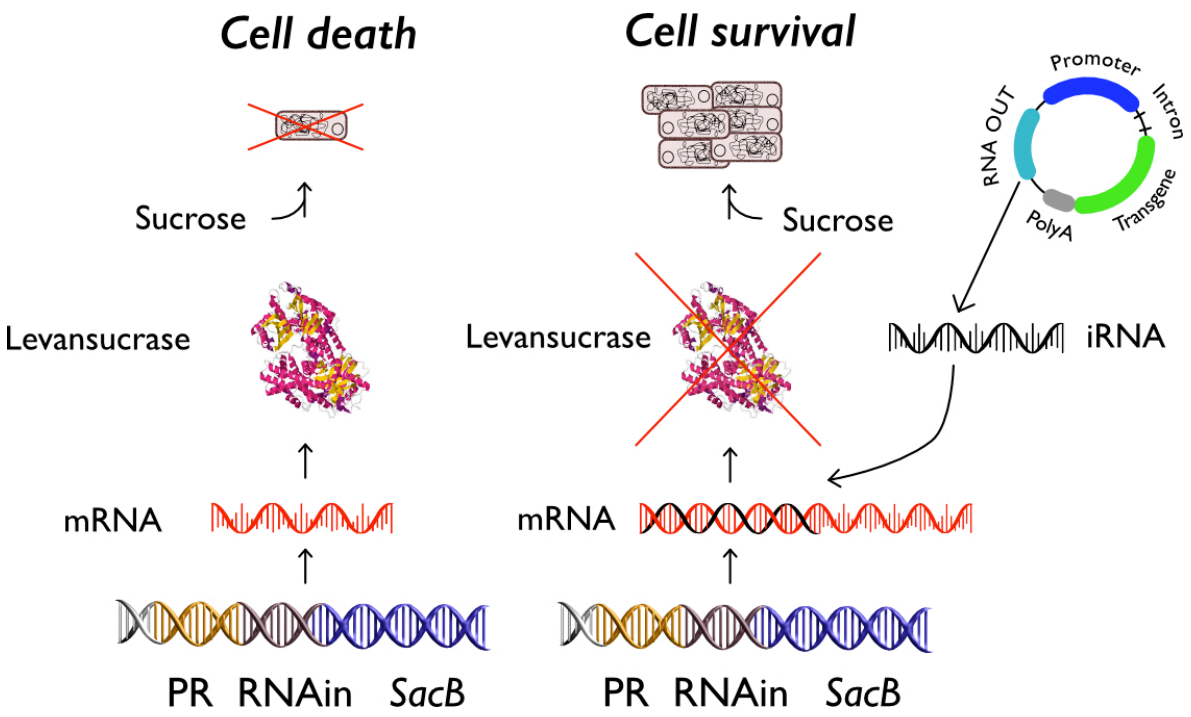


Figure 33. **Selection of recombinant plasmid with the RNA-OUT system (NTX ™).** For the selection of bacteria and propagation of the plasmids, the nano vectors produced from NTX present a small interfering RNA (iRNA) instead of an antibiotic selection marker. The plasmids are propagated in a engineered strain of *E.Coli* which express the enzyme Levansucrase (SacB). In the genome of the bacteria between the promoter (PR) and the gene there is a sequence (RNAin) that is complementary to the iRNA expressed from the plasmid. When this particular strain of bacterial cells are transformed with a NTX vector, it produces the iRNA, that can bind the complementary RNAin sequence on the RNA responsible for the production of the levansucrase. This RNA-RNA interaction inhibit the translation of the protein. In absence of the iRNA, the protein is produced and in presence of sucrose this enzyme induces bacterial death. If the translation is inhibited the protein is not formed and the bacteria can grow and expand in presence of sucrose.

The isolation of the nano-vector DNA is then made via normal DNA preparation. Compared to other minicircle producing systems, the manufacturing of NTX nano vector technology generates a high yield of DNA and its preparation is very simple and cheap. We cooperated with NTX to develop a Nano-S/MAR vector for persistent and stable expression in dividing cells. The primary limitations of other minicircle production procedures is that the purification of the minicircles from the producer plasmid vectors is time consuming and inefficient. In order to generate minicricles an intramolecular recombination in the so called “producer” plasmid is induced and the desired minicircle vector has to be isolated and purified using specifically designed and proprietary columns.

Minicircle production requires also an additional purification step to remove concatenamers which are generated during the intracellular recombination event. In the RNA-OUT system from NTX there is no need of intramolecular recombination and the exclusion of the bacterial backbone. With this system is possible to generate large quantities of pure supercoiled DNA vector without the need of additional steps of purification. The pS/MARt vectors were retrofitted with the nano bacterial backbone from NTX to generate the Nano-S/MARt vector system. (Figure 34)

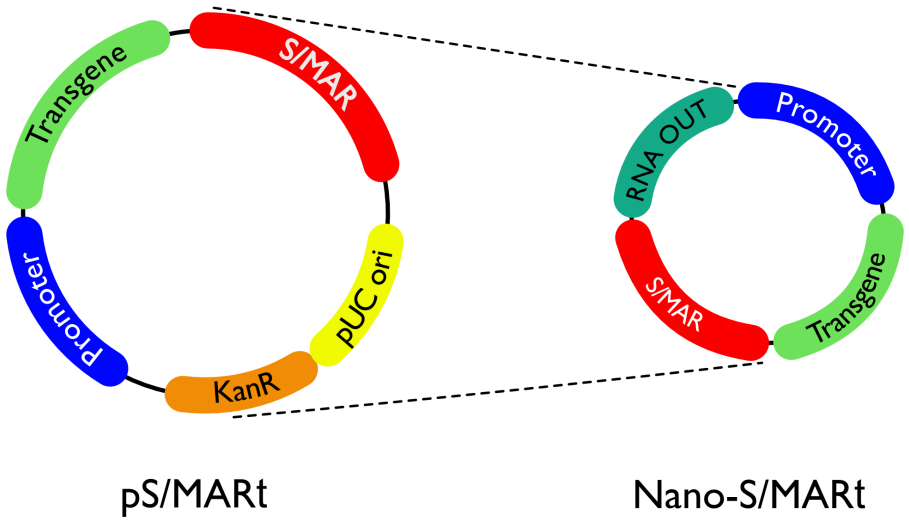


Figure 34. **Schematic representation of pS/MARt and Nano-S/MARt vector.** The bacterial backbone of pS/MARt containing the Kan resistance and the pUC Ori was swapped for the RNA OUT system previously described. This process reduces the amount of potentially toxic bacterial sequences by ~30% and the size of a typical vector by 2 kb.

The RNA-OUT system significantly reduced the size of the vector (Nano-S/MARt is 1 kb smaller than pS/MARt) which resulted in a higher number of transfecting DNA molecule per cell. The establishment of the vector is a stochastic process and if only the vectors that reach the correct site in the nucleus are able to establish, the delivery of a higher number of DNA molecule per cell, together with the reduced toxicity due to the absence of the bacterial backbone, should increase the cells establishment rate. To test the efficacy of this new class of plasmid, named NP (Nano-Plasmid), pS/MARt.1 vectors and pS/MARt were retrofitted into NTX™ technology generating:

Nano-S/MARt, Nano-S/MARt-Ele40 and Nano-S/MARt-UCOE. The vectors were compared in a colony forming assay (Figure 35). DNA was delivered to HEK293T cells and 24 h after transfection the positive cells were FACS sorted (FACS Aria). 100 cells were plated into a 6 cm dish and three replicates were generated per vector.

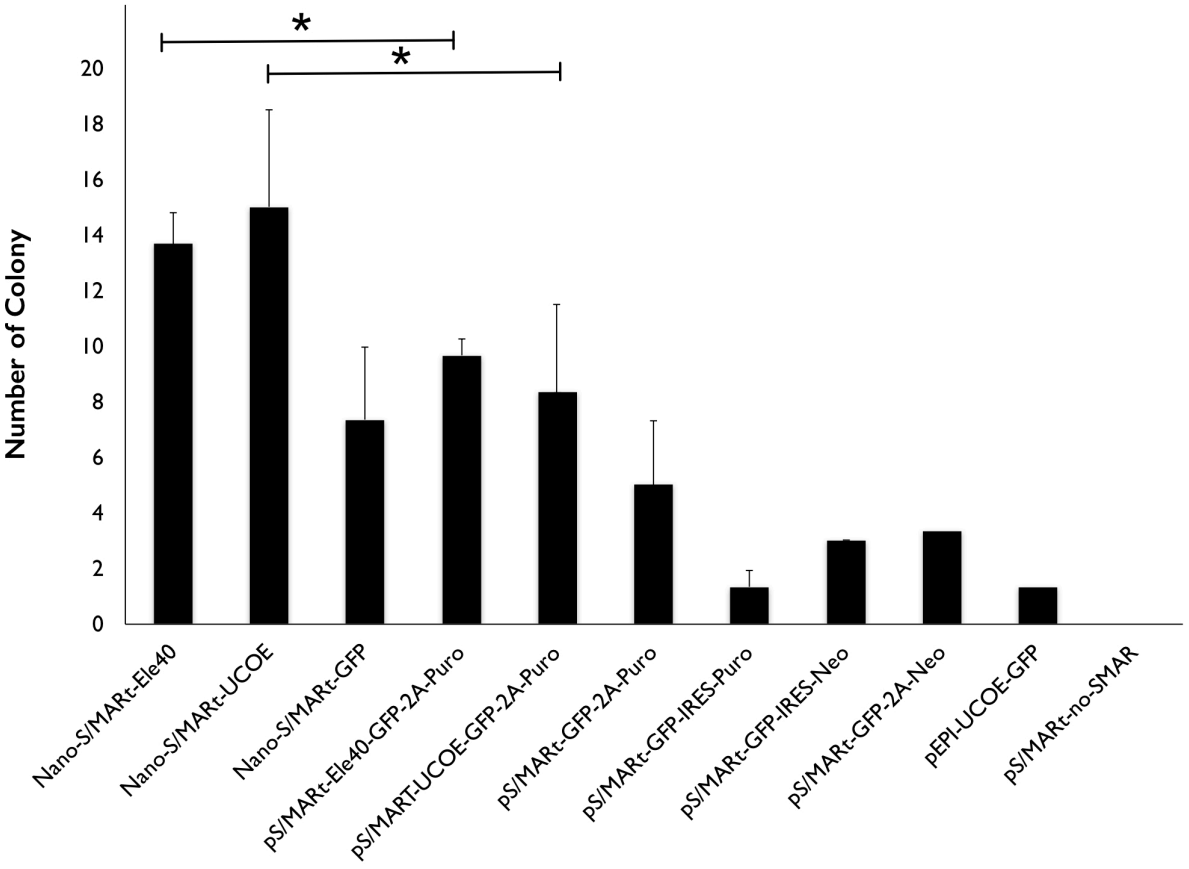


Figure 35. **Colony forming assay pS/MARt and Nano-S/MARt.**
The Nano-S/MARt version of the vectors previously tested was evaluated in a colony forming assay. The analysis was performed with a t-test and the data were considered significant with a $p < 0.05$ (* $p < 0.05$). The transfection was standardised on the number of DNA molecule. Where necessary the number of DNA molecule was adjusted with a non-expressing pUC plasmid. Nano-S/MARt-Ele40 and Nano-S/MARt-UCOE produce the highest number colonies compare to all the version of pS/MARt that carry a bacterial backbone.

As expected the Nano-Vectors have a higher efficacy in generating established cells when compared to the respective pS/MARt. The experiment showed that the bacterial backbone had an effect and it might be the cause of the induction of the innate immunity or of chromatin induced silencing.

3.16 Application of Nano-S/MARt for the generation of Isogenic Patient Derived (PDX) pancreatic cancer cells

The efficacy of generating genetically engineered cell lines with the pS/MARt technology was previously demonstrated (figure 11). In order to evaluate the efficacy of the new Nano-S/MARt platform a Patient Derived Pancreatic cancer cell line (Paco-2) was selected. These cells were isolated from an aggressive pancreatic tumour, they form reliable xenograft models and they carry a deletion in the Smad4 genetic locus. Paco-2 cells are typically very difficult to transfect and transduce due to their high "inflammatory" state (Noll, Eisen et al. 2016). The same approach used for genetically modifying Capan-1 cells was applied for the Paco-2 cells testing the new nano vector system. The workflow was first set up generating a line expressing the reporter gene GFP and then, it was used to try the rescue of Smad4. These cells were derived from a patient with pancreatic ductal adenocarcinoma and upon isolation they were expanded in NSG mice. The established tumours were then kept in culture in a well-defined medium. In contrast to Capan-1 which are grown in medium supplemented with FBS, the Paco-2 cells are cultured without TGFβ. The rescue of the functionality of Smad4 leads to the reactivation of the TGFβ pathway, thus avoiding the risk of inducing alternative mutations.

Two different engineered cell lines were generated with the Nano-S/MARt technology. In the first the vector provided the GFP expression and in the second Smad4 expression was restored together with the reporter gene GFP. The modified cells were analysed via fluorescent microscopy, FACS and Western Blot (Figure 36). In the experiment performed with the Capan-1 no differences were observed between single cell clonal expansion and mixed populations, therefore for this experiment only mixed populations were expanded.

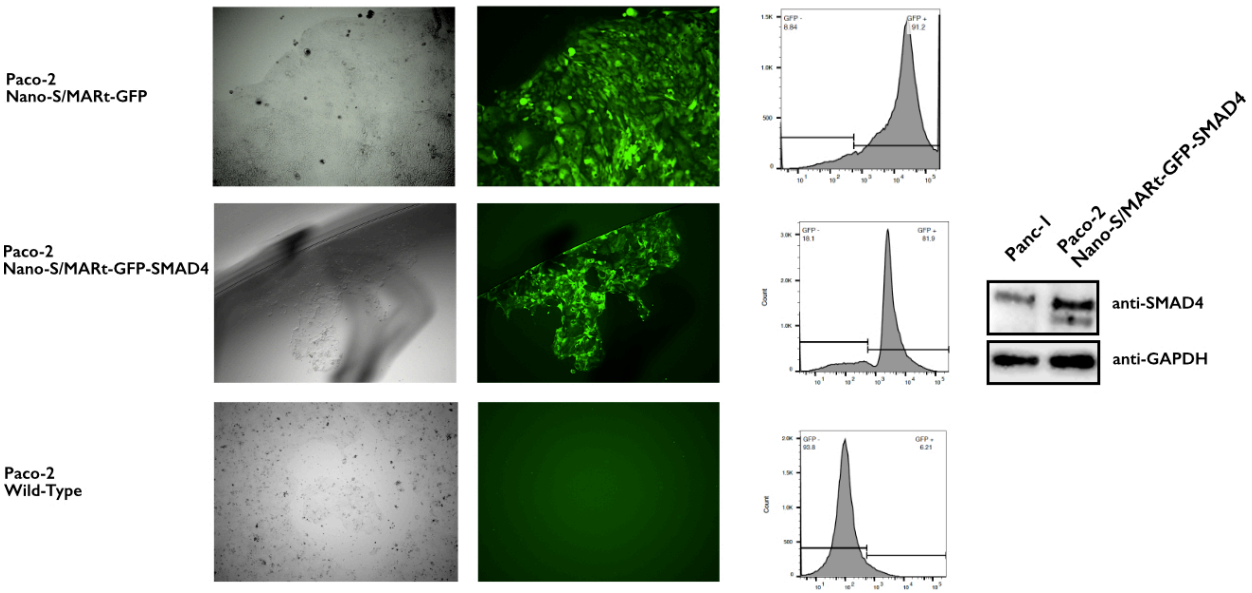


Figure 36. **Analysis of the genetically engineered Paco-2 cell lines with the NanoS/MARt vectors.** Paco-2 cells were transfected with NanoS/MARt-GFP and NanoS/MARt-Smad4-GFP and selected in presence of 0.5 μg/ml Puro for 1 week. After one single administration of Puromycin the cells were expanded and grown as mixed populations. The presence of positive cells was first determined via fluorescent microscopy and the populations were also analyzed by Flow Cytometry for the expression of the reporter gene GFP. The expression of the tumor suppressor Smad4 was determined with Western Blot and it was compared to the expression of the same gene in the Panc-1 cells. The experiments proved the functionality of the transgenic expression cassette *in vitro.*

The molecular integrity of the Nano-Vectors in modified Paco2 cells was determined via Southern Blot (Figure 37).

The analysis revealed that upon genomic DNA extraction and digestion with a single cutter, only one band of the same size of the respective linearized maxi prep control was detected. The absence of smires or alternative bands demonstrated the integrity as well as the stability of the new Nano technology.

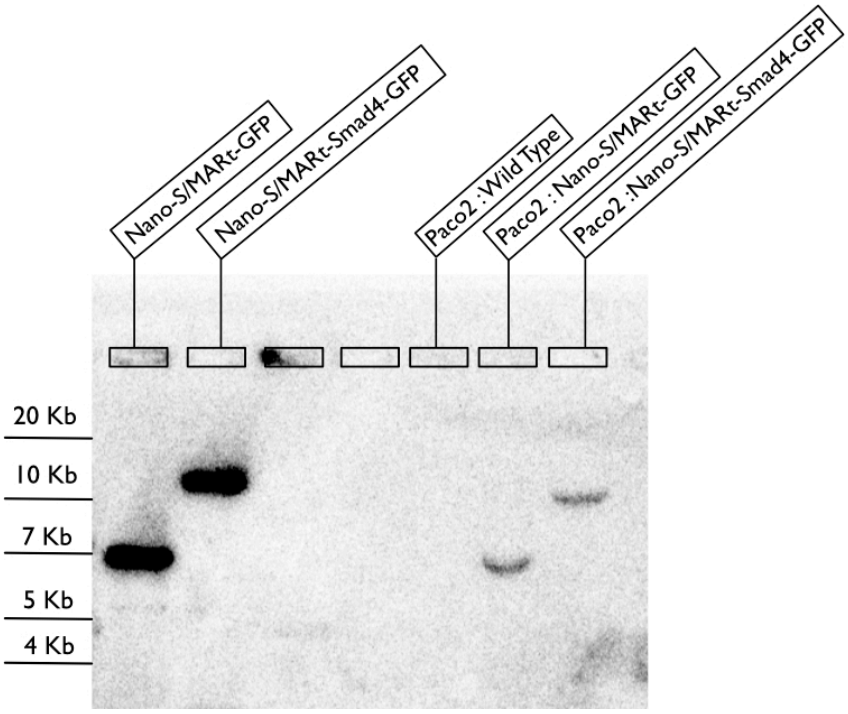


Figure 37. **Southern Blot analysis of the vector Nano-S/MAR-GFP and Nano-S/MAR-Smad4-GFP in Paco-2 cells.**
The maxipreps of the Nano-Vectors (first two lanes) and the total DNA from the modified Paco-2 cells were digested with the restriction enzyme *Bam*HI and separated on a 1% agarose gel before being transferred on a nylon membrane. The GFP reporter gene was used to generate the radioactive probe. The plasmids isolated via total DNA extraction from established cells show an identical size to their respective reference controls. The absence of alternative bands or smires confirmed their episomal maintenance

Microarray analysis was performed on RNA from cultured cells was extracted and analysed on a IlluminaHuman-12 Chip (Microarray analysis performed from the Genomic and Proteomic core facility, DKFZ). Four biological replicates were prepared per cell line and the differences in the gene

expression was analysed with the Partek Genomic Suite software (Thermo Fisher). The RNA expression levels were introduced into the non-rooted hierarchical clustering tool of the software and they were also tested in a Principal Component Analysis (PCA) assay (Figure 38).

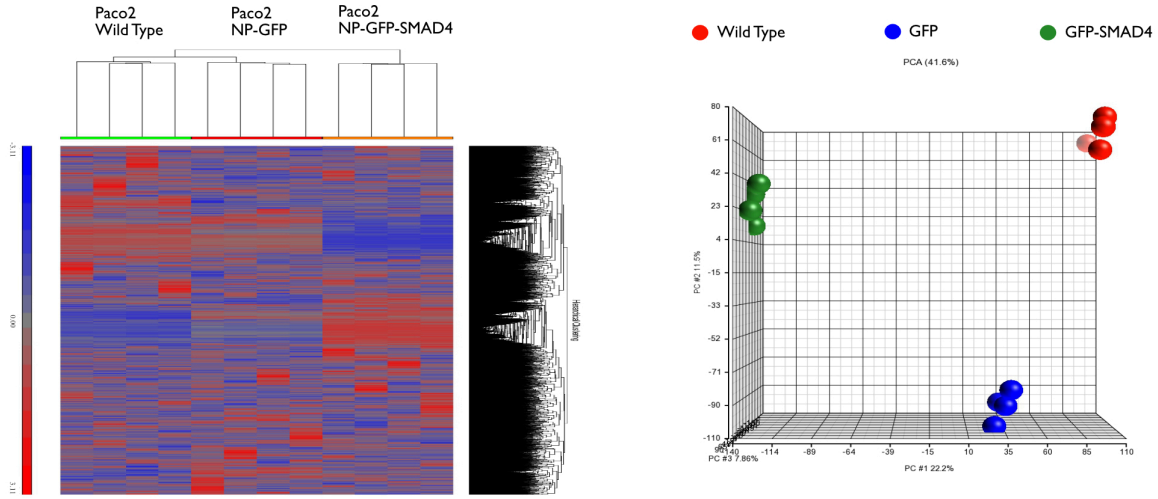


Figure 38. **Paco2 microarray analysis. Un-rooted hierarchical clustering and PCA.** Total RNA from pancreatic tumors was extracted (RNAeasy Kit, Qiagen) and the expression profiles were determined by microarray hybridization with a Human Illumina-12 Chip. Genome wide RNA expression values were analyzed with the Partek Genomic Suite software in a non-rooted hierarchical clustering test (left panel) and a PCA assay (right panel). Pancreatic tumors from three mice per group were considered for the study.

The un-rooted hierarchical clustering showed that the samples produced 3 clusters that corresponded to the different cell lines. The phylogenic tree generated by the software was composed of two branches. In one, the wild type unmodified cells clustered with the Paco-2-GFP divided into two nodes. In the other, the replicates for Paco2-Smad4 were grouped. The analysis of the PCA confirmed the differences between the cell lines. In contrast to the results obtained with the Capan-1, in this data set, Paco2 cells modified with the Nano-S/MAR-GFP vector and the parental unmodified cells clustered in the same branch. The GFP and the Smad4 expressing cells underwent the same selection process but the use of the new vector induced significantly fewer changes at the transcriptional level. To investigate which genes were changing, the RNA profiles

from these cell lines were compared with a cut off of >2 fold and <-2 fold with a FDR=0.1 (Figure 39). Only 5 genes appeared up-regulated in Paco-2 expressing the reporter gene GFP and modified with the Nano-S/MAR, whereas several hundreds were perturbed when Smad4 was re-introduced.

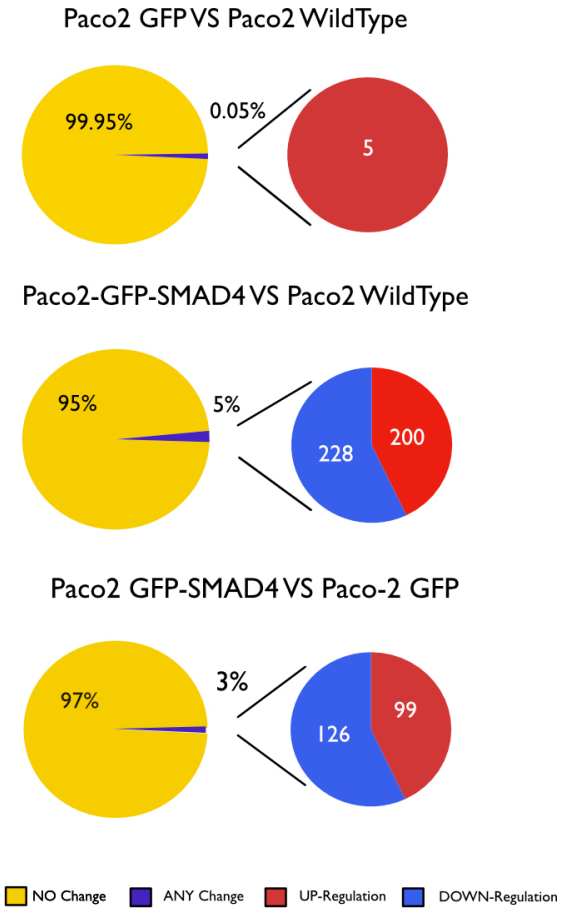


Figure 39. **Genome wide RNA profiles analysis on Nano-S/MARt modified Paco2 cells.** RNA from cultured cells was extracted and used for microarray analysis on a IlluminaHuman.12 chip. The array was performed in the DKFZ Genomic and Proteomic Core Facility and also the normalization across the samples was performed there. For the analysis, RNA from 4 independent extractions were analyzed per cell line. The gene expression was evaluated with the Partek Genomic Suite Software (Thermo Fisher). A cut off of > 2 fold and <-2 fold, FDR = 0.1 was applied to study the expression differences.

GSEA analysis were used to prove Smad4 functionality. The enrichment in two pathways which activity is strictly dependent from Smad4 were found. The EMT transition, also found in Capan-1-

Smad4 modified cells, and the TGFβ-pathway demonstrated the functionality of the Nano Vector technology (Figure 40).

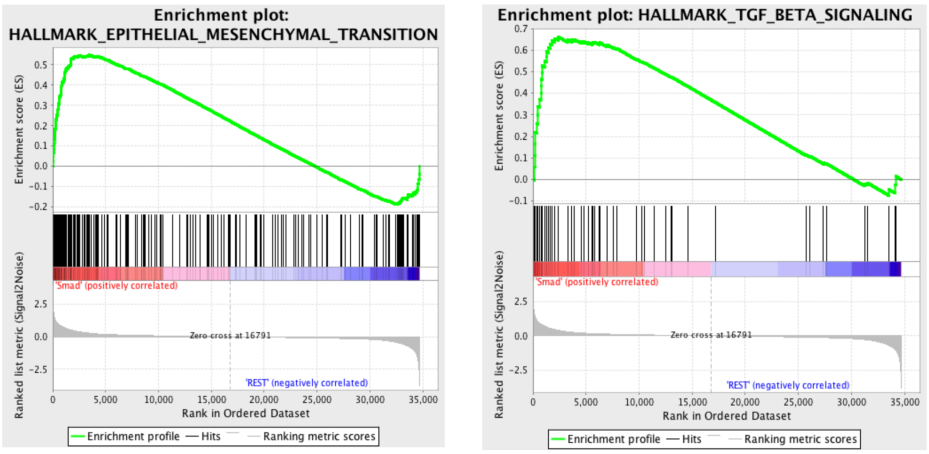


Figure 40. **GSEA analysis of Paco-2 Smad4-GFP modified cells with the Nano-S/MAR vector system.**
The GSEA analysis shows a strong enrichment for the hallmark underling the epithelial to mesenchymal transition and the activation of the TGFβ pathways in Paco-2 cells where the functionality of the tumor suppressor gene Smad4 was restored when compared to wild type Paco-2 cells and cells expressing the reporter gene GFP.

$0.5x10^5$ cells were then injected orthotopically into NSG mice and the xenografted tumours in the pancreas were evaluated for morphology, proliferation (Ki67 staining) and Smad4 expression (Figure 41). Tumours formed from cells engineered with the Nano-S/MARt providing the expression of the reporter gene GFP presented the same morphology to those formed with the unmodified parental Paco-2 cells. They were characterised by highly differentiated ductal structures and showed high recruitment of fibroblasts that stained positive for Smad4. In mice injected with cells where Smad4 was rescued no tumour masses were detectable. The histological analysis confirmed the presence of small, dormant and inert patches of cells which resulted positive for Smad4 expression but that did not show proliferation activity (negative cells for Ki67).

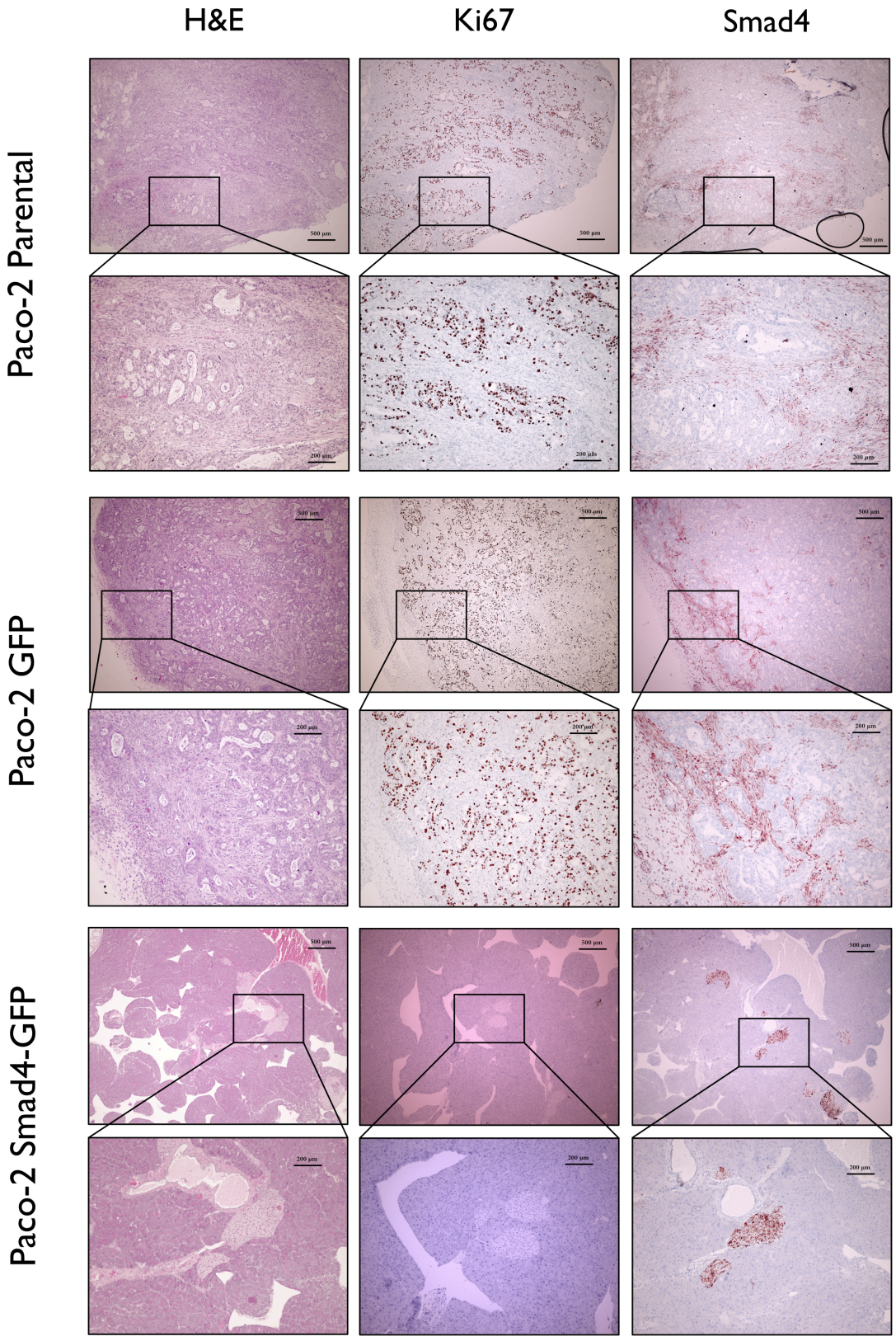


Figure 41. **Immunohistological assessment of primary tumors formed from the modified and parental Paco2 cells.**
Primary tumors obtained from the orthotopic injection of the modified and not modified Paco-2 cells were assessed for the phenotype, the expression of Smad4 and the proliferation via staining of Ki67.

The Paco-2 cells modified with the reporter gene maintained their original behaviour demonstrating that the episomal Nano Vector was not having any effect on the cells behaviour nor on their molecular integrity. Therefore all the observed effect in the Smad4 rescued cells were attributed to the functionality of the tumour suppressor gene.

3.17 Nano-spliced-S/MAR vector: a DNA vector that mimics a mammalian genomic endogenous gene

It was mentioned previously that in order to be functional in a plasmid the S/MAR region has to be placed in a transcriptionally active expression cassette. However, it was also demonstrated that transcripts that contain this sequence or those generated from cDNA are less stable than those produced from a cell's genome. Normally, an RNA that encodes for a protein is composed of characteristic intro-exon structures and it undergoes a splicing process in which the non-coding sequences are removed. This mechanism is mediated by proteins that remain bound to the mature mRNA, stabilising it until it is translated into a polypeptide. In order to mimic the RNA maturation process that normally takes place in eukaryotic cells, the sequences necessary for the splicing were added before and after the S/MAR motif in the pS/MARt vector. The idea was to generate a pre-mRNA that started from the leader sequence of the expression cassette after the promoter and ended after the S/MAR and presented all the sites necessary to recruit the splicing proteins (Figure 42).

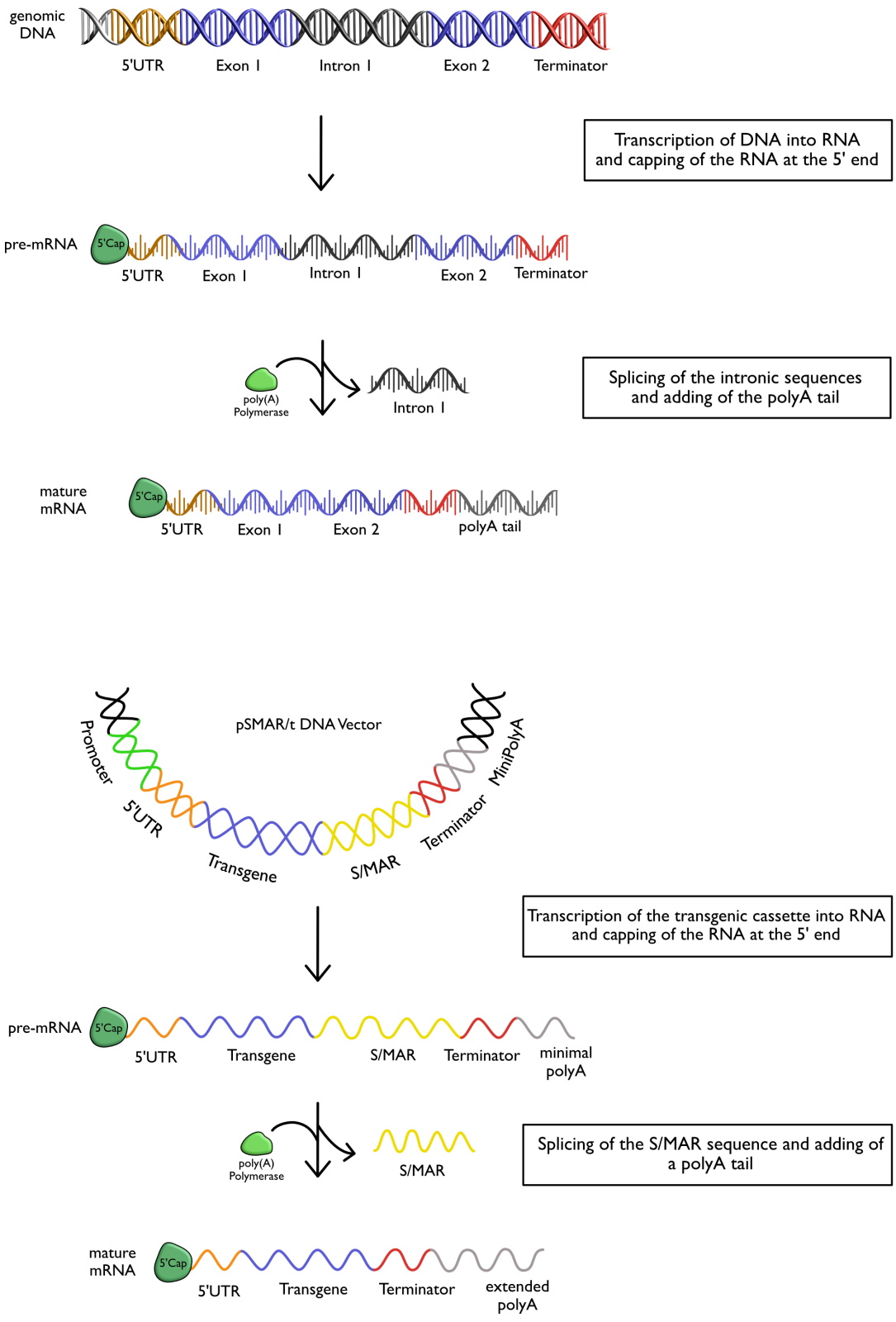


Figure 42. **Schematic representation of the RNA maturation process.**
In the upper panel a representation of the process of RNA maturation in a eukaryotic cell is represented. In the lower one a representation of how the pS/MARt vector will mimic this procedure is displayed.

The vector was named Nano-S/MAR-splice and it was evaluated in Hek293T cells in comparison with the most up to date respective Nano-S/MAR vector. Cells were established with the standard protocols which required the application of Puromycin (0.5 μg/ml) for one week before expansion for at least 30 days. The established populations were analysed for the expression of the reporter gene GFP via FACS and the GFP RNA levels were evaluated via qPCR (Figure 43).

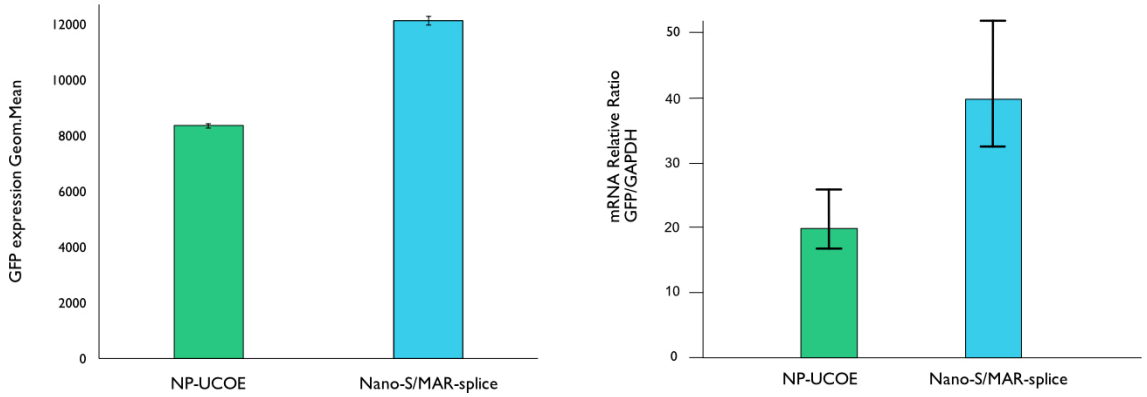


Figure 43. **Nano-S/MAR-splice GFP expression Vs Nano-S/MAR.**
The GFP expression was evaluated in the established population via FACS and the RNA levels were then confirmed via qPCR. The expression of the reporter gene was compared to the expression of the endogenous gene GAPDH. The fluorescent intensity as well as the relative RNA levels in cells established with the Nano-S/MAR-splice vector are higher than in cells treated with the respective un-spliced version of the plasmid.

The reporter gene expression in cells established with the Nano-S/MAR-splice vector was higher than in cells established with the non-spliced MAR plasmid. The maturation of the RNA with the binding of the splicing proteins increased the mRNA stability and corresponding expression. The efficacy in establishing cells was also tested in HEK293T through colony forming assay and it was demonstrated (Figure 13) that this plasmid had the highest efficacy in generating established cells producing the highest number of colonies.

This preliminary data encouraged further experiments in which the molecular events that undergo the transgenic RNA generated from this vector have yet to be validated.

3.18 pS/MARter: the ultimate design of S/MAR DNA plasmid vectors.

Broll et al. (Broll, Oumard et al. 2010) demonstrated that a minicircle harbouring a human β-interferon S/MAR sequence undergoes spontaneous rearrangements and that these structural changes take place within the MAR element. They concluded that during cell replication and segregation of the chromosomes the fragile MAR sequence was readily damaged and was reduced to a minimally sized (200-300bp) attachment sequence. They hypothesised this to be an evolutionary process because it was found that the breaking point was always in the 5' region of the genetic element.

Since 1977 it has been known that the nuclear matrix organizes the 25 million nucleosomes in a single mammalian nucleus forming approximately 60.000 chromatin loops. This is achieved by the binding of nuclear matrix proteins to characteristic DNA landmarks in introns as well as proximal and distal flanking 5' and 3' ends of genes. (Paulson and Laemmli 1977, Gasser and Laemmli 1986). MAR, Ori and homeotic protein binding sites share common DNA sequence motifs (Boulikas 1992). In particular, the ATTA and ATTTA motifs which constitute the core elements recognized by the homeobox domain from different species frequently occur in the matrix attachment sites of several genes as well being present in the sequences of different origins of replication such as yeast and viruses. The sequence of the β-INF MAR isolated by Lipps and collaborators in 1999 and used as anchoring site for the DNA plasmid, its replication and episomal maintenance (Figure 44) was analyzed for the presence of these recurrent motifs.

TAATGAATGTCTAAGTTAATGCAGAAACGGAGAGACATACTATATTCATGAACTAAAAGACTTAATATTGTGA
AGGTATACTTTCTTTCCACATAAATTTGTAGTCAATATGTTCACCCCAAAAAAGCTGTTTGTTAACTTGCCAA
CCTCATTCTAAAATGTATATAGAAGCCCAAAAGACAATAACAAAAATATTCTTGTAGAACAAAATGGGAAAGA
ATGTTCCACTAAATATCAAGATTTAGAGCAAAGCATGAGATGTGTGGGGATAGACAGTGAGGCTGATAAAATA
GAGTAGAGCTCAGAAACAGACCCATTGATATATGTAAGTGACCTATGAAAAAAATATGGCATTTTACAATGGG
AAAATGATGATCTTTTTCTTTTTTAGAAAAACAGGGAAATATATTTATATGTAAAAAATAAAAGGGAACCCAT
ATGTCATACCATACACACAAAAAAATTCCAGTGAATTATAAGTCTAAATGGAGAAGGCAAAACTTTAAATCTT
TTAGAAAATAATATAGAAGCATGCCATCATGACTTCAGTGTAGAGAAAAATTTCTTATGACTCAAAGTCCTAA
CCACAAAGAAAAGATTGTTAATTAGATTGCATGAATATTAAGACTTATTTTTAAAATTAAAAAACCATTAAGA
AAAGTCAGGCCATAGAATGACAGAAAATATTTGCAACACCCCAGTAAAGAGAATTGTAATATGCAGATTATAA
AAAGAAGTCTTACAAATCAGTAAAAAATAAAACTAGACAAAAATTTGAACAGATGAAAGAGAAACTCTAAATA
ATCATTACACATGAGAAACTCAATCTCAGAAATCAGAGAACTATCATTGCATATACACTAAATTAGAGAAATA
TTAAAAGGCTAAGTAACATCTGTGGCAATATTGATGGTATATAACCTTGATATGATGTGATGAGAACAGTACT
TTACCCCATGGGCTTCCTCCCCAAACCCTTACCCCAGTATAAATCATGACAAATATACTTTAAAAACCATTAC
CCTATATCTAACCAGTACTCCTCAAAACTGTCAAGGTCATCAAAAATAAGAAAAGTCTGAGGAACTGTCAAAA
CTAAGAGGAACCCAAGGAGACATGAGAATTATATGTAATGTGGCATTCTGAATGAGATCCCAGAACAGAAAAA
GAACAGTAGCTAAAAAACTAATGAAATATAAATAAAGTTTGAACTTTAGTTTTTTTTAAAAAAGAGTAGCATT
AACACGGCAAAGCCATTTTCATATTTTTCTTGAACATTAAGTACAAGTCTATAATTAAAAATTTTTTAAATGT
AGTCTGGAACATTGCCAGAAACAGAAGTACAGCAGCTATCTGTGCTGTCGCCTAACTATCCATAGCTGATTGG
TCTAAAATGAGATACATCAACGCTCCTCCATGTTTTTTGTTTTCTTTTTAAATGAAAAACTTTATTTTTTAAG
AGGAGTTTCAGGTTCATAGCAAAATTGAGAGGAAGGTACATTCAAGCTGAGGAAGTTTTCCTCTATTCCTAGT
TTACTGAGAGATTGCATCATGAATGGGTGTTAAATTTTGTCAAATGCTTTTTCTGTGTCTATCAATATGACCA
TGTGATTTTCTTCTTTAACCTGTTGATGGGACAAATTACGTTAATTGATTTTCAAACGTTGAACCACCCTTAC
ATATCTGGAATAAATTCTACTTGGTTGTGGTGTATATTTTTTGATACATTCTTGGATTCTTTTTGCTAATATT
TTGTTGAAAATGTTTGTATCTTTGTTCATGAGAGATATTGGTCTGTTGTTTTCTTTTCTTGTAATGTCATTTT
CTAGTTCCGGTATTAAGGTAATGCTGGCCTAGTTGAATGATTTAGGAAGTATTCCCTCTGCTTCTGTCTTCTG
AAAGAGATTGTAGAAAGTTGATACAATTTTTTTTTCTTTAAATATTTGATA

Figure 44. **DNA sequence of the β-Interferon MAR from Homo Sapiens**.
The DNA binding motives ATTA is presented in green and the ATTTA in dark yellow.

Although in this MAR the motifs are particularly enriched (18 motifs in 1955 bp), they seem randomly distributed within the sequence and no conserved structure could be identified. We wondered whether a better MAR could be found and we identified that the ApoB gene might be a good candidate. In comparison to the β-interferon MAR, the one that is part of the human

apolipoprotein B gene is almost entirely composed of a contiguous stretch of 555 bp comprising of a mosaic of the ATTA-ATTTA (Figure 45).

TTAAGAGCCTAAAGAGCATACATGTATGCTGGGTCACAGAATTGTTAGGAATCTACATCTCCACCCCTCCTCT
CCCAATCATACTTTCTTGGCCTTGAATGGATCCTGGCAGAGCTCCAGGGAGACATCTGGGGTCCGTATTGCCA
TGAAGCCCCTGGGGCTAGGACTCCCTAGCCATTCCTTCTCCACTCCTGGCAGGCTGAGTGAAATAAAGGACTT
GTTATTTCATCTCGAGGCCTACCGGAGAGCCTTGCCTTGCAAAGGCAGACAGTCAGTGAGGAAGACTATGTGG
CACATGAAGACACCAGAGGTGTTCCTCAGGATCAAAGTATGTACAAGCCTTTGTGAATATTTTTTCCTTCTCA
CTTGGCAAATACAATTCCTGAGATCAATAACCTCGTCTTTTTAATTTTTTCCTCGTCTTTTTAACTATTTATA
AAATATTGAATTATAAAATATGTAATTATAAATACTTTAATTATAAAATATGTAATTATAAATACTTTAATTA
TAAAATATGTAATTATAAATACTTTATAAAATATGTAATTATAAAATATGTAATTATAAACATTTTAATTATA
AAATATGTAATTATAAACATTTTAATTATAAAATATGTAATTATAAACATTTTAATTATAAAATATGTAATTA
TAAACATTTTAATTATAAAATATGTAATTATAAACATTTTAATTATAAAATATTTAATTATAAACATTTTAAT
TATAAAATATTTAATTATAAATATTTTAATTATAAAATATTTAATTATAAATATTTTAATTATAAAATATTTA
ATTATAAATATTTTAATTATAAAATATTTAATTATAAATACTTTAATTATAAAATATTTAATTATAAATATTT
TAATTATAAAATATTTAATTATAAATATTTTAATTATAAATATTTTAATTATAAAATATTTAATTATAAAAAC
ACAATTACCTCATCTTTTTAAATATTTTTGCAAAATATTTCCCTCCATAATTTCTCCGTTTCCATTTTTATTC
TGTTACTTAAATCACACTATGTGTTCTAGAGGTTTTGCTGTGCCAGAACATTTTATCAATGCCCTCGTTTCAC
TGTCTTTCAATACAAATGAGCCACATTCAGTGGTATGATACACAATAAAGACTCCATTTATTTGTTCCTCCTC
CCCCAAGTTTAGCAAAATAACTCAGATCCTGATTTTCTTTAACTTGCAAAAAATGCCATCCTTCTGAGTTCAG
AGACCTTCCGAGCCCTGGTGCCAGCTTTGGTGCAGGTCCAGTTCATATGTGCTTCTGCTTATAGTCTACTGCC
TACTGCAAGGCTGGCTCACTGTATGGTTTTATCAATATAGGCAGTTTGAATTTTTTCTGTGCTATGTGAAAGT
TCAATTGGAAAAGAAGAATAAATGAAGATTTCTTTTAAAAAATTAGAGGATGATAGTAAGTTCTCCTGGAGCA
AGCTTCATGTAGGGGTTCATGACTGTGGTTGATTGCAGCTTTTTCAGTAACTCCGTGATGTATATCAGAAATG
TGTGGTAGTTTTGAATGGACAGGTCAATCAATCTTTTGGATTCAGCAATAAATTTTTCATAGTAATCAGAGAG
TTGGTCTGAAAAATCTTGCAGTTTATATCTAAACT

Figure 45. **DNA sequence of the Apolipoprotein B MAR from Homo Sapiens**.
The ATTA motif is displayed in green, whereas the dark yellow show the ATTTA sequence. In light blue is shown the DNA sequence that was use for the cloning of this DNA stretch into the pS/MARt backbone.

45 motifs are present in this 1600 bp DNA sequence and a repetitive conserved structure can be detected. The 11 bp sequence TAAATATTTTA divides two ATTA domains and the this very defined conserved structure may form an ideal binding site for the formation of stable DNA-protein complexes. Due to the high CpG content of the 5' and 3' end of the sequence, only the core

repetitive domain of this MAR was isolated and cloned into pS/MARt backbone substituting the original MAR element. Besides one or two trials, the possibility of testing a novel sequence has not been exploited because it was demonstrated that alternative smaller MARs were not as efficient as the β-interferon one. The highly conserved structure of the Apolipoprotein MAR, suggested that this sequence might also act as a anchoring sequence for a DNA vector. Its potential in mediating the extrachromosomal replication of the vector was then tested. The episome harboring this novel MAR was named pSMARter and its capability in genetically modifying dividing cells was tested in HEK293T. In a first preliminary test, cells transfected with pS/MARter were cultured in presence of Puromycin (0.5 μg/ml) for 1 week and expanded in absence of selection. The established cells were tested for the expression of the reporter gene GFP via fluorescent microscopy and the molecular integrity of the new plasmid was investigated via plasmid rescue (Figure 46). The cells resulted positive for the expression of the reporter gene GFP and the episomal maintenance of the vector was demonstrated. The gDNA from the modified HEK293T cells was extracted and it was used to transform DH10B *E.coli.* 12 bacterial colonies were picked, digested with the enzyme *Bam*HI (a single cutter in the plasmid sequence) and analyzed on an agarose gel. All the colonies showed the same restriction pattern and it was equivalent to the one obtained from the digestion of the maxi preparation used to transfect the cells at the beginning of the procedure.

This experiments represent for the first time that a novel, functional alternative MAR sequence is able to mediate the episomal maintenance and extrachromosomal replication of plasmid DNA in the nucleus of actively dividing cells. The possibility of using this new anchoring sequence as a base for a novel vector platform was exploited and its efficacy in establishing cells was compared to pS/MARt and to the Nano-S/MAR series of plasmids in a colony forming assay (Figure 13). The vector based on the novel MAR sequence established cells with a lower efficiency than the Nano-S/MAR-splice plasmid but it was as efficient as the other Nano vectors and better than the pS/MARt

series. Considering that this plasmid is still in a "normal" version with a bacterial backbone, the introduction of this new anchoring sequence dramatically increases its efficacy in establishing cells.

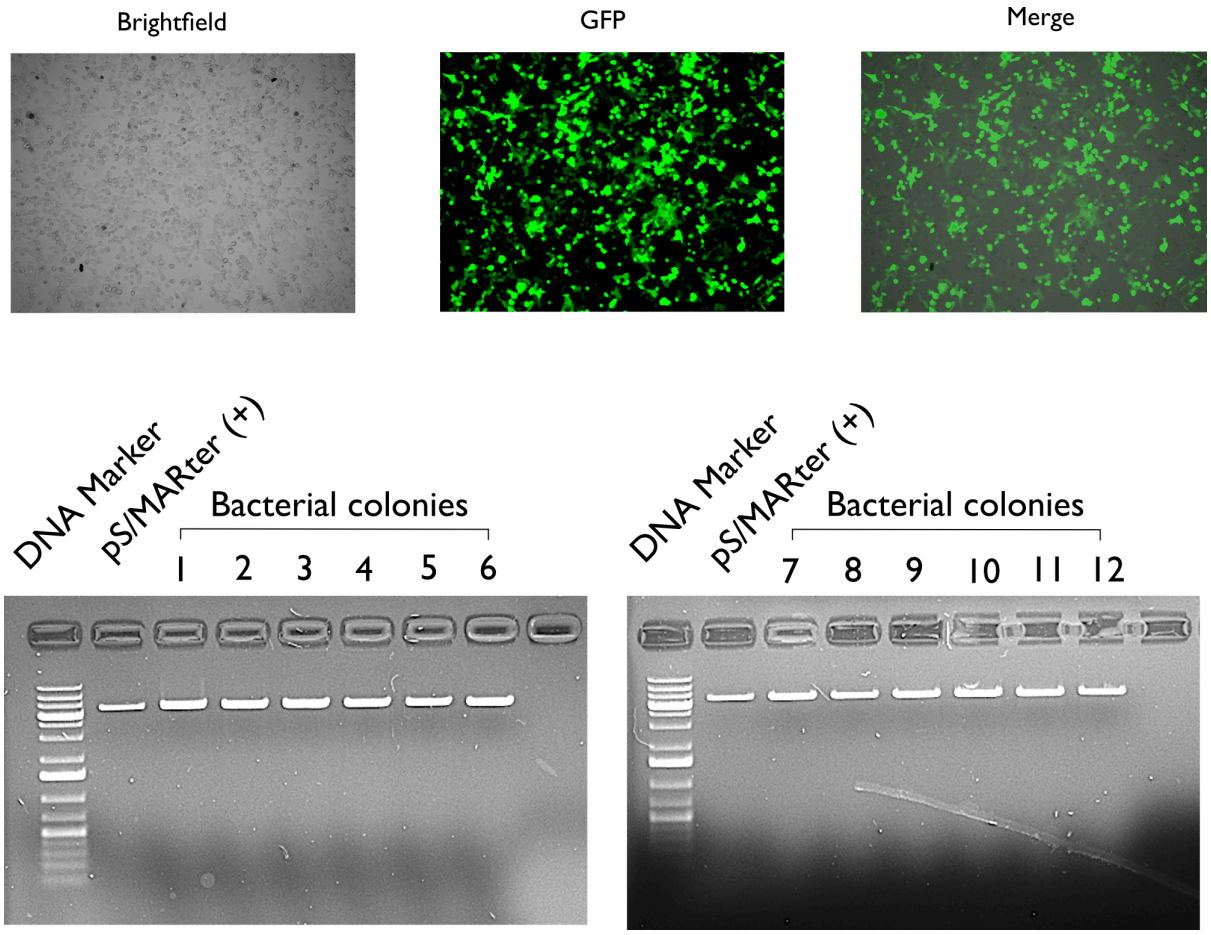


Figure 46. **Testing of pS/MARter in HEK293T cells and analysis of the molecular integrity of the vector upon cell establishment.**
Hek293T cells were transfected with the plasmid pS/MARter and selected with Puro (0.5 μg/ml) for 1 week. After the selection round the cells were grown in absence of Puro and expanded. An established mix population was tested via fluorescent microscopy for the expression of the reporter gene GFP and the molecular integrity of the plasmid was assessed via plasmid rescue. For the plasmid rescue the gDNA from established HEK293T was extracted with the Blood&Tissue DNAeasy kit (Qiagen) and transformed into DH10B *E.Coli*. The bacterial were grown on LB-Agar plates with Kanamycin. The resulting colonies were picked and grown in liquid LB medium with Kan overnight and the plasmid DNA was extracted with the MiniprepKit (Qiagen). For the analysis of the vector, the DNA mini preparations were digested with the restriction enzyme *Bam*HI (Thermo Fisher) and the restriction pattern was addressed on a 1% agarose gel. As control the DNA used for transfecting the cells at the beginning of the establishment procedure was digested with the same enzyme and run as a reference. All the rescued plasmid show an identical restriction pattern to the one generate with the control vector.

3.19 An ultimate test of S/MAR vector activity – Genetic Engineering of primary human T Cells

Immunotherapy is currently one of the most exciting and successful therapeutic strategies for the treatment of cancer. Recently, the FDA approved Kymriah™, the first cancer therapy with *in vitro* engineered T Cells for the treatment of Acute Lymphocytic Leukemia (ALL). Currently the genetic modification of these cells is achieved thorough the integrative lentivirus or the sleeping beauty transposon systems. Although the T cells modification via viruses is considered safe, severe side effects in the long term such as those already described by *Hacein, Von Kalle et al.* (Hacein-Bey-Abina, Von Kalle et al. 2003) represent the major risk for this treatments.

The new S/MAR technology was challenged for its capability of generating genetically modified human primary T cells. As a proof of concept the sustained and stable expression of the reporter gene GFP was investigated. pS/MARt was delivered into freshly isolated Peripheral Mononucleated Blood Cells (PBMC) and the GFP expression was monitored for 12 days (Figure 47).

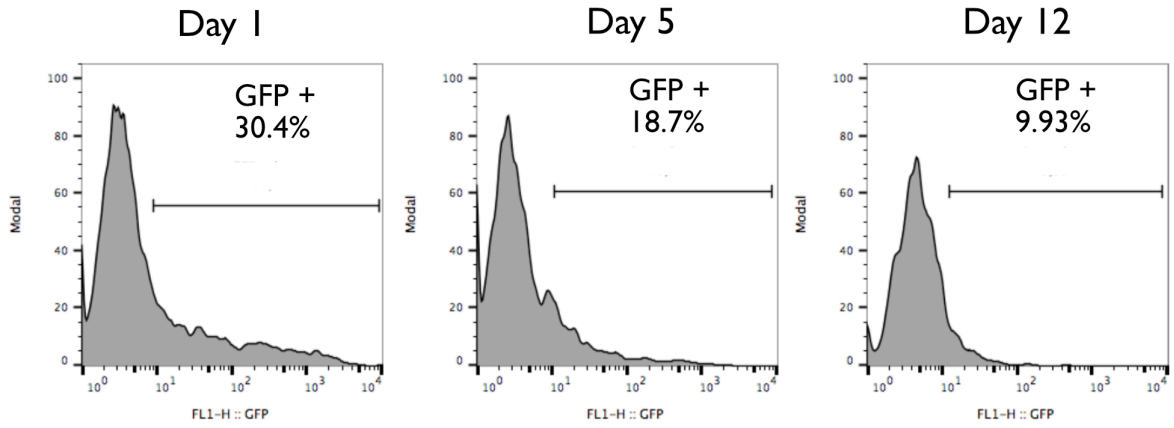


Figure 47**. CD8+ cells expressing the reporter gene GFP.**
pS/MARt vectors were delivered into freshly isolated PBMC via electro transfer (Amaxa 4D Nucleofector, Lonza) and the GFP expression was monitored over a 12 days time window. The plasmid is not able to sustain the stable expression of the reporter gene that drop dramatically within 12 days of delivery.

The DNA vectors were delivered via electro transfer (Amaxa 4D Nucleofector, Lonza) and 6 h after the electroporation the cells were cultured in wells coated with human anti CD3/CD28 and interleukine-2 (IL-2). This culturing system allowed the expansion and the proliferation of only cytotoxic lymphocytes CD8+. The number of cells expressing the reporter gene over was monitored once a week via FACS . The efficiency of the DNA delivery was approximately 30% and the number of cells expressing the transgene decreased constantly within the first two weeks of delivery. The experiment was repeated with plasmid in which the expression of the reporter gene GFP was driven from a different promoter such as, PGK and EF1α but the same trend was observed (data not shown). pS/MARter and Nano-S/MARt were then challenged in these difficult primary cells. The same approach was used and upon DNA delivery into freshly isolated PBMC, the GFP expression was monitored for 34 days (Figure 48).

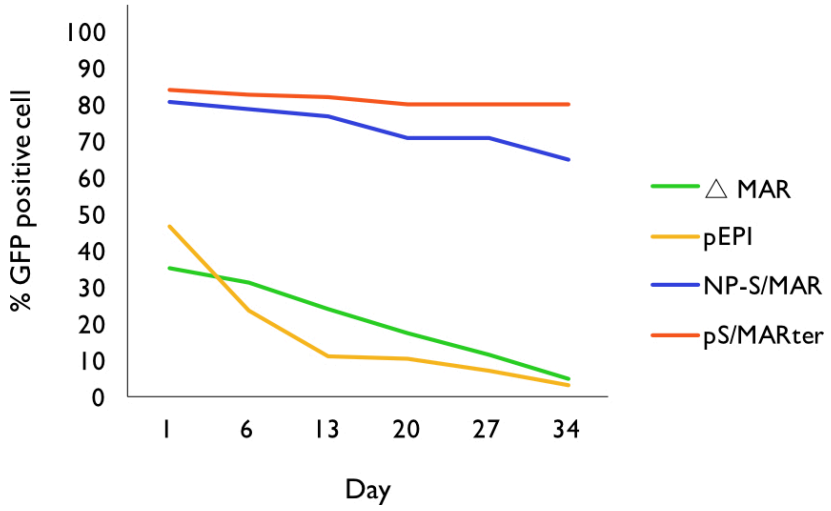


Figure 48. **GFP expression in primary human T cells.**
Four different vectors were tested for their capability in sustaining the expression of the reporter gene GFP in human T cells. The plasmids were delivered to freshly isolated PBMCs via electro-transfer (Nucleofector Device Y, Lonza) and the cells were cultured in presence of IL-2 (5μg/ml, Biolegend) for 34 days . Every 7 days the cells were checked for the transgene expression and their growth was stimulated via addition in the media of the antibody anti-CD28(Bioegend) and anti- CD3 (Biolegend). The Nano-S/MAR vector is able to sustain the expression of the reporter gene for 34 days in these cells but although it has a small decline during the last week. Instead pS/MARter provided more stable and persistent transgene expression in freshly isolated PBMC for at least 34 days.

These plasmids were able to sustain the expression of the reporter gene for the whole period of the experiment, whereas the plasmids lacking the MAR and pEPI showed a dramatic decline within

10 days of delivery. The GFP reporter gene was swapped for the transgenic T Cell Receptor (TCR) MART1. This TCR is able to recognize and bind cells expressing the peptide MelanA, a characteristic epitope in melanoma cancer cells. In contrary to what observed with the reporter gene, when the vector expressing the TCR MART1 was introduced into T cells, the receptor expression was sustained at high levels only for a short period of time (Figure 49). The possibility of the silencing of the transgene induced by the promoter was tested. Plasmids in which MART1 expression was driven by the CMV, EF1α, CAG, PGK and ROSA26 promoter were tested but they resulted in the same expression trend.

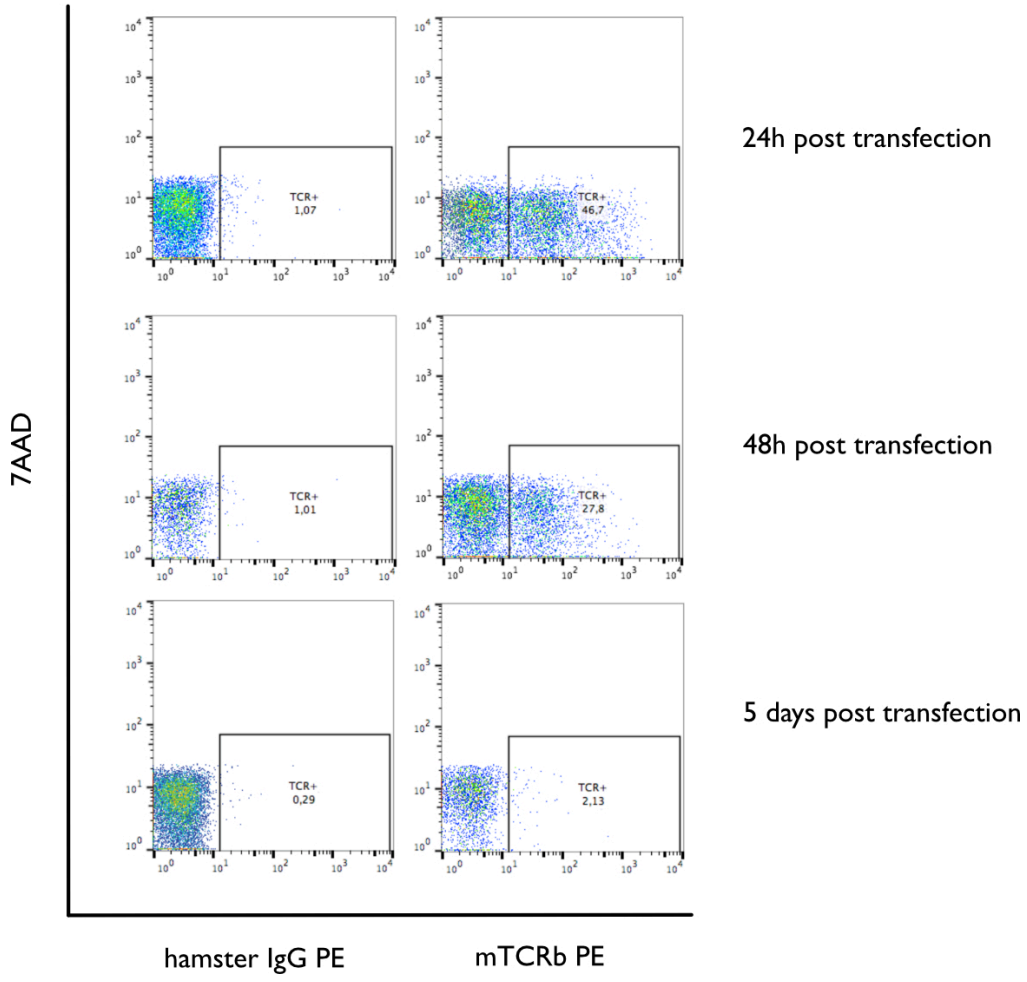


Figure 49. **MART1 TCR expression in human T cells.**
pS/MARter expressing the TCR receptor MART1 was delivered to freshly isolated PBMCs. The TCR expression was monitored via FACS. The expression of the receptor was measured with the antibody mTCRb PE (Biolegend) in relation to isotype antibody control (Hamster IgG PE, Biolegend). The plasmid can not provide a stable expression of the T Cell Receptor that decreases constantly 5 days of delivery.

The fact that the vector is able to sustain the expression of the transgene GFP for over a month demonstrate that the technology is suitable for genetic engineering of this cells but, the short expression of the TCR demonstrate also that further optimizations is required.

This DNA vector technology would provide several advantages over the currently used systems. It is cheaper, easier and more safe to produce and manipulate. The high safety profile is given from their reduced toxicity as they avoid potential side effects induced by random integration. It is proposed a novel technology for the safe introduction of T-Cell receptors or Chimeric Antigen Receptor (TCRs or CARTs) into naive Human T-Cells with the scope of using them for autologous-immunotherapy.

4 Discussion

4.1 Current state of art for the persistent modification of cells. Drawbacks and limitations.

The potential to genetically engineer mammalian cells with non-integrative systems represents an attractive field of research. Many viruses have been attenuated and utilized as vectors for the transfer of genetic material into cells. This process takes advantage of the infectious nature of viruses and modified lentivirus, adenovirus and others are widely used for persistently modifying cells in research and also for therapeutic applications. The inherent nature of these viral vectors can and does lead to potential problems such as unexpected cellular proliferation induced by the random integration of the viral genome (Hacein-Bey-Abina, Von Kalle et al. 2003) and/or cell transformation. It was reported from *Henderson et al.* (Henderson, Rowe et al. 1991) that some viral proteins such as the latent membrane protein 1 (LMP1) from the EBV virus is able to upregulate the expression of the cellular endogenous genes *bcl-2* inducing an uncontrolled proliferation. Even episomally maintained viruses such as Adenovirus were shown to cause severe side effects inducing strong adaptive immune responses against the viral capsid (Gregory, Nazir et al. 2011). However, despite the intrinsic risks related to the use of engineered infectious agents, cells genetically modified with viruses are still wildly use for basic research and in clinical trials. Recently, the FDA approved the autologous treatment of leukemia with lymphocytes modified with an integrative lentivirus.

The necessity of developing alternative less toxic vectors has driven the creation of novel approaches such as CRISPR/Cas9 and sleeping beauty transposons. These alternative systems in some cases do not require delivery via viruses which potentially reduces their capacity for developing severe side effects. However, in the sleeping beauty transposons, to produce a

mammalian cassette that is constitutively expressed the genetic material has to be integrated into the cellular genome. The nature of the integration remains random and the expression of the gene of interest can be very different within cells from the same population. Integrated DNA can be silenced and it can induce *cis* activation or repression of surrounding genes. The CRISPR/Cas9 system instead, is a technique that allows the precise editing of cellular genome. It was shown to be very efficient when it was applied in single nucleotide modification (Cong, Ran et al. 2013) but it showed its limitations when it was tested for the re-introduction or deletion of genes. It was estimated that the efficiency in re-introducing a gene in its locus was about 5% (Paquet, Kwart et al. 2016). The possibility of off target effects and ethical discussions regarding the application of genome editing techniques in humans still represent major problems.

4.2 pEPI: a non viral system that lacks stability

The autonomous replicon pEPI vector which was first described in 1999 (Piechaczek, Fetzer et al. 1999) represented an opportunity to develop alternative vectors for the field of cell engineering and gene therapy. It opened the possibility of modifying cells with a DNA plasmid platform that was able to provide sustained transgene expression and autonomous extrachromosomal replication. This prototype vector was described to be efficient in genetically modified providing persistent transgenes expression in a range of cells *in vitro*, and it was also shown to be able to sustain the long term expression of genes when delivered directly *in vivo* in the liver of mice.

When we started to develop and apply this vector platform with the aim of generating isogenic cell lines, we realized that the efficacy of this technology in generating persistently modified cell lines was very low. The major problem faced was related to the instability of the plasmid once delivered into the cells. We demonstrated that a significant proportion of cells in a population that was kept under antibiotic selection for several weeks was negative for GFP transgene expression (Figure 7)

when tested via Flow Cytometry. We decided then to investigate whether the problem of such a low number of positive cells was related to the vector stability. We hypothesized that once delivered into the cells the pEPI plasmid was undergoing random rearrangement. A plasmid rescue assay (Figure 8), showed that the plasmids that were retrieved from established cells didn't have the same molecular size as the control vector. pEPI derived DNA was still maintained in its episomal form but surprisingly the rescued plasmids were smaller with a different restriction pattern which confirmed the hypothesis of instability. Probably the constant selection pressure applied to those cells, selected those vectors that had a functional antibiotic cassette but not transgene expression.

4.3 The potential of minicircle technology

Minicircles have been shown to provide higher and more sustained transgene expression (Bigger, Tolmachov et al. 2001). The generation of S/MAR minicircles further improved the DNA vector technology for its application in animal studies and it also increased its efficacy *in vitro*. These class of vectors are produced in bacterial cells but they lack the bacterial backbone. The introduction of two Flp sites before and after the mammalian expression cassette allow, upon *Cre recombinase* induction, the production of a vector that comprises exclusively the promoter-transgene-S/MAR cassette. The prototype pEPI as well as the minicircle establish within a certain number of vector copies but the expression levels of the minicircle vectors are higher when compared to normal plasmids. When S/MAR minicircle were delivered to the liver of mice (Argyros, Wong et al. 2011) they were shown to provide a longer and higher level of expression of transgenes which leads to the suggestion that epigenetic features and DNA composition strongly influence the regulation of the expression from a non-viral episome. Minicircles then, minimize the potential innate immune reaction due to a reduction in their CpG content which is typically very abundant in the pUC Ori or in the antibiotic resistance gene. Intracellular nucleic acid sensors like the TLR9 can bind to

unmethylated CpG islands which leads to an innate immune response. Small interfering RNAs then can be generated from features of the bacterial backbone, such as the Ori, and they can lead also to transcriptional inactivation (Chen, He et al. 2003). Although this class of DNA vectors was demonstrated to be particularly efficient, restricted access to this technology, limited their application and development.

4.4 pS/MARt - a stable DNA Vector episomally maintained in dividing cells

Initially the aim of this project was to generate a range of luciferase labelled isogenic cell lines with the S/MAR plasmid pEPI and to utilize those for testing experimental drug compounds *in vitro* and in *in vivo* xenograft models. The focus of the DNA vector lab was, however, on the development of this DNA vector as a platform for a range of applications. As we became aware of the limitations and inefficiency of the original prototype pEPI vector such as low establishment rate and its instability issues we shifted the attention of this project towards the development of a new S/MAR DNA vector platform which could overcome these limitations.

Primarily, the idea was to produce a tool which could be used to consistently and easily generate genetically modified cell lines of high purity and homogenicity, in which the number of non-expressing cells was reduced to a insignificant fraction.

The first modification we made to the vector was rather simple but proved significant to the functionality of the S/MAR. In the prototype pEPI vectors the mammalian selection marker was combined with the bacterial gene and this was expressed from within the bacterial backbone. To all extents of purpose, the expression of this gene had no correlation to the activity of the S/MAR and as we discovered it could be active providing drug resistance even in the absence of the transgene cassette. We decoupled the expression of the transgene and linked it to the expression of the drug

resistance Puromycin (Figure 9). This small change overcame several limitations of the previous vector system. It allowed the possibility of fine regulation of the transgene without limiting the functionality of the S/MAR motif. It also improved the establishment of the cellular clones making their formation part of an active process (Figure 10) which can be directly controlled by levels of selection pressure applied to the cells. Our next aim was to boost and sustain the expression of the cassette that drives the S/MAR. Genomic elements, such as the anti-repressive element 40 (Kwaks, Barnett et al. 2003) were introduced into the vector to provide insulating boundaries within the construct that prevent its silencing through the spread of inactive chromatin from the nearby bacterial backbone. It was introduced before the mammalian promoter and the protection from this element provided to the expression cassette, increased the S/MAR vector establishment efficiency by 10 fold when compared to the pEPI vector prototype (Figure 11) as well as the transgene expression by 3 fold. The anti-repressive element 40 was discovered as a genetic element that is able to block chromatin- associated repressors. How this element counteract distinct repressors is still unclear but it is likely that it may interfere with the spreading of histone deactetylation and methylation patterns that are characteristic of a repressive state by recruiting DNA binding proteins such as transcription factors and others. This boundary element was selected over enhancer sequences or two adjacent promoters because it was reported that these combinations in some cases resulted in transcriptional interference with negative effects (Shearwin, Callen et al. 2005).

We named this updated, more efficient, genetic tool pS/MARt. This new DNA vector was not only suitable for generating populations of cells marked with reporter genes but also allowed the generation of genetically modified isogenic cancer lines which required more sophisticated control of transgene expression.

4.5 Generation of isogenic pancreatic cancer cells expressing the reporter gene Luciferase

With the new vector platform, Luciferase labelled pancreatic cancer cells were generated and the reliability of the genetic modified cells was confirmed *in vitro* and *in vivo* (Figure 19). These cancer cell lines showed a proliferation rate that was comparable to the unmodified controls and when used in orthotropic xenograft studies formed tumours that were phenotypically identical to the one formed from the parental control cell lines. As a proof of principle the cell line BxPC-3 was xenografted into nude mice and the cells modified with the episomal vector pS/MARt expressing the reporter gene Luciferase formed a tumour that had the same morphology as the parental BxPC-3 cells. The stable expression of the reporter gene Luciferase in the xenografted cells was evaluated via *in vivo* imaging and immunohistochemistry and we could conclude that the modified cells generated a reliable non-invasive imaging platform that was also used to test a novel pancreatic antic cancer drug (Figure 20 and 21). All the cells in the tumour masses contributed to the transgene expression avoiding the risk of mosaicism induced by the random silencing of foreign genetic material. The cells modified with the S/MAR technology are currently used in the evaluation of new anti-cancer compounds by Pharmatest Service, Turku (Finland).

4.6 SMAD4 and its role in pancreatic cancer development

A significant challenge in cell biology is the restoration of repressive or suppressive genes into cells which can function more effectively in their absence. The restoration of tumour suppressive genes into cancer cells which have lost them as part of their tumorigenic development present such a challenge. In pancreatic cancer one of the fundamental mutations which leads to aggressive tumour formation and metastasis is the loss of the tumour suppressor SMAD4. This gene's normal function

is to induce programmed cell death upon stimulation from TGFβ. SMAD4 protein normally shuttles from the nucleus to the cytoplasm. When a TGFβ molecule binds its receptor it induces the phosphorylation of the cytoplasmic proteins SMAD2 and SMAD3. They then bind SMAD4 forming a trimeric complex that translocates into the nucleus and acts as a transcription factor. This activation induces the transcription of several genes, most of them involved in the cell cycle arrest such as p21. The loss of the gene prevents the formation of the complex and the blocking of this inhibiting pathway leads to uncontrolled cell proliferation.

We initially elected to rescue the functionality of SMAD4 in the pancreatic cancer cell line Capan-1. These cells represent a typical pancreatic cancer line which has characteristic mutations in four key genes: KRAS, p53, CDKN2A and SMAD4. This allowed us to evaluate the impact of one of these important genes by resurrecting its functionality. The idea of challenging pS/MARt in these cells arose because we wanted to test our vector system in cells that were hard to manipulate with standard techniques. Indeed, the DKFZ Genomic and Proteomic Core Facility reported difficulties even in generating stable Luciferase labelled Capan-1 cells with their lentiviral system. This cancer line was reported to be particularly difficult to transduce and even under constant selection pressure the transgene expression was not stable. The viral genome probably underwent chromatin induced silencing which made the modified cells un-suitable for long term studies.

Our primary focus was on the possibility of expressing a fundamental tumour suppressor gene in a representative pancreatic cancer cell line and we were interested in the reliability of the tumour model rather than the functionality of Smad4 in the context of pancreatic cancer development, which is beyond the expertise of the lab. The results clearly demonstrated that when injected orthotopically in the pancreas of nude mice the cells in which the functionality of Smad4 was restored formed undifferentiated cellular masses that had a completely different morphology to those tumours formed from parental Capan-1 and Capan-1 Luciferase expressing cells. The introduction of the reporter gene Luciferase didn't change the cells behaviour, in fact they retained

their ability to form tumours characterized by well differentiated ductal structures and metastasis (Figure 26). When the SMAD4 tumour suppressor gene was restored the masses looked less differentiated (with a more mesenchymal-like phenotype) with a higher recruitment of stroma. The cells also lost their metastatic potential and the "primary" cell masses appeared dormant but they were still able to proliferate (Figure 27). The rescue of Smad4 in these cells, caused slower tumour development but not a complete cell cycle arrest. The mutations in the other three onco- and tumour suppressor genes in Capan-1 cells probably prevent the induction of the apoptosis. We also believe that the culturing conditions required for these cells includes Serum that contains high quantities of TGFβ may have induced secondary aberration that allowed the cells to adapt their growth upon Smad4 restoration.

The DNA vector showed high efficacy in generating stably expressing cancer cell lines and this opened up other possibilities that we wanted to explore that were not possible with other vectors (or previous versions of this vector platform). The vector stability and as a consequence the low efficiency in generating stably expressing cells limited the widely application of this non-integrative vector platform in the past years. The long and tedious selection procedure together with cell populations in which the transgene expression was not homogenous limited the application of the previous version of the S/MAR vector in more intriguing cells.

4.7 The Genetic Engineering of Primary Cells

One application which was particularly appealing was the generation of isogenic/transgenic primary cells. We knew from communication with collaborators and previous work in our lab that the pEPI plasmid had been tested several times for generating genetically engineered primary cells such as stem cells, but without success. In order to evaluate our vector's function in this application we demonstrated that we could immortalize mouse lung fibroblasts with pS/MARt expressing the

transforming factor T-antigen (Figure 22). We then reprogrammed these fibroblast cells into IPSCs to understand better whether the pS/MARt vector could sustain expression through the reprogramming process. Without applying selection, we were able to immortalize and expand primary fibroblasts and the introduction of the reporter gene GFP in the mammalian expression cassette allowed us to follow the reporter gene expression over time and demonstrate the vector functionality throughout the whole process. For the first time we were able to demonstrate that a non-integrative DNA vector could provide sustained transgene expression during reprogramming into IPSCs and through differentiation. The potential of persistently expressing transgenes in stem-cells opened up a brand new field of research for the lab with the potential of using our new vector platform for stem cell therapy. Stem Cells modified with this vector technology can be used for the treatment of several genetic diseases such as Severe Combined Immunodeficiency X-linked (SCID) where patient bone marrow stem cells could be corrected *ex vivo* and re-injected with the aim of fixing the immune system. Patients affected by SCID do not produce functional lymphocytes and therefore they are more sensitive to infectious agents and tumour development. A safe, non-integrative system that can repair the mutations and produce "cured" CD34+ cells would undoubtedly provide a novel therapy for the treatment of this disease.

The successful test in mouse primary fibroblasts encouraged us to apply the pS/MARt technology in more difficult and intriguing primary cells. Since the establishment of cells in culture was significantly more efficient, compared to the previous version of the vector we decided to challenge the technology and its capability of generating transgenic mice by direct injection. pS/MARt expressing GFP was administrated by pro-nuclear injection to 1-cell stage zygotes (Figure 23) and for the first time, by monitoring the embryos *in vitro,* we were able to demonstrate that the vector is capable of sustaining the expression of a transgene, at least in the first crucial stages of the embryonic development. Some of the zygotes were also implanted into pseudo-pregnant mice and the lab is currently undertaking the analysis of the new-born transgenic mice in order to determine whether

this new DNA vector is able to sustain its transgene expression throughout the whole gestation period and if it is maintained episomally. If GFP expression is detected, the possibility of germ line transition will be also tested in order to determine if this new vector technology can be used to generate transgenic mice.

4.8 Genetic Correction of Alkaptonuria

One obvious application of this vector system is as a tool for gene therapy. Our lab has been involved in several preclinical studies investigating the application of DNA vectors for the treatment of genetic liver disease. We have previously shown that prototypes of this vector system could provide sustained life-long transgene expression in the livers of mice from a single administration (Argyros, Wong et al. 2011, Wong, Argyros et al. 2011). We have also shown the genetic correction of the liver disease Phenylketonuria (PKU) by a single administration of a minicircle DNA vector (Viecelli, Harbottle et al. 2014). In this study we showed that the application of a plasmid DNA had no effect on the metabolic defect in mice and the vector was rapidly silenced. In our study, in collaboration with Prof Jonathan Jarvis in Liverpool we have investigated the possibility of using pS/MARt as a genetic therapeutic treatment for the metabolic disorder Alkaptonuria (AKU). We demonstrated for the first time that the Liver specific pS/MARt-P3-HGD (Figure 24) provided sustained and corrective expression of the HGD gene following hydrodynamic delivery to the livers of a mouse model of AKU mice (Figure 24). In our experience such a result has not been previously possible using bacterial plasmid vectors and has only been accomplished using minicircles. This is most likely because the big bacterial backbone is responsible for the spread of inactive chromatin and contribute to vector silencing. The minimalisation of the CpG islands together with the introduction of the genomic insulator overcame these problems demonstrating that our DNA Vector can also sustain the long term expression of corrective genes when delivered

in vivo. Hydrodynamic delivery is, of course not realistically applicable to the treatment of human disease. But, there are several groups that are working on the development of complexes such as nano-particles that might help the delivery of DNA vectors directly in humans for the treatment of disease.

4.9 From pS/MARt to Nano-S/MARt - a novel vector without bacterial backbone.

Although preliminary tests on primary cells with pS/MARt were encouraging, some cells such as primary human cells are often typically refractory to transfection with traditional bacterial plasmids. The disadvantages of vectors based on bacterial sequences were already discussed. Therefore, to improve the capability and application of our vector system that could potentially be clinically relevant we decided to generate a range of vectors based on pS/MARt in which the toxic extraneous bacterial sequences were reduced to a minimum. In all previous reports it was shown how minicircles outperformed their corresponding plasmids but their application and development is currently restricted. However, a collaboration with the American company Nature technology allowed us to swiftly generate a new class of vectors that are based on the pS/MARt platform but lack a bacterial backbone. We have called this new range of constructs Nano-S/MARt Vectors. When compared to minicircle technology, the manufacturing of Nano-S/MAR vectors is simpler, cheaper and quicker. They are prepared via normal DNA preparation in bacteria that generates a high yield. In comparison, the purification of minicircles from their producer plasmid vectors is time consuming and inefficient. In order to generate minicircles an intramolecular recombination of the "producer" plasmid is induced and the desired minicircle vector has to be isolated and purified using specifically designed and proprietary columns. Minicircle production also requires an additional purification step to remove concatemers which are generated during the intracellular

recombination event. In the RNA-OUT (Figure 33) system from NTX there is no need for intramolecular recombination and the exclusion of the bacterial backbone. With this system is possible to generate large quantities of pure supercoiled minimally sized DNA nano-vector without the need of additional steps of purification. This new generation of nano vectors performs better in any application (Figure 17-18-35) than their respective traditional plasmids confirming once again how the removal of the bacterial backbone significantly improves the plasmid efficacy. The cells looked to be less damaged in the days that followed the DNA delivery and the establishment efficiency was improved perhaps as a direct consequence of the delivery of more DNA molecules per cell. The Nano-Vectors are ~2kb (Figure 34) smaller than the corresponding plasmid and since the vector establishment is considered a stochastic process, the delivery of more DNA molecule per cell increases the chance for the vector to reach the appropriate nuclear compartment.

4.10 Nano Vectors for the Genetic Modification of Patient Derived Pancreatic Cancer Cells

We decided to further develop the Isogenic cell work and utilised these Nano vectors to genetically restore human, patient derived pancreatic cancer cells with SMAD4. These cells have been intensively studied and characterised by our collaborators in HiStem (Noll, Eisen et al. 2016) who demonstrated that they are particularly difficult to modify with other vector systems (personal communication). We showed that the nano-vector could successfully sustain the expression of the reporter gene GFP and functionally restore SMAD4 (Figure 36). The cells were tested *in vitro* for the transgenes expression and they were sub sequentially injected orthotopically into nude mice. The restoration of functional Smad4 in Paco2 cells induced the develop of masses that appeared less differentiated and were less aggressive when compared to those formed with unmodified and with GFP expressing cells. The immunohistochemical (Figure 41) analysis showed that the genetic

modification of these cells with a Nano-vector driving the expression of the reporter gene GFP had no impact on the behaviour of the cells. They formed tumours that displayed the same aggressive phenotype characterised by ductal structures and active proliferation as demonstrated with the Ki67 staining. The restoration of Smad4, instead induced profound changes. The injected cells did not develop tumours but formed only inert aggregates of cells that were not actively replicating (as demonstrated by the negative staining for Ki67). The rescue of Smad4 in these cells was able to induce cell cycle arrest but it failed to trigger the apoptotic pathway. This was also expected since these tumour cells spontaneously became immortalised and resistant to the cell death program during their isolation from primary human tumours.

The impact of the Nano vectors on cells was investigated at the molecular level by measuring their genome wide RNA expression levels. The cells modified with the GFP expressing Nano vector surprisingly showed that even though they underwent the selection process and they were growth for months in presence of the episomal vectors, only 5 genetic changes (Figure 39) in their expression profile. We showed that in Capan-1 cells modified with pS/MARt (that have a bacterial backbone) around 300 genes (Figure 29) were either up-or down modulated and we believed that such a dramatic change in these expression profiles is due to the absence of bacterial sequences in the Nano vector.

4.11 Nano Vector Splice - a DNA Vector that mimics human genes

Ultimately, we have completely updated the pEPI vector system to generate the pS/MARt and Nano-SMART vectors and we have demonstrated that they perform better in every measure than the original (Figure 13-17 and 18). We first reduced the bacterial backbone removing the f1 Ori from pEPI which was initially described as a fundamental feature for episomal maintenance. We

boosted the activity of the mammalian expression cassette by insulating it from the bacterial backbone and by linking the expression of a selection marker directly to it. These changes resulted in the efficient pS/MARt vector that was shown to be capable of modifying cells more quickly and efficiently. The vector was then further updated by removing its bacterial backbone and swapping it for the minimal selection system based on RNA interference developed by Nature Technology. When, the updated vector was compared to pS/MARt it was shown that the new minimally sized version of the vector outperformed the original in any application.

One aspect of DNA vector design which potentially has an impact on its functionality that had not previously been considered is the structure and sequence of its mRNA. We wanted to generate an expression system that has and intron-exon structure that is as similar as possible as those typical in eukaryotic cells. We decided to mimic such and expression cassette in our Nano (Figure 42). We felt that the DNA structure of our plasmid was as stable as possible and this was confirmed by Southern Blots therefore we worked in generating a vector system in which the RNA message was also improved. It was reported (Hicks, Yang et al. 2006) that pre-mRNA that do not undergo splicing have a shorter life time than those that undergo this process that leaves the nucleic acid covered and protected from so called splicing proteins. *Broll et al.* (Broll, Oumard et al. 2010) reported that when the β-interferon MAR was included in the transcripts it was less stable and that lead to weaker transgene expression. However, they didn't change the structure of the expression cassette probably because, active transcription is necessary in the MAR sequence to make it functional. To overcome this problem, we decided to add splicing sites before and after the MAR generating what we called Nano-S/MAR-splice vector.

The establishment efficiency of this new vector as well as the expression profile of the population were significantly better when compared to the un-spliced control (Figure 43). We believe that the maturation process improved the RNA stability and it doesn't create uncoated RNA molecules that could be the recognized as a foreign entity such as RNA viruses by intra-cellular sensors.

4.12 A novel, more compact and efficient MAR sequence

Following the development of our new vector system only one of the original features of the original pEPI vector remained unchanged and that was the almost sacrosanct "un-touchable" component of the vector the S/MAR motif from the human β-interferon gene which provides the vector with its most unique functionality of an episomally maintained DNA vector. We decided that this element could also be improved so we exchanged it.

Several MAR elements were evaluated and it was noticed that the primary limitation of the original S/MAR was the lack of a well characterized structure. We thought that a DNA sequence that was composed of repetitive modules could form much better binding sites for nuclear proteins that may facilitate the episomal anchoring to the nuclear matrix. The proteins that mediate the anchoring of our episome to the nuclear scaffold are still unclear but we thought that a more organized sequence would improve the recognition and the tethering of the DNA. Analysis of the core MAR sequences from several different organisms and species showed that in what was considered a "real" DNA origin of replication and also "putative" ones the recurrent motif ATTA or ATTTA was enriched. We then assessed the human genome looking for a region where these motifs were highly represented. We found that the Apolipoprotein B gene presents a MAR exclusively constituted by the repetition of two short domains. The original pEPI MAR was replaced by this more compact and potentially more efficient one. The newly designed vector was tested in comparison with all previous versions and we found that it established cells with the same efficiency as the Nano Vectors, despite having a bacterial backbone. The transgene expression within populations established with this version of the plasmid is however weaker.

4.13 The Genetic Engineering of Human T-Cells

As an ultimate test of this vector we applied it to genetically engineer primary human T-cells. These cells as part of the immune system are designed to react against non-self entities and we expected that they would be particularly sensitive in the recognition of foreign DNA. We showed that pS/MARt could be successfully applied in several applications but when challenged with the persistent modification of T-Cells it proved to be incapable of mediating their modification. Those cells represent arguably the most challenging application for our vector system and although pS/MARt was not incapable of modifying these cells we managed to overcome the problem by generating and applying the Nano-vectors and the pS/MARter with the novel MAR motif. We showed that pS/MARter could sustain the expression of the GFP transgene for over a month (Figure 48) at sustained high levels without undergoing the decline that pEPI or the same plasmid without the MAR showed. However, when we tried to modify these cells with a transgenic TCR, we saw that even pS/MARter was not able to sustain the transgene expression beyond 6 days of delivery. We are currently performing several analyses in order to determine whether the loss of the expression is due to the silencing of the vector or to the post-translational down-modulation of the TCR. To be expressed on the T Cells surface the α and β chains of a TCR need to be complexed with CD3 adaptor molecules. The lack of these adaptors can potentially block the translocation of the complex to the membrane and induce degradation of the protein. Currently we are investigating via intracellular staining of the TCR whether the receptors are expressed but not translocated into the membrane or if the transfected cells die as a consequence of the over-expression.

The possibility of applying our episomal vectors for autologous therapy where T Cells are modified and reprogrammed to target tumor cells represents undoubtedly one of the most interesting and

intriguing fields of gene therapy research and we are continuing to improve our understanding of how we can generate a novel, alternative vector system suitable for this application.

5 Conclusions

In conclusion, the performance of our ultimate vector pS/MARter in the establishment of cells was better than most other vectors generated during this process. The number of colonies obtained in the colony forming assay test was comparable to the Nano Vectors despite that vector carrying the new MAR motif is still in a version with the bacterial backbone. The Nano-S/MAR-splice vector remains, however, the most efficient construct in all the aspects tested. We believe that the generation of pS/MARter, the ultimate vector produced in this study, takes the development of plasmid vectors with bacterial backbones as far as possible with our current understanding of vector genetics. We believe that an updated Nano version of this construct that comprises the novel splicing feature will improve its efficacy and application even further.

Table 17. SMAD4 dependent up regulated genes with at least 2 fold increase.

ADAM19	ME1
CDH3	LRRC33
CFP	MGC102966
CHCHD2	MICB
CHCHD9	MYEOV
CLEC11A	PPP2R2C
COL16A1	PRSS1
COL17A1	MMP1
COL18A1	RRAD
CSK	SERPINE2
CTHRC1	MMP3
CXCR7	MMP9
DPYSL3	SLC6A10P
DSE	SOSTDC1
EGFR	MTSS1
FEZ1	MYL9
FGFBP1	MYO1B
GJB2	NUAK1
GLIPR1	OLR1
GLS	SPPL2A
GPX3	PANX2
HBEGF	PI3
HCP5	ACTN1
HOXA10	AIF1
HS.554203	PLAU
HS.567963	PPP2R2B
ICAM2	ANTXR1
IGSF3	APOD
IL1RL1	PRNP
IL20RB	PRSS2
IRX5	BASP1
KRT14	PTGS1
KRT6B	RAB37
KRT81	RPTN
LAMA3	C12ORF54
LAMC2	C18ORF45
LGALS7B	C6ORF15
LOC100131139	C6ORF85
LOC375295	CALD1
LOC401817	S100A2
LOC650200	SAA1
LOC651397	SAA2
LOC652002	SAA4
LOC728324	CASP1
LOC728910	CCL28
MAOA	CD276

SCGB3A1	THBS4
SERPINB3	TIMP2
SERPINE1	TNC
SFN	TNFSF10
SNAI2	TPST1
TAGLN	TUBB2A
THBS1	

Table 18. SMAD4 dependent down regulated genes with at least 2 fold decrease

ACSL5	FAM149A
BMP4	FOS
BTBD16	GPD1L
C10ORF81	IGFBP3
C1ORF192	IL17RB
C1ORF194	IQGAP2
C3ORF15	KIAA1147
C5ORF41	LOC391019
C7ORF57	PAQR8
C7ORF63	PGC
CASC1	PLTP
CEACAM1	PPP1R15A
CTXN1	PPP1R16A
CXORF57	RNASE1
DEFB1	RPL14
DGAT1	RPL15
DHRS3	SCG5
ERBB3	SLC38A5
FOXJ1	SLC6A20
GABRP	STEAP1
GSTK1	STXBP6
IGFBP4	SUCLG2
KLK11	TM4SF4
KRT23	TNFRSF19
LMTK3	TOP1MT
LOC649821	TOP2B
LOC728820	TRAK1
LY6E	ZNF256

Table 19. Genes at least 2 folds up-regulated in pancreatic tumors formed from the Capan-1 Luciferase labelled cells upon orthotopic injection compared to tumors generated with unmodified Capan-1 cells.

ALCAM	COL4A5
ALDH1A3	CST6
ALDH3B2	CTSD
ALOX5	CXCL6
ANKRD33	DHRS9
ANXA1	DMBT1
ANXA8	DUSP6
ANXA8L2	EDARADD
AQP3	EFEMP1
BLCAP	EGLN3
BST2	ELF5
C14ORF4	F3
C20ORF114	FAM83A
C9ORF169	FCGBP
CAV1	FGFR3
CD14	FNBP1
CD82	GABRP
CDA	GATS
CFB	GNA15
CFH	GPR116
CFI	HEG1
CLIC3	HS.407903
HSPBL2	LOC100133817
IFI16	LOC283392
IFI27	LOC644760
IFI44	LOC645553
IFI44L	LOC645638
IFI6	LOC652846
IFIT1	LOC653879
IGFBP6	LOC728969
INHBB	LRG1
IRF9	LTBP2
IRX3	LY6D
JAG2	LYNX1
KLK6	LYPD2
KLRC2	MAGED1
KRT13	MAMDC2
KRT16	MMP10
KRT5	MMP12
KRT6A	MT1A
KRT6C	MUC16
LEPROT	MUC4
LGALS8	MX1
LOC100129681	NOTCH3

NTN4	SGK1
OAS2	SGSM2
PADI3	SHISA2
PALMD	SIPA1L2
PCDHB2	SLC16A3
PDZK1IP1	SLC1A3
PPP1R3C	SLC4A11
PROM2	STAT2
PROS1	SYTL4
PRSS23	TACSTD2
PTGES	TCN1
PTPN20	TEAD2
PYGL	TMEM173
RARRES1	TMEM45A
RARRES3	TRIM29
RNASET2	TSPAN1
RXRA	TSPAN9
S100A8	VAMP5
S100A9	VGLL1
SCCPDH	VTCN1
SFTA1P	WFDC2
SGK	

REFERENCES

Argyros, O., S. P. Wong, C. Fedonidis, O. Tolmachov, S. N. Waddington, S. J. Howe, M. Niceta, C. Coutelle and R. P. Harbottle (2011). "Development of S/MAR minicircles for enhanced and persistent transgene expression in the mouse liver." J Mol Med (Berl) **89**(5): 515-529.

Argyros, O., S. P. Wong, M. Niceta, S. N. Waddington, S. J. Howe, C. Coutelle, A. D. Miller and R. P. Harbottle (2008). "Persistent episomal transgene expression in liver following delivery of a scaffold/matrix attachment region containing non-viral vector." Gene Ther **15**(24): 1593-1605.

Atianand, M. K. and K. A. Fitzgerald (2013). "Molecular basis of DNA recognition in the immune system." J Immunol **190**(5): 1911-1918.

Baiker, A., C. Maercker, C. Piechaczek, S. B. Schmidt, J. Bode, C. Benham and H. J. Lipps (2000). "Mitotic stability of an episomal vector containing a human scaffold/matrix-attached region is provided by association with nuclear matrix." Nat Cell Biol **2**(3): 182-184.

Bigger, B. W., O. Tolmachov, J. M. Collombet, M. Fragkos, I. Palaszewski and C. Coutelle (2001). "An araC-controlled bacterial cre expression system to produce DNA minicircle vectors for nuclear and mitochondrial gene therapy." J Biol Chem **276**(25): 23018-23027.

Bode, J., T. Schlake, M. Rios-Ramirez, C. Mielke, M. Stengert, V. Kay and D. Klehr-Wirth (1995). "Scaffold/matrix-attached regions: structural properties creating transcriptionally active loci." Int Rev Cytol **162A**: 389-454.

Boulikas, T. (1992). "Homeotic protein binding sites, origins of replication, and nuclear matrix anchorage sites share the ATTA and ATTTA motifs." J Cell Biochem **50**(2): 111-123.

Broll, S., A. Oumard, K. Hahn, A. Schambach and J. Bode (2010). "Minicircle performance depending on S/MAR-nuclear matrix interactions." J Mol Biol **395**(5): 950-965.

Brooks, A. R., R. N. Harkins, P. Wang, H. S. Qian, P. Liu and G. M. Rubanyi (2004). "Transcriptional silencing is associated with extensive methylation of the CMV promoter following adenoviral gene delivery to muscle." J Gene Med **6**(4): 395-404.

Chen, C., X. D. Fang, J. Zhu, X. F. Wu, Z. C. Zhang, J. X. Gu, Z. Y. Wang and C. W. Chi (1999). "The gene expression of coagulation factor VIII in mammalian cell lines." Thromb Res **95**(2): 105-115.

Chen, Z. Y., C. Y. He, A. Ehrhardt and M. A. Kay (2003). "Minicircle DNA vectors devoid of bacterial DNA result in persistent and high-level transgene expression in vivo." Mol Ther **8**(3): 495-500.

Cockerill, P. N. and W. T. Garrard (1986). "Chromosomal loop anchorage sites appear to be evolutionarily conserved." FEBS Lett **204**(1): 5-7.

Cong, L., F. A. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, P. D. Hsu, X. Wu, W. Jiang, L. A. Marraffini and F. Zhang (2013). "Multiplex genome engineering using CRISPR/Cas systems." Science **339**(6121): 819-823.

David, C. J., Y. H. Huang, M. Chen, J. Su, Y. Zou, N. Bardeesy, C. A. Iacobuzio-Donahue and J. Massague (2016). "TGF-beta Tumor Suppression through a Lethal EMT." Cell **164**(5): 1015-1030.

Deer, E. L., J. Gonzalez-Hernandez, J. D. Coursen, J. E. Shea, J. Ngatia, C. L. Scaife, M. A. Firpo and S. J. Mulvihill (2010). "Phenotype and genotype of pancreatic cancer cell lines." Pancreas **39**(4): 425-435.

Deniger, D. C., J. Yu, M. H. Huls, M. J. Figliola, T. Mi, S. N. Maiti, G. F. Widhopf, 2nd, L. V. Hurton, R. Thokala, H. Singh, S. Olivares, R. E. Champlin, W. G. Wierda, T. J. Kipps and L. J. Cooper (2015). "Sleeping Beauty Transposition of Chimeric Antigen Receptors Targeting Receptor Tyrosine Kinase-Like Orphan Receptor-1 (ROR1) into Diverse Memory T-Cell Populations." PLoS One **10**(6): e0128151.

Gasser, S. M. and U. K. Laemmli (1986). "Cohabitation of scaffold binding regions with upstream/enhancer elements of three developmentally regulated genes of D. melanogaster." Cell **46**(4): 521-530.

Gregory, S. M., S. A. Nazir and J. P. Metcalf (2011). "Implications of the innate immune response to adenovirus and adenoviral vectors." Future Virol **6**(3): 357-374.

Haan, K. M., A. Aiyar and R. Longnecker (2001). "Establishment of latent Epstein-Barr virus infection and stable episomal maintenance in murine B-cell lines." J Virol **75**(6): 3016-3020.

Haase, R., T. Magnusson, B. Su, F. Kopp, E. Wagner, H. Lipps, A. Baiker and M. Ogris (2013). "Generation of a tumor- and tissue-specific episomal non-viral vector system." BMC Biotechnol **13**: 49.

Hacein-Bey-Abina, S., C. Von Kalle, M. Schmidt, M. P. McCormack, N. Wulffraat, P. Leboulch, A. Lim, C. S. Osborne, R. Pawliuk, E. Morillon, R. Sorensen, A. Forster, P. Fraser, J. I. Cohen, G. de Saint Basile, I. Alexander, U. Wintergerst, T. Frebourg, A. Aurias, D. Stoppa-Lyonnet, S. Romana, I. Radford-Weiss, F. Gross, F. Valensi, E. Delabesse, E. Macintyre, F. Sigaux, J. Soulier, L. E. Leiva, M. Wissler, C. Prinz, T. H. Rabbitts, F. Le Deist, A. Fischer and M. Cavazzana-Calvo (2003). "LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1." Science **302**(5644): 415-419.

Hackett, P. B., Jr., E. L. Aronovich, D. Hunter, M. Urness, J. B. Bell, S. J. Kass, L. J. Cooper and S. McIvor (2011). "Efficacy and safety of Sleeping Beauty transposon-mediated gene transfer in preclinical animal studies." Curr Gene Ther **11**(5): 341-349.

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

Hagedorn, C., M. N. Antoniou and H. J. Lipps (2013). "Genomic cis-acting Sequences Improve Expression and Establishment of a Nonviral Vector." Mol Ther Nucleic Acids **2**: e118.

Hagedorn, C., A. Gogol-Doring, S. Schreiber, J. T. Epplen and H. J. Lipps (2017). "Genome-wide profiling of S/MAR-based replicon contact sites." Nucleic Acids Res.

Hagedorn, C., H. J. Lipps and S. Rupprecht (2010). "The epigenetic regulation of autonomous replicons." Biomol Concepts **1**(1): 17-30.

Hayes, F. (2003). "The function and organization of plasmids." Methods Mol Biol **235**: 1-17.

Hemmi, H., O. Takeuchi, T. Kawai, T. Kaisho, S. Sato, H. Sanjo, M. Matsumoto, K. Hoshino, H. Wagner, K. Takeda and S. Akira (2000). "A Toll-like receptor recognizes bacterial DNA." Nature **408**(6813): 740-745.

Henderson, S., M. Rowe, C. Gregory, D. Croom-Carter, F. Wang, R. Longnecker, E. Kieff and A. Rickinson (1991). "Induction of bcl-2 expression by Epstein-Barr virus latent membrane protein 1 protects infected B cells from programmed cell death." Cell **65**(7): 1107-1115.

Hicks, M. J., C. R. Yang, M. V. Kotlajich and K. J. Hertel (2006). "Linking splicing to Pol II transcription stabilizes pre-mRNAs and influences splicing patterns." PLoS Biol **4**(6): e147.

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

Hsieh, W. S., M. V. Lemas and R. F. Ambinder (1999). "The biology of Epstein-Barr virus in post-transplant lymphoproliferative disease." Transpl Infect Dis **1**(3): 204-212.

Jenke, A. C., I. M. Stehle, F. Herrmann, T. Eisenberger, A. Baiker, J. Bode, F. O. Fackelmayer and H. J. Lipps (2004). "Nuclear scaffold/matrix attached region modules linked to a transcription unit are sufficient for replication and maintenance of a mammalian episome." Proc Natl Acad Sci U S A **101**(31): 11322-11327.

Jenke, B. H., C. P. Fetzer, I. M. Stehle, F. Jonsson, F. O. Fackelmayer, H. Conradt, J. Bode and H. J. Lipps (2002). "An episomally replicating vector binds to the nuclear matrix protein SAF-A in vivo." EMBO Rep **3**(4): 349-354.

Kalos, M. and R. E. Fournier (1995). "Position-independent transgene expression mediated by boundary elements from the apolipoprotein B chromatin domain." Mol Cell Biol **15**(1): 198-207.

Kawai, T. and S. Akira (2007). "Signaling to NF-kappaB by Toll-like receptors." Trends Mol Med **13**(11): 460-469.

Kershaw, M. H., M. W. Teng, M. J. Smyth and P. K. Darcy (2005). "Supernatural T cells: genetic modification of T cells for cancer therapy." Nat Rev Immunol **5**(12): 928-940.

Kipp, M., F. Gohring, T. Ostendorp, C. M. van Drunen, R. van Driel, M. Przybylski and F. O. Fackelmayer (2000). "SAF-Box, a conserved protein domain that specifically recognizes scaffold attachment region DNA." Mol Cell Biol **20**(20): 7480-7489.

Kozak, M. (1987). "An analysis of 5'-noncoding sequences from 699 vertebrate messenger RNAs." Nucleic Acids Res **15**(20): 8125-8148.

Kwaks, T. H., P. Barnett, W. Hemrika, T. Siersma, R. G. Sewalt, D. P. Satijn, J. F. Brons, R. van Blokland, P. Kwakman, A. L. Kruckeberg, A. Kelder and A. P. Otte (2003). "Identification of anti-repressor elements that confer high and stable protein production in mammalian cells." Nat Biotechnol **21**(5): 553-558.

Lederberg, J. (1952). "Cell genetics and hereditary symbiosis." Physiol Rev **32**(4): 403-430.

Li, S., S. D. Hursting, B. J. Davis, J. A. McLachlan and J. C. Barrett (2003). "Environmental exposure, DNA methylation, and gene regulation: lessons from diethylstilbesterol-induced cancers." Ann N Y Acad Sci **983**: 161-169.

Liu, W., B. Liang, H. Liu, Y. Huang, X. Yin, F. Zhou, X. Yu, Q. Feng, E. Li, Z. Zou and L. Wu (2017). "Overexpression of nonSMC condensin I complex subunit G serves as a promising prognostic marker and therapeutic target for hepatocellular carcinoma." Int J Mol Med.

Luke, J. M., A. E. Carnes and J. A. Williams (2014). "Development of antibiotic-free selection system for safer DNA vaccination." Methods Mol Biol **1143**: 91-111.

Lutfalla, G., L. Armbruster, S. Dequin and R. Bertolotti (1989). "Construction of an EBNA-producing line of well-differentiated human hepatoma cells and of appropriate Epstein-Barr virus-based shuttle vectors." Gene **76**(1): 27-39.

Mali, P., L. Yang, K. M. Esvelt, J. Aach, M. Guell, J. E. DiCarlo, J. E. Norville and G. M. Church (2013). "RNA-guided human genome engineering via Cas9." Science **339**(6121): 823-826.

Mansur, D. S., G. L. Smith and B. J. Ferguson (2014). "Intracellular sensing of viral DNA by the innate immune system." Microbes Infect **16**(12): 1002-1012.

Mesner, L. D., J. L. Hamlin and P. A. Dijkwel (2003). "The matrix attachment region in the Chinese hamster dihydrofolate reductase origin of replication may be required for local chromatid separation." Proc Natl Acad Sci U S A **100**(6): 3281-3286.

Mirkovitch, J., M. E. Mirault and U. K. Laemmli (1984). "Organization of the higher-order chromatin loop: specific DNA attachment sites on nuclear scaffold." Cell **39**(1): 223-232.

Noll, E. M., C. Eisen, A. Stenzinger, E. Espinet, A. Muckenhuber, C. Klein, V. Vogel, B. Klaus, W. Nadler, C. Rosli, C. Lutz, M. Kulke, J. Engelhardt, F. M. Zickgraf, O. Espinosa, M. Schlesner, X. Jiang, A. Kopp-Schneider, P. Neuhaus, M. Bahra, B. V. Sinn, R. Eils, N. A. Giese, T. Hackert, O. Strobel, J. Werner, M. W. Buchler, W. Weichert, A. Trumpp and M. R. Sprick (2016). "CYP3A5 mediates basal and acquired therapy resistance in different subtypes of pancreatic ductal adenocarcinoma." Nat Med **22**(3): 278-287.

O'Neill, L. A., D. Golenbock and A. G. Bowie (2013). "The history of Toll-like receptors - redefining innate immunity." Nat Rev Immunol **13**(6): 453-460.

Ohgane, J., S. Yagi and K. Shiota (2008). "Epigenetics: the DNA methylation profile of tissue-dependent and differentially methylated regions in cells." Placenta **29 Suppl A**: S29-35.

Paquet, D., D. Kwart, A. Chen, A. Sproul, S. Jacob, S. Teo, K. M. Olsen, A. Gregg, S. Noggle and M. Tessier-Lavigne (2016). "Efficient introduction of specific homozygous and heterozygous mutations using CRISPR/Cas9." Nature **533**(7601): 125-129.

Paulson, J. R. and U. K. Laemmli (1977). "Morphogenetic core of the bacteriophage T4 head. Structure of the core in polyheads." J Mol Biol **111**(4): 459-485.

Pelletier, J. and N. Sonenberg (1989). "Internal binding of eucaryotic ribosomes on poliovirus RNA: translation in HeLa cell extracts." J Virol **63**(1): 441-444.

Piechaczek, C., C. Fetzer, A. Baiker, J. Bode and H. J. Lipps (1999). "A vector based on the SV40 origin of replication and chromosomal S/MARs replicates episomally in CHO cells." Nucleic Acids Res **27**(2): 426-428.

Plasterk, R. H. (1993). "Molecular mechanisms of transposition and its control." Cell **74**(5): 781-786.

Rupprecht, S., C. Hagedorn, D. Seruggia, T. Magnusson, E. Wagner, M. Ogris and H. J. Lipps (2010). "Controlled removal of a nonviral episomal vector from transfected cells." Gene **466**(1-2): 36-42.

Ryan, M. D., A. M. King and G. P. Thomas (1991). "Cleavage of foot-and-mouth disease virus polyprotein is mediated by residues located within a 19 amino acid sequence." J Gen Virol **72 (Pt 11)**: 2727-2732.

Sarukhan, A., C. Garcia, A. Lanoue and H. von Boehmer (1998). "Allelic inclusion of T cell receptor alpha genes poses an autoimmune hazard due to low-level expression of autospecific receptors." Immunity **8**(5): 563-570.

Sepkowitz, K. A. (2001). "AIDS--the first 20 years." N Engl J Med **344**(23): 1764-1772.

Shearwin, K. E., B. P. Callen and J. B. Egan (2005). "Transcriptional interference--a crash course." Trends Genet **21**(6): 339-345.

Sixbey, J. W. and J. S. Pagano (1985). "Epstein-Barr virus transformation of human B lymphocytes despite inhibition of viral polymerase." J Virol **53**(1): 299-301.

Snudden, D. K., J. Hearing, P. R. Smith, F. A. Grasser and B. E. Griffin (1994). "EBNA-1, the major nuclear antigen of Epstein-Barr virus, resembles 'RGG' RNA binding proteins." EMBO J **13**(20): 4840-4847.

Stinchcomb, D. T., M. Thomas, J. Kelly, E. Selker and R. W. Davis (1980). "Eukaryotic DNA segments capable of autonomous replication in yeast." Proc Natl Acad Sci U S A **77**(8): 4559-4563.

Tao, J., X. Zhou and Z. Jiang (2016). "cGAS-cGAMP-STING: The three musketeers of cytosolic DNA sensing and signaling." IUBMB Life **68**(11): 858-870.

Tsukamoto, H., D. Wells, S. Brown, P. Serpente, P. Strong, J. Drew, K. Inui, S. Okada and G. Dickson (1999). "Enhanced expression of recombinant dystrophin following intramuscular injection of Epstein-Barr virus (EBV)-based mini-chromosome vectors in mdx mice." Gene Ther **6**(7): 1331-1335.

Uematsu, S., S. Sato, M. Yamamoto, T. Hirotani, H. Kato, F. Takeshita, M. Matsuda, C. Coban, K. J. Ishii, T. Kawai, O. Takeuchi and S. Akira (2005). "Interleukin-1 receptor-associated kinase-1 plays an essential role for Toll-like receptor (TLR)7- and TLR9-mediated interferon-{alpha} induction." J Exp Med **201**(6): 915-923.

Vaysse, L., L. G. Gregory, R. P. Harbottle, E. Perouzel, O. Tolmachov and C. Coutelle (2006). "Nuclear-targeted minicircle to enhance gene transfer with non-viral vectors in vitro and in vivo." J Gene Med **8**(6): 754-763.

Verma, I. M. and N. Somia (1997). "Gene therapy -- promises, problems and prospects." Nature **389**(6648): 239-242.

Viecelli, H. M., R. P. Harbottle, S. P. Wong, A. Schlegel, M. K. Chuah, T. VandenDriessche, C. O. Harding and B. Thony (2014). "Treatment of phenylketonuria using minicircle-based naked-DNA gene transfer to murine liver." Hepatology **60**(3): 1035-1043.

Wade-Martins, R., Y. Saeki and E. A. Chiocca (2003). "Infectious delivery of a 135-kb LDLR genomic locus leads to regulated complementation of low-density lipoprotein receptor deficiency in human cells." Mol Ther **7**(5 Pt 1): 604-612.

Wilson, J. B., J. L. Bell and A. J. Levine (1996). "Expression of Epstein-Barr virus nuclear antigen-1 induces B cell neoplasia in transgenic mice." EMBO J **15**(12): 3117-3126.

Wong, S. P., O. Argyros, C. Coutelle and R. P. Harbottle (2011). "Non-viral S/MAR vectors replicate episomally in vivo when provided with a selective advantage." Gene Ther **18**(1): 82-87.