HARVARD STEM CELL INSTITUTE

FH KREMS UNIVERSITY OF APPLIED SCIENCES/AUSTRIA

AUSTRIAN MARSHALL PLAN FOUNDATION

Final report

"Generation of induced pluripotent stem cells from fibroblast cell lines and gene editing in pluripotent stem cell"

By

Konrad Chudzik

Bachelor programme

Medical and Pharmaceutical Biotechnology

External Supervisor: Laurence Daheron, PhD

Internal Supervisor: Dr. Reinhard Klein

Acknowledgments

First of all, I would like to thank Harvard University and Harvard Stem Cell Institute for hosting my internship and providing me with resources necessary for this project. Particularly, Laurence Daheron, my supervisor, who put enormous amount of effort and time to teach me as much as possible during those six months and whose patience, knowledge and experience made my internship an enriching and rewarding experience. Also my coworkers from iPS core facility: Sarah Fahmy, Caroline Becker, Lauren Jasko, Ying Shao, Ping Xu and Xin Jiang, whose advice, help and companionship had contribute to making my internship an amazing experience.

Secondly, I would like to thank the IMC FH Krems, which empowered me to do an internship and supported me throughout the whole process. Particularly, my external supervisor, Reinhard Klein and Barbara Entler, both of whom provided me with a much needed guidance, advice and help during those six months.

Last, but not least, I would like to thank the Marshall Plan Foundation for providing me with a scholarship that helped me fund the internship.

My internship and project would not be possible without the effort, time and resources of the before mentioned institutions and persons, all of which truly deserve the credit for the following work and to whom I owe the experience of a life-time that my internship has been.

Table of contents

Table of figures

List of Abbreviations

hES - human embryonic stem

ES - embryonic stem

Oct ¾ - octamer-binding transcription factor ¾

Sox 2 - sex determining-region Y-box 2

Klf4 - Krueppel-like factor 4

DNA - Deoxyribonucleic acid

RNA - Ribonucleic acid

Gag - group-specific antigen

cDNA - complementary DNA

LTR - long terminal repeat

HEK - Human Embryonic Kidney

iPS – induced pluripotent stem

ICC - Immunocytochemistry

qPCR - Real-time quantitative polymerase chain reaction

PCR - Polymerase chain reaction

EB - embryonic body

FACS - Fluorescence-activated cell sorting

GCM2 - Glial Cells Missing Homolog 2

MGH - Massachusetts General Hospital

eGFP - enhanced green fluorescent protein

GFP - Green fluorescent protein

IRESs - internal ribosomal entry sites

IRES - internal ribosomal entry site

P2A - Peptide 2A

DSBs - double-strand breaks

DSB - double-strand break

indel - insertion or deletion

NHEJ - nonhomologous end-joining

CRISPR - clustered regularly interspaced short palindromic repeats

PAM - protospacer-adjacent motif

TIDE - Tracking of Indels by DEcomposition

MEF - Mouse Embryo Fibroblast (MEF)

DMEM - Dulbecco's Modified Eagle Medium

FBS - Fetal bovine serum

EDTA - Ethylenediaminetetraacetic acid

VPA - Valproic Acid

FGF - Fibroblast growth factor

bFGF - Basic fibroblast growth factor

MOI - Multiplicity of infection

CIU - Viral Coefficient

PBS - Phosphate-Buffered Saline

DAPI - 4',6-diamidino-2-phenylindole

IgM - Immunoglobulin M

IgG - Immunoglobulin G

sgRNA – single-stranded guide RNA

SAM - synergistic activation mediator

1 Introduction

1.1 Generation of induced pluripotent stem cells from lymphoblastoid cell lines

1.1.1. Cell culture

Cell culture, as an instrument of research, has been a source of major break-through for many fields, particularly those investigating human development and disease pathogenesis. The possibility of studying human cells in vitro offers unique insight for biomedical research. However, primary human cells have limited proliferation capabilities in vitro. This puts severe limitations on studying regeneration or formation of tissue. Cells with unlimited proliferation could be derived from malignant tissues or from primary somatic cells by genetic modification, but these cells do not have in vivo counterpart and therefore are not optimal tools for research. [1,3]

1.1.2. Human embryonic stem cells

The derivation of human embryonic stem (hES) cells in 1998 has been a major breakthrough in that regard. These cells are derived from the inner cell mass of blastocyst. Human blastocysts can be obtained for research purposes from excess embryos from in vitro fertilization clinics. Those cells opened a completely new array of possibilities, compared to the use of primary human cells since they can proliferate indefinitely and can differentiate into any cell type of the body. Therefore, they could provide an unlimited source of cells to facilitate the development of new treatment for diseases such as Parkinson's disease or diabetes and the exploration of human development. Still there are two major issues related to the derivation of hES cells: first of all they are derived from human blastocysts, which is obviously not an abundant resource and puts a limit on the number of available cells; secondly, there are ethical concerns related to the derivation procedure since the embryos is destroyed during the process. These ethical issues have led some countries to introduce laws restraining human embryos' access; the most prominent example would be George H. Bush administration in the US. [1-3]

1.1.3. Induced pluripotent stem cells

Another important breakthrough for the stem cell field and use of cell culture has come with Shinya Yamanaka's discovery in 2006. He showed that pluripotent stem cells, capable of immortal growth and differentiation into all three germ layers, can be derived directly from fibroblasts by overexpressing four transcriptional factors, Oct4, Klf4, Sox2 and cMyc, often referred to as "Yamanaka factors" as a tribute to their discoverer. Since then, this reprogramming method has been tested in many laboratories worldwide and proven to be a robust and reliable method for deriving induced pluripotent stem cells from human somatic cells. Those induced pluripotent stem (iPS) cells exhibit many similarities with embryonic stem (ES) cells, such as morphology, expression of markers and function. This means that iPS derivation does not only provide a unique opportunity to obtain virtually unlimited human cell in culture without the ethical and technical difficulties related to the ES cells, but also allows the derivation of patient-specific cells a unique and valuable tool to get insight into specific diseases and a potential source of autologous cells for transplant . [1,3]

1.1.4. Reprogramming transcriptional factors

The reprogramming of somatic cells using the four aforementioned transcription factors was discovered together by Yamanaka and Takahashi in 2006 and Yamanaka received a Nobel prize in the field of physiology or medicine in 2012 for this discovery. Takahashi and Yamanaka selected twenty four genes that were known to be crucial in establishing and maintaining the unique identity of

embryonic stem cells. Then they tested the impact of overexpression of those factors by retroviral transduction into mouse embryonic fibroblast in different combinations overtime. In the end, they reached the conclusion that the cocktail of four factors: octamer-binding transcription factor ¾ (Oct ¾), sex determining-region Y-box 2 (Sox 2), Krueppel-like factor 4 (Klf4) and c-Myc, was sufficient to reprogram adult somatic cells into a pluripotent state. [3]

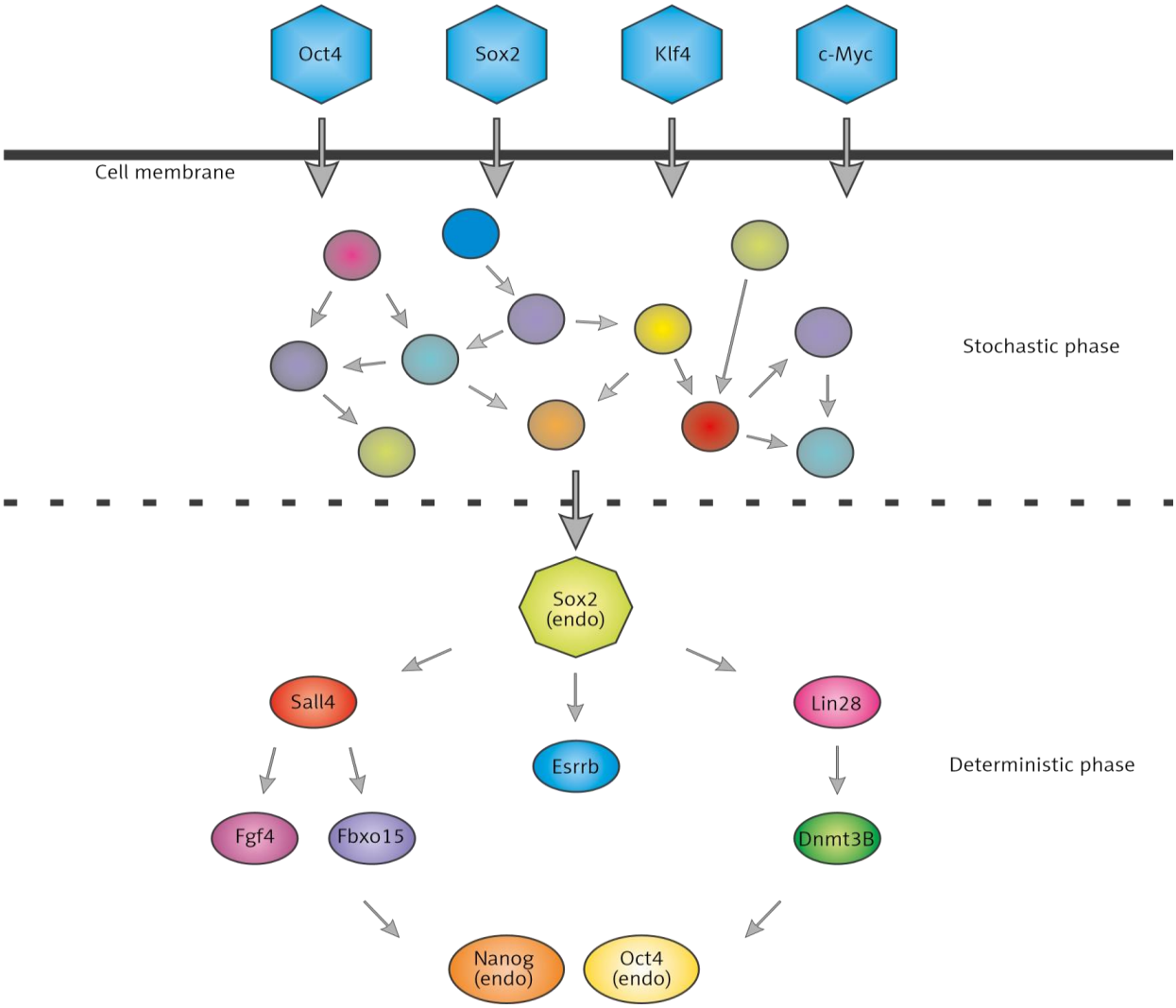


FIGURE 1 EARLY STOCHASTIC AND LATE HIERARCHIAL PHASES OF REPROGRAMMING

The initial research also indicated that the factors: Oct ¾ and Sox 2 are responsible for the generation of mature iPS cells, while the other two factors: Klf-4 and c-Myc play important role in accelerating the process of reprogramming and that they significantly increase the efficiency of the whole process [3]. Despite the fact that the reprogramming method discovered by Yamanaka has been widely used in many laboratories around the world, still there is only limited information about the molecular actions of the factors [4].

Buganim et al. [5] showed that the process of reprogramming has a stochastic and a hierarchical phase. Interestingly, a cascade, triggered by the reprogramming factors, leads to the creation of pre-iPSc from which only a small percentage will turn into fully reprogrammed stem cells. Here is a summary of what is known about the role of Yamanaka's factors::

C-Myc plays a role during the initial phase of reprogramming by activating the pluripotency markers. It is responsible for activating genes involved in cell transformation and increase in proliferation. It facilitates exogenous Oct ¾ and Sox 2 binding to DNA, by combining with histone acetyltransferase complex and inducing global histone acetylation. [4]

Klf4 plays a part in the initial phase of the reprogramming, as well as in the later, deterministic phase, thus playing a dual role. In the first phase, it inhibits the expression of certain genes, while later on, it induces expression of some of the pluripotent genes, such as NANOG, by repression of p53 protein. [4]

Sox 2 is a factor that plays an important role during all stages of reprogramming process. It is believed that exogenous Sox2 is active during the stochastic stage, while the activation of endogenous Sox2 triggers the hierarchical phase. This transcription factor plays a role in variety of ways: it has an impact on the expression of fibroblast growth factor 4 or DNA cytosine 5-methyltransferase 3ß amongst other. It is interacting with a Wdr5 chromatin modifier. It also works together with other pluripotency factors, e.g. NANOG, by binding to the same DNA sequence, further stimulating its expression. [4]

Oct 3/4 is the most crucial reprogramming factors, found in all the different reported reprogramming cocktails. It plays a major role in early development, as it impacts gene expression, by reorganizing the chromatin: it binds to closed chromatins, paving the way for other transcription factors, and bringing chromatin regulatory complexes to regions in which they are able to act. [4]

1.1.5. Reprogramming factors delivery

Knowing that the cocktail of these four transcriptional factors can be used to fully reprogram adult somatic cells into a pluripotent state, the next important point is the delivery of these factors to the cell. One of the possible delivery methods is to use the machinery of a viron particle to integrate sequences of transcriptional factors of interest into the host genome. I used a retrovirus and sendai virus viron particles in my work.

1.1.5.1. Retrovirus

Retroviruses belong to the family of Retroviridae, which contains enveloped viruses with RNA genome. They replicate through a reverse transcription once inside the host cell - which is a unique ability as most viruses operate by standard transcription of DNA into RNA, followed by translation into proteins. There are morphological and biological differences between specific retroviruses virons, but there are also significant similarities [6]. Virons have a diameter of ~ 100 nm and their components are:

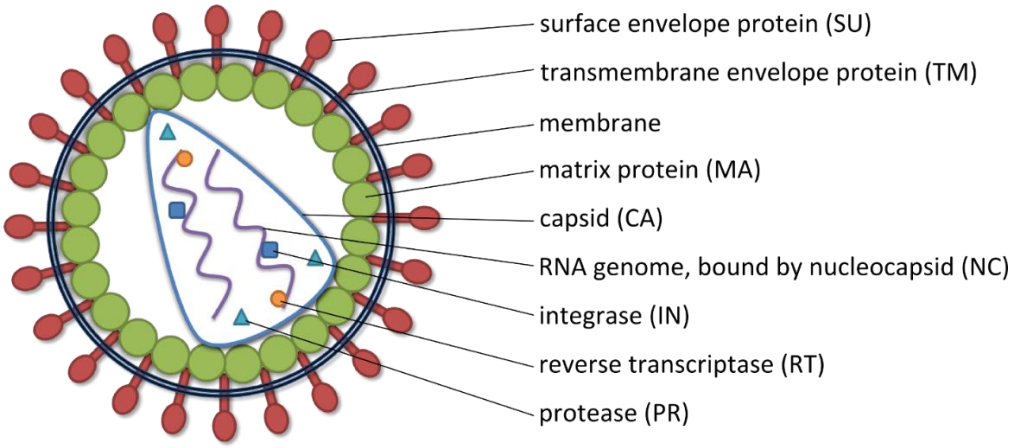


FIGURE 2 SCHEMATIC REPRESENTATION OF A RETROVIRUS PARTICLE STRUCTURE

-Proteins with specific functions:

The group-specific antigen (gag) proteins are the major component of the capsid of the virion. Pol Proteins are responsible for synthesis of viral DNA, by reverse transcription, and integration of this reverse-transcribed DNA into the host genome. Env proteins are absolutely crucial in the retrovirus life cycle: the surface component of the Env proteins allows the retrovirus to bind to its target membrane and the membrane-anchored-trans-membrane-component then facilitates the membrane fusion, therefore these particular proteins are responsible for the infectious abilities of the virus. Retroviruses also contain proteases which differ between specific viruses. Their function is to facilitate proteolytic cleavages enabling maturation of gag and pol proteins. [6,7]

1.1.5.2. Retrovirus lifecycle

Retrovirus lifecycle can be divided into five main parts: fusion, reverse transcriptase, integration, transcription and budding. The fusion phase begins with the surface components of env protein binding to the receptor on the membrane of the host cell. After the binding, the membranes of the virion and the cell fuse together and the contents of the virus enter the host cell. [6,7]

The first step of the life cycle after the virus enters the host cell is the reverse transcriptase, which means that a complementary strand of cDNA is created from the retrovirus RNA. The RNA is degraded after the process is finished and the integration phase begins: first the cDNA strand is further replicated to create two complementary cDNA strands, between which a weak bond is established, and they are transferred together into the host nucleus. The integrase facilitates the integration of those cDNA retrovirus strands into the host DNA. [6,7]

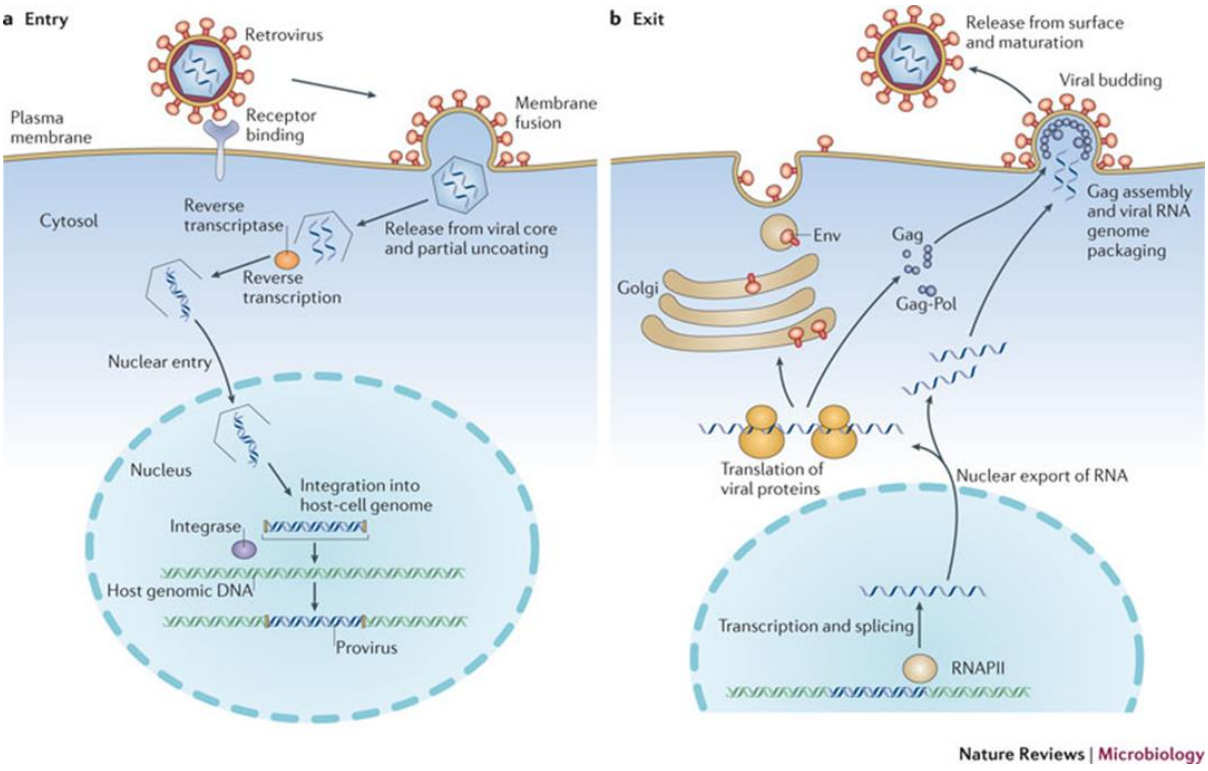


FIGURE 3 ILLUSTRATION OF DIFFERENT EVENTS IN THE LIFE CYCLE OF RETROVIRUSES

After the retrovirus genome is integrated into the host DNA, the cell either stay dormant, which means that for the time being the virus genome will not be transcribed and virion particles will not be produced, or transcription of this region can occur, thus marking the fourth phase of the retrovirus life cycle. The transcribed viral RNA is translated into proteins. The Viral RNA is then transported into the cytoplasm where it will combine with the viral proteins to assemble into new virions. In the final step - the budding – the virus leaves the cell and enters the extracellular space as a fully matured retrovirus capable of infecting other cells, thus closing the retrovirus life cycle. [6,7]

1.1.5.3. Sendai virus and its replication mechanisms

Sendai virus is also known as murine parainfluenza virus type 1 or hemagglutinating virus of Japan. It belongs to a family Paramyxoviridae, subfamily Paramyxovirinae and genus Respirovirus. It is a single-stranded RNA virus (15,384 bases), with 150-250 nm in diameter, which infects respiratory tracts of mostly mouses and rats - transmitting airborne and by direct contact. [13]

There are six wild-type sendai virus genes, which are coding for viral proteins needed to form a viron particle [13]:

- Fusion protein, which enables the viron evelope to fuse with the membrane of the host cell upon entering.

- Hemagglutinin Neuraminidase is responsible for recognizing a receptor on the surface of target cell, which is sialic acid.

- Matrix protein form the matrix inside the viron particle supporting the envelope structure of the virus.

- Nucleocapsid protein is responsible for forming core nucleocapsid complexes with the genome RNA.

- Phosphoprotein, which is a small subunit - that together with the large one - forms RNA polymerase.

- Large protein, which is a large subunit - that together with the small one - forms RNA polymerase.

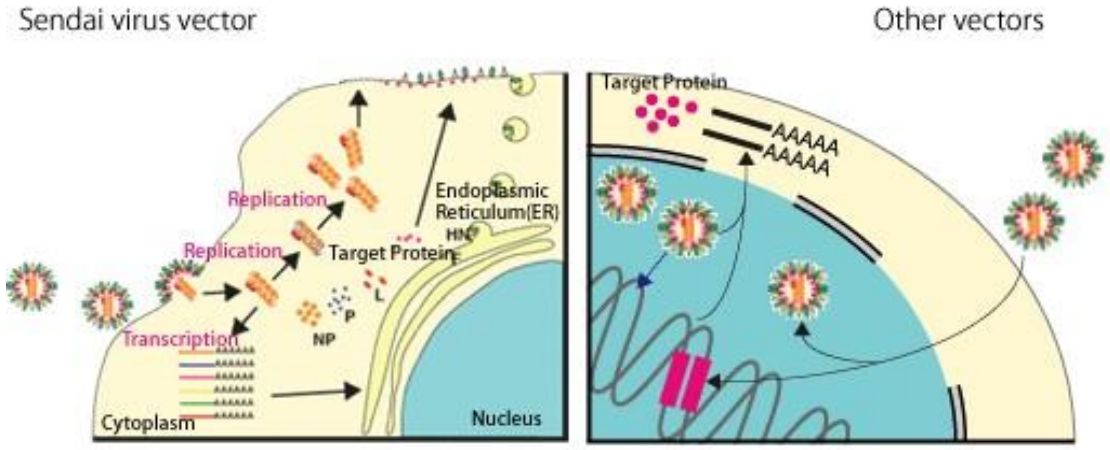


FIGURE 4 COMPARISON OF THE LIFECYCLES OF NON-INTEGRATING SEV VECTORS AND OTHER, INTEGRATING VECTORS

Sendai virus has a cytoplasmic mode of replication, which means that it does not integrate into the host genome, but carry out its replication cycle in the cytoplasm of the host cell. First the Hemagglutinin Neuraminidase recognizes the sialic acid on the surface on the cell membrane of host cell and attach to it. After attachment fusion protein causes the viron envelope to fuse with the membrane of the host cell and as a result of this ribonucleocapsid enter the cytoplasm of the host cell. In the cytoplasm the viral message RNAs are capped, polyandenylated and sequential transcription starts. When enough viron proteins are present an encapsulation begins, which leads to the budding off the virus by interaction between the matrix proteins and plasma membrane. The replication of the virus ends, by making a full cycle, when a mature viron particle, able to infect other cells, is released from the host cell into inter-cellular space. [13]

This means that the sendai virus does not integrate genomic information into the host genome, which might be an important factor, when using the reprogrammed cells for research or clinical purposes as an integrated virus genome might have a negative impact on the cell behavior.

1.1.6. Transfection and transduction

Transfection is the introduction of nucleic acids into eukaryotic cells by non-viral methods using methods such as electroporation or lipofection. This method results in transient expression of the genes of interest since the plasmid is diluted as the cells divide. Biomedicine market is abundant in transfection reagents facilitating the delivery of DNA into cells. [8] Comparing different kind of transfection or transfection reagents would go far beyond the scope and extent of this work. Specific information about those reagents, their method of use, limitations, efficiency, protocols etc. are provided by the manufacturers.

Transduction is a process during which foreign genetic information is introduced into a host cell via viral vector. [9] For instance, a modified retrovirus, in which the gag-pol-env genes are replaced by the gene of interest in the RNA virus, could be used to overexpress this gene in cells. A major advantage of retroviral transduction is the prolonged expression of the gene of interest due to integration into the genome of the host cells. For the retrovirus reprogramming described in this work both, the transfection and transduction, are used. First 293T cells are transfected with viral vectors in order to generate viral particles that will be then used to transduce fibroblasts. The sendai virus reprogramming is carried out by direct transduction.

The viruses will contain the four Yamanaka reprogramming factors. For safety reason, the viral vectors I use in laboratories are modified to produce replication defective viruses, which are viruses that are capable of entering the host cell, carrying out their replication cycle, which involves expressing the genomic information they are carrying – involving in this case reprogramming factors – but are unable to infect other cells. This is important for a variety of reasons, most importantly for safety and uncontrolled gene expression.

1.1.6.1. Retrovirus plasmid design

Various commercially available plasmids can be used with the proper transfection method, the important part for this step is to express high level of proteins into 293T cells. Our genes of interest - in this case reprogramming factors - should be inserted into a vector that contains the retrovirus long terminal repeat (LTR), which is a part of retrovirus genome containing promoter and enhancers, with 5' prime end portion of genome responsible for gag proteins, which contains in it the packaging signal (commonly marked by Ψ sign).

Along with this vector, plasmids containing information necessary for the expression of gag, pol and env proteins need to be transfected into the cells (but strictly without the packaging signal).

This co-transfection will lead the 293 cells into producing replicative deficient viruses that can be used to transduce somatic cells for the reprogramming process. This way the reprogramming factors will be expressed in the somatic cells, but the cells won't produce retroviruses.

1.1.6.2. Sendai virus design

The crucial part of the design of the sendai virus used for reprogramming is that their gene responsible for encoding the fusion protein is deleted, making the replicated particles incapable of infecting cells. In other words, the deletion of F protein encoding sequence, leads to production of replicative deficient viruses. [13]

1.1.7. Human Embryonic Kidney 293 cells

Human Embryonic Kidney 293 is a cell line derived from an aborted human embryo kidney cells in The Netherlands by introduction of adenovirus 5 DNA. Since then, it has proven to be easy to culture and to transfect. It has been used widely in laboratories around the world, since it exhibits extraordinary transfectability. [10,11]

293T cell line is a variant of HEK 293 family, which contains SV40 Large T-antigen. This T-antigen is able to facilitate episomal replication of plasmids with matching SV40 origin of replication. This enables an amplification of transfected plasmids and prolonged expression of target gene. All this factors combine to make HEK 293 and 293T cell line a widely used line for the production of retroviral vectors and are used as a base for establishing retroviral packaging cell line. [10,11]

The transfection of 293T cells with retroviral vectors offers an easy, cheap and efficient method to produce replication defective virus.

1.1.8 Fibroblast

Fibroblasts are the most common cells of connective tissue in the human body. They secrete various important extracellular matrix components including collagene. They play an important part in creating and maintaining the structural framework for many cells and tissues and also wound healing. [12]

Virtually any adult somatic cell in the human body can be reprogrammed to a state of induced pluripotency. I am focusing on fibroblast, since these were the cells I was working on. Fibroblasts are easy to obtain through skin biopsy and have been shown to be readily reprogrammed. However, different type of cells can be used (such as blood cells) but different protocols and reprogramming methods might be necessary.

1.1.9. Cell line characterization

In the process of reprogramming cells that were successfully reprogrammed are picked based on their morphology and/or expression of fluorescence markers. However, lines can differ greatly in their properties and characterization of the cell line should be carried out in order to confirm its pluripotency, its ability to differentiate and other characteristic of the cell that are crucial to its function and might have been distorted during the reprogramming. Also properties of the induced pluripotent stem cell line might differ based on the cell line it was derived from. [1, 14]

I used following methods to characterize the derived iPS cell lines.

1.1.9.1 Immunocytochemistry (ICC)

Immunocytochemistry (ICC) is a laboratory technique that can be used to assess the presence and localization of a specific protein or antigen in cells. First a primary antibody binds to the target antigen and then a secondary antibody with a conjugated fluorophore bind to it, therefore a fluorescence microscope can be used to observe the presence and location of the specific protein or antigen by observing the location of fluorescing secondary antibody. [15]

NANOG, OCT4, Tra-1-60, SSEA3, and SSEA4 are all pluripotency markers for which primary antibodies I used in ICC. [1]

1.1.9.2. Real-time quantitative polymerase chain reaction (qPCR) for pluripotency and differentiation markers

Polymerase chain reaction (PCR) is a molecular biology technology that can be used to greatly amplify - across several orders of magnitude - strands of genomic information. A PCR reaction must contain: a DNA template, which contains the region that will be amplified; two primers that are complementary to the 3' end of the target DNA fragment; DNA polymerase; Deoxynucleoside triphosphates, which are building blocks for newly synthesized DNA strands and Buffer solution to provide an optimal environment for the amplification. [16]

It usually consists of 20-40 repeated temperature changes cycles with three temperature steps. These are the temperature steps:

1) Denaturation - this step disrupts the hydrogen bonds between complementary bases, breaking the double-stranded structure of the DNA, thus yielding single-stranded DNA. A temperature of 94-98 degrees for a period of 20 to 30 seconds is used to obtain this effect.

2) Annealing - during this step the two primers should bind to their target sequences. The temperature during this step can differ greatly from around 50 degrees Celsius up to 65 and last between 20 to 40 seconds. The temperature have to be adjusted accordingly to the sequence of the primers: both a too low temperature and too high temperature can lead to unsuccessful PCR or distorted results. The more accurate the temperature chosen for the annealing step the more accurate and reliable will the results be.

3) Elongation - during this step the polymerase is synthesizing new DNA strands. The temperature and time of this step have to be adjusted to the specific conditions - the temperature will differ based on the polymerase used, while the time has to be adjusted to the length of the amplified DNA fragment.

These steps are accompanied by non-cycling steps: the initialization step, the final elongation and possibly final hold. [16]

This PCR method that aim at amplification of target DNA can be also used to quantify the initial amount of targeted DNA sequence in the sample, when used in the so-called real-time setting, because the amplification is monitored in real-time, throughout the whole reaction. This can be done with a DNA-binding dye that will bind to all double-stranded DNA in the sample, causing fluorescence. With the amplification of the DNA more dye will bind, leading to an increase in the fluorescence proportionally to the amount of DNA amplified, which depends on the initial amount of the DNA in the sample. The fluorescence can be measured by a detector after each cycle. [17]

I used qPCR to quantify the expression of pluripotency markers –Oct4, Nanog, Sox2, hTERT and Rex1 – and differentiation or lineage markers (after two weeks of differentiation) – AFP, Brachyury, Sox17, Pax6, Map2 and GATA2. [1, 14]

1.1.9.3. Karyotyping

Karyotype. coming from greek word karyon - "nucleus" - and typos - "general form", is the number and arrangements of chromosomes in the nucleus. In humans the basic number of chromosomes 46 in the somatic cells and 23 in the germ lines. Karyotyping is the action of determining the number of chromosomes and their arrangement in the sample cell - it is done for the purpose of cell line characterization in order to determine whether the cell have gained any chromosome abnormalities, which might have negative impact on the cell behavior. [18]

1.1.10. Established cell line expansion and usage

Established induced pluripotent cell lines that express pluripotency markers, differentiation markers – after two weeks of embryonic body (EB) differentiation and have a normal human karyotype, have the potential to be infinitely expanded in cell culture, through continuing proliferation in suitable conditions. Furthermore, such iPS cell lines have a number of possible research applications.

The iPS cells can be used in regenerative medicine, toxicology, disease models, drug screening and number of other usage. Also an important feature is that patient specific iPS can be derived, which allows for unique opportunities, in each of those areas: in *in vitro* drug screening, for example the impact of a drug on neurons derived from iPS lines generated from patients with amyotrophic lateral sclerosis; in disease modeling, as the patient derived iPS cells can be differentiated into any cell type allowing the researchers to study the disease in the dish; in development of therapies, especially in

cases of hereditary diseases as it allows for an in vitro testing of various autologous therapies and correction of gene defects before the reconstitution of tissue. [1, 19]

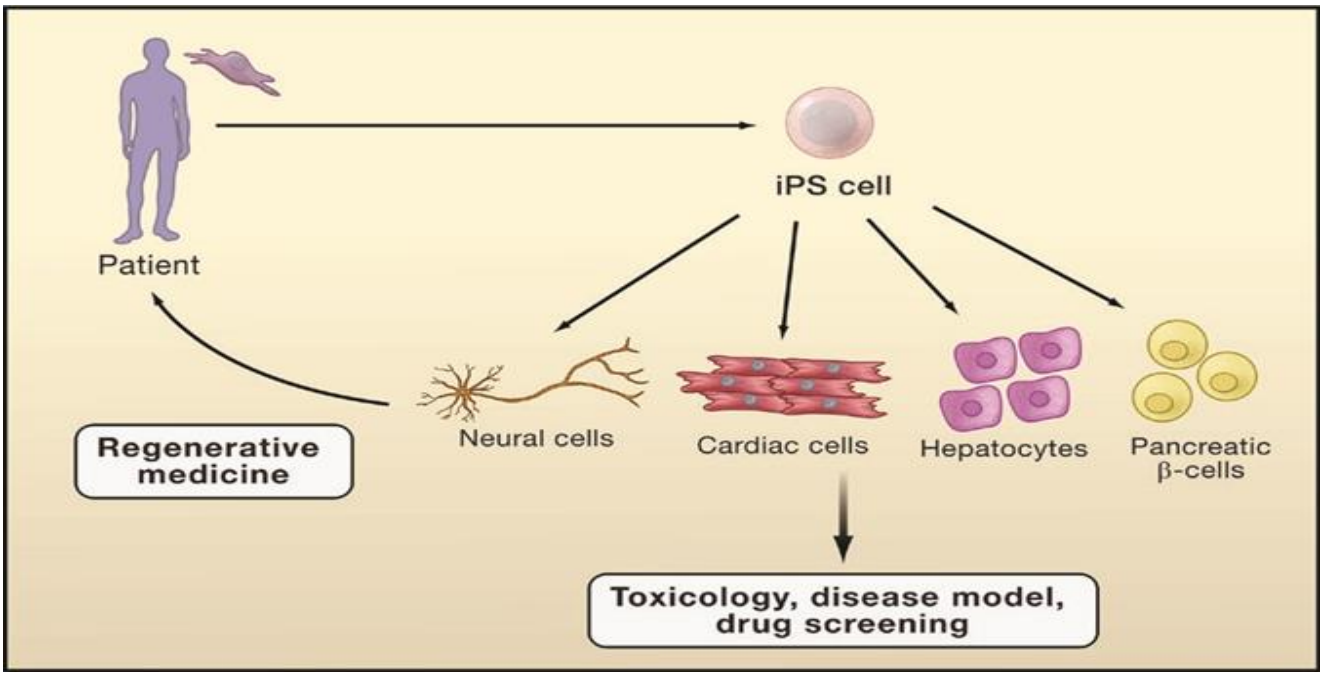


FIGURE 5 APPLICATIONS OF IPS CELL TECHNOLOGY

1.2 Gene editing in pluripotent stem cells

1.2.1. What is genome editing and why use it on iPS?

Genome editing refers widely to techniques that allow making specific, targeted changes in the genetic information of cells. It might range from deletion of one base pair to insertion of whole genes. It might cause a genetic loss-of-function or gain-of-function. It might be used to mimic genetic disorders in vitro or test genetic therapies. It can be theoretically used to cause any change in genomic information including creation of new ones or deletion of existing ones, the only practical limitation is the available technology and imagination of the architect.

Why this kind of techniques are especially useful when dealing with iPS is because they have enriching effect on the usage of those cells. In development of therapies and in vitro testing of various autologous therapies, genome editing methods allow the correction of the gene defects before the reconstitution of tissue. By introducing a specific mutation in wild type iPS cells or repairing a mutation in diseased iPS cells,this technology offers unique opportunity to study the effect of the mutation on the disease and to test new drugs. With genome editing tools, it is possible to knock-out specific genes in iPS cells, to investigate their functions and the results of their alterations in vitro. The connection of having pluripotent, immortalized cells in culture with the possibility to regulate their genomic information greatly expands the research opportunities. Every advance in the technology or understanding of the theory behind it expands those possibilities even further. [1, 19]

1.2.2. Gene reporter cell line

Reporter gene is a gene that can be attached to a gene interest and report on its activity - meaning that it indicates the expression of the gene in a detectable way. Usually, the reporter gene provides

an easy, reliable way to quantify or confirm the expression of the gene of interest. A gene reporter cell line is a cell line that has been engineered to have a reporter gene. [24]

1.2.3. GCM2

Glial cells missing homolog 2 (GCM2) is a gene found on homo sapiens sixth chromosome that codes for a transciptional factor. This transcriptional factor is believed to act as a master regulator of parathyroid development and its mutation is linked to hypoparathyroidism. Therefore, it is believed that the GCM2 has to be expressed in order for a cell to differentiate into a parathyroid state. [25]

The Harvard Stem Cell Institute iPS core facility was contacted by an affiliated lab of Harvard Medical School and Massachusetts General Hospital (MGH) that is working on developing a reliable protocol for differentiation of iPS cells into parathyroid cells. To facilitate their work, they are interested in obtaining a gene reporter line for the gene GCM2. The GCM2 protein is a transcriptional factor , located in the nucleus, thus it cannot be used to detect parathyroid cells by a standard go-to method, FACS. This would require a permeabilization of the cells, which would lead to cell death..

Engineering a cell line with a gene reporter attached to the GCM2 gene would provide a much faster, cheaper and more efficient method of assessing the expression of the gene by FACS or using a fluorescent microscope.

1.2.4. Targeted allele and untargeted allele

The GCM2 gene reporter cell line is supposed to facilitate the development of differentiation protocol for parathyroid cells from IPS cells. Since the GCM2 protein is believed to be a crucial factor in such differentiation, it was important, when designing the reporter gene and the method of introduction into the genome, not to impair the gene itself or its protein output.

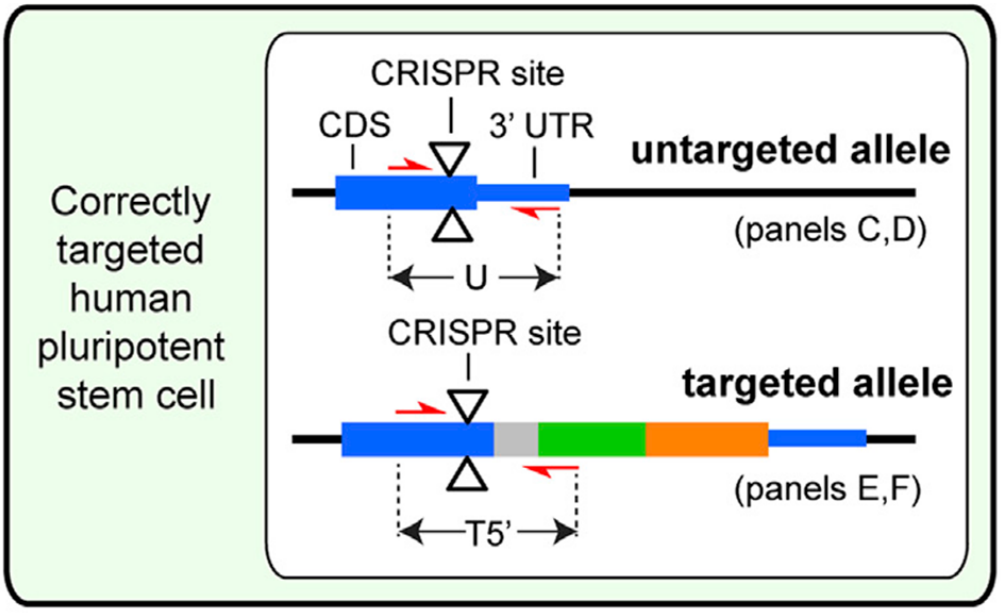


FIGURE 6 SCHEMATIC OF A CORRECTLY TARGETED HUMAN PLURIPOTENT STEM CELL CLONE SHOWING UNTARGETED AND TARGETED ALLELES, CRISPR-CAS9 CLEAVAGE SITES (TRIANGLES), AND THE SEQUENCED AMPLICONS.

To achieve this goal, we decided to use the viral 2A peptide for bicistronic expression of GCM2 and GFP. This strategy should allow the expression of GCM2 to remain undisturbed. Both alleles could

be targeted but to minimize the risk of modification of GCM2 expression, we preferred to select clones with one targeted allele and one intact allele. Therefore it was crucial to design an appropriate insert and to choose the best methods for delivery and screening.

1.2.5. Design of the insert

Sequence of the insert for the gene reporter cell line was designed by the MGH lab. It begins with a 5' and 3' arm homologous to the GCM2 gene, between those there are three distinctive parts with specific functions: P2A, eGFP and Puromycin drug resistance cassette.

1.2.5.1. P2A

There are different ways to link the expression of a fluorescent marker to a specific gene in cells. The most widely used is an insertion of internal ribosomal entry sites (IRESs) between the gene of interest and the fluorescent marker. However, IRES insert has some limitations: the gene located after the IRES has much lower translation efficiency than the one placed before it. The MGH lab wanted to have the same expression of GCM2 and gene reporter and therefore they decided to use a 2A peptide (P2A). [26]

2A is a small peptide, on average between 18-22 amino acids, that is subject to self-cleavage at a very high rate. It can be inserted between two genes that are then co-transcribed and co-translated. After translation, however, the P2A will be cleaved and the proteins coded by the two genes will be separated and expressed at the same level.. [26]

1.2.5.2. eGFP

Green fluorescent protein (GFP) is a protein that exhibits bright green fluorescence when exposed to light from the blue to ultraviolet spectrum. It is composed of 238 amino acids and is frequently used as a mean of gene expression reporter. It has been used in variety of organisms, such as bacteria, plants and fish among others. [27]

In this insert a derivative of GFP is used called enhanced green fluorescent protein (eGFP) that was discovered in 1995 and that has a mutation that increases its folding efficiency at 37 degrees Celsius making its use more practical in mammalian cells. [28]

1.2.5.3. Puromycin drug resistance cassette

Puromycin is an aminonucleoside antibiotic that is naturally produced by and was derived from the bacterium Streptomyces alboniger. By causing premature chain termination during translation it is a potent translational inhibitor in eukaryotic cells. It has a fast mode of action and can cause relatively quick cell death even at low concentrations. Resistance to it can be conferred by the puromycin N-acetyl-transferase gene that is naturally present in Streptomyces. The addition of this cassette will facilitate the selection of clones in which the reporter gene is inserted.

1.2.6. CRISP/Cas9

To insert the gene reporter in human iPS lines, we choose the CRISPR/Cas9 methodology.A genome editing system derived from clustered regularly interspaced short palindromic repeats (CRISPR) associated RNA guided endonuclease Cas9, which is a part of microbial antiphage defense system, has enabled easy and robust genome editing of DNA sequences in almost any organism of choice. [31]

A variety of studies ranging from 1992 to 2003 has shown that introduction of double-strand breaks (DSBs) into DNA can stimulate homologous recombination (HR) events that can be used as a tool in genome editing. It was also shown that in the absence of an exogenous homology repair template the DSBs can lead to insertions or deletion (indel) mutations by nonhomologous end-joining (NHEJ), which is an error-prone repair pathway. [31]

CRISPR nuclease Cas9 targeting system is based on Watson-Crick base pairing, by which a short guide RNA recognizes target DNA. The guide sequences , as part of natural antiviral defense, typically correspond to phage sequences, however they can be easily replaced by a sequence of interest to retarget the Cas9 nuclease. This allows a large-scale, high-efficiency, site-specific targeted by introduction of short guide RNAs. [31]

An important feature of the CRISPR/Cas9 system is that the targeting site needs to be immediately upstream of a protospacer-adjacent motif (PAM) that dictates the DNA target search mechanisms. The PAM sequence is absolutely necessary for target binding and this sequence is dependent upon the species of Cas9. For the most commonly used Cas9 (from Streptococcus pyogenes) PAM is a three nucleotide sequence NGG, however others PAM were discovered and tested using additional species of Cas9. It facilitates self-versus non-self-discrimination of Cas9 and determines the overall DNA targeting space of Cas9. [31]

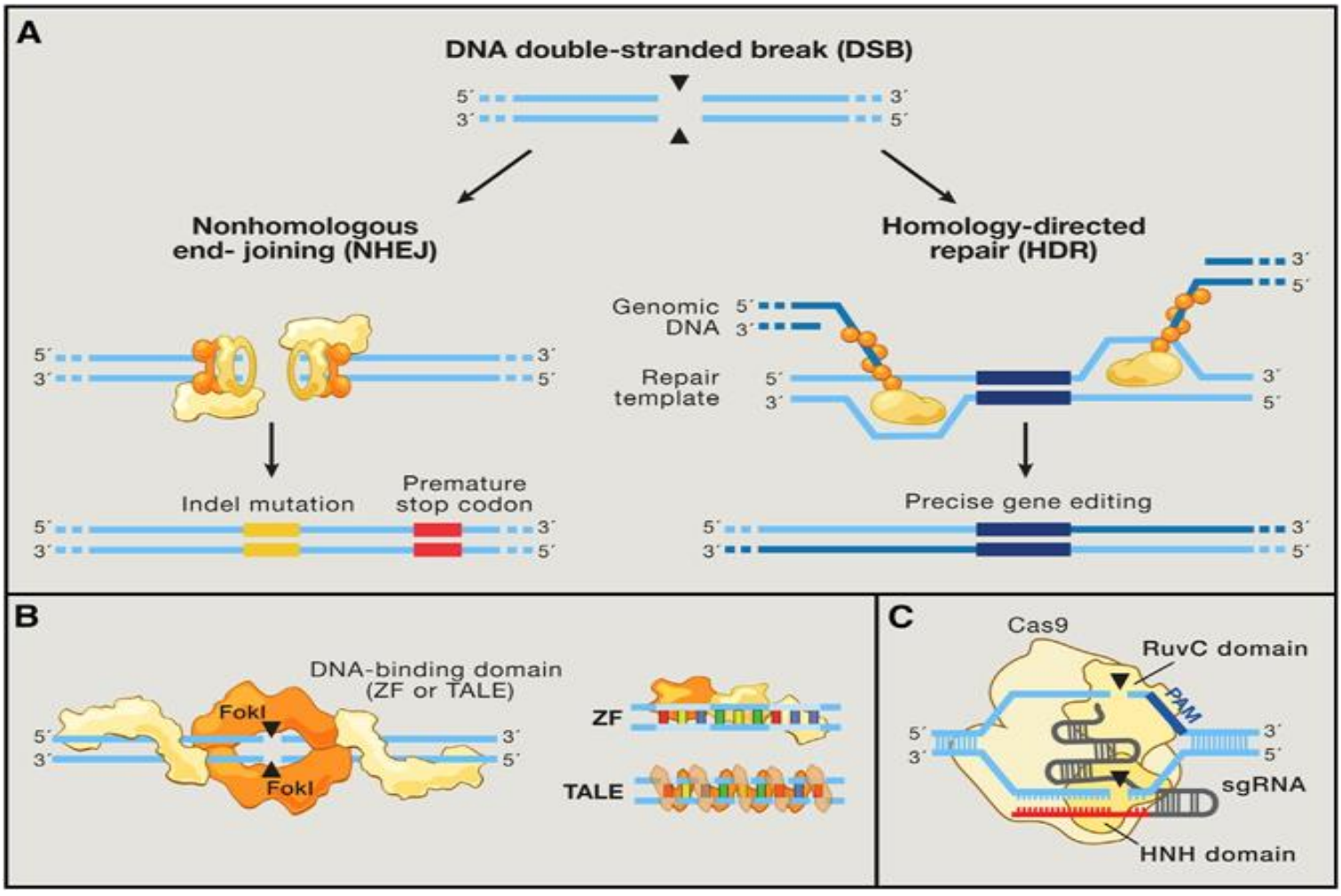


FIGURE 7GENOME EDITING TECHNOLOGIES EXPLOIT ENDOGENOUS DNA REPAIR MACHINERY

To use the Crisp/Cas9 system in our project, it was important to design guides RNA that would provide sufficient targeting efficiency, while minimizing the possible off-target effects. We also had to design suitable screening technique that would allow checking whether precise gene editing had occured, with template insert in place in one allele, and noindel mutation as an effect of NHEJ, in the other allele. [31]

1.2.7. Guide RNA design

Guide design was done by the MGH lab using a CRISP design web tool created by Zhang Lab at Massachusetts Institute of Technology. The web-tool uses a specific algorithm to scan a sequence of choice for possible guides (20 nucleotides followed by a PAM sequence) and then scan for possible off target matches throughout the selected genome (human genome for this project). Then the web-tool assign scores to the guides based on the possible off-target effects genome-wide and by target specificity.

Because the guides designed by this web-tool differ in their properties, it was decided to choose few guides and then test them for their ability to target the right sequence. For the testing of the guides, we used the TIDE assay.

1.2.7.1. TIDE

Genome editing tools are based on the introduction of DSB on specific sites and their repair by NHEJ or HDR. The DSB is induced by introducing endonuclease with programmable sequence specificity into a pool of cells. The efficiency of different endonucleases and the amount of off-target effects can differ greatly, therefore it is usually necessary or advisable to measure their properties. Tracking of Indels by DEcomposition (TIDE) is a simple, cost-effective method to measure the cutting efficiency. It requires only two parallel PCR reactions and a pair of standard capillary sequencing analyses. Those sequences are then analyzed using a software with specially developed decomposition algorithm that identifies the major induced mutations in the targeted genome editing site and then determines the frequency of those mutations in a cell population. [32].

To facilitate the process, we use HEK293T cells (cells that are easy to transfect) to test the effciency of the designed guides. The guide with the highest efficiency can then be used for the targeting of the Human iPS lines.

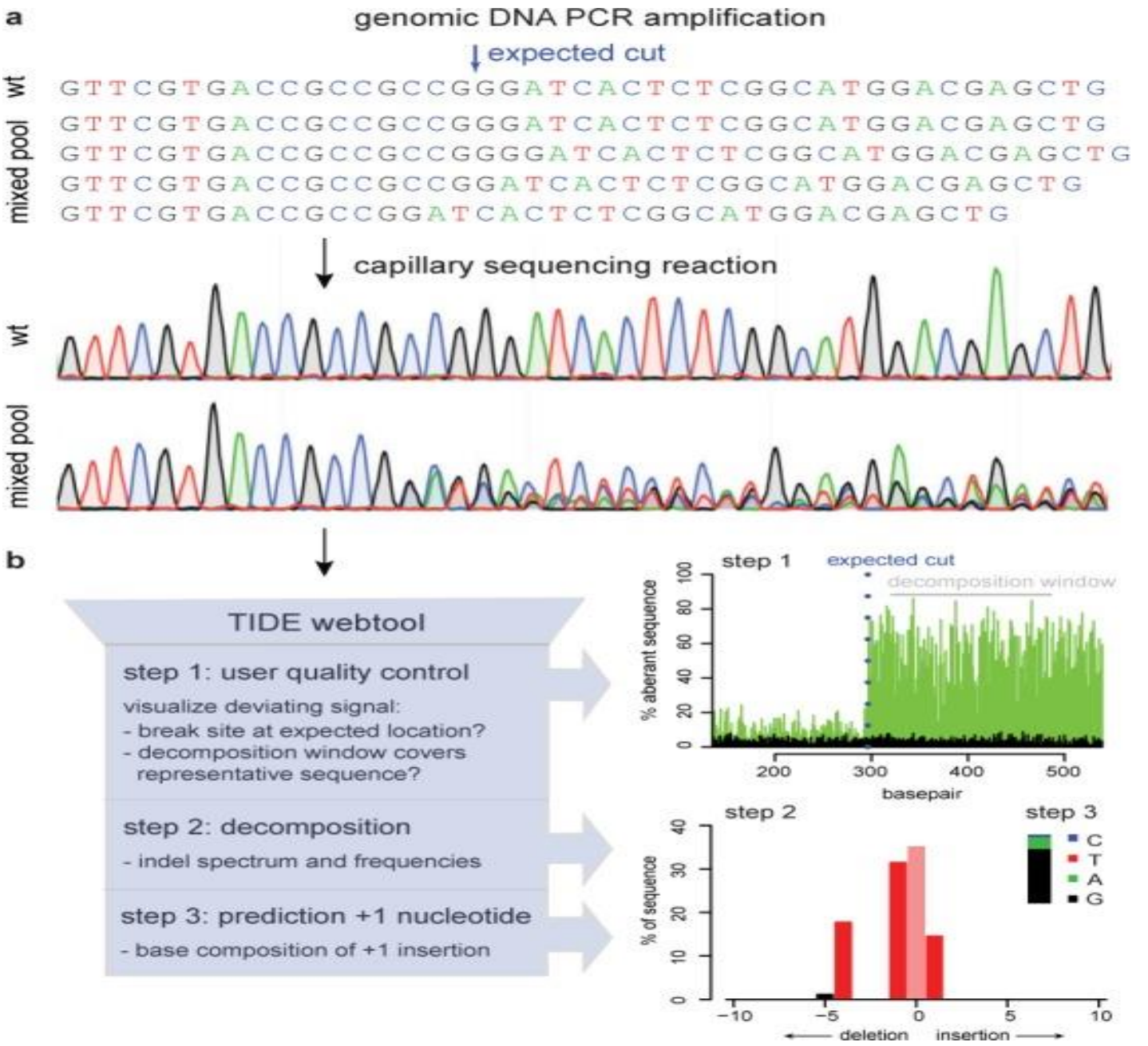


FIGURE 8 ASSESSMENT OF GENOME EDITING BY SEQUENCE TRACE DECOMPOSITION

1.2.8. Screening – searching for the right clone

After a pool of iPS cells has been transfected with the selected guide, the Cas9 endonuclease and the template insert, the next step is to screen those cells in order to find the correctly targeted clones, which in this case, is a presence of template insert at the correct location (targeted allele) and no indel mutations at the other allele of the gene (untargeted allele). To do this a multiple step screening process has been designed.

1.2.8.1. Puromycin selection

First of all, since the template contains a puromycin resistance gene, cells can be treated with puromycin, which should lead to death of all cells without drug-resistance cassette. However, this step has to be preceded by a kill-curve experiment, which is an experiment where cells are subjected

to increasing amounts of antibiotic up to a point where all cells die in a specific amount of time. It is recommended to perform such a dose-response experiment for each combination of antibiotic and cell line, since each cell line can respond differently to different dosage. Successful kill-curve experiment allows determining the amount of antibiotic that has to be used to carry out an effective selection in a given amount of time.

1.2.8.2. PCR screening for targeted allele

After eliminating non-resistant cells from the pool, the next step is to determine whether the remaining clones have the gene reporter inserted at the right location. To do this, the MGH lab designed two sets of primers to amplify a sequence only if the insert was present at the right location. One of the primers (forward/reverse) is annealing on the gene itself and the other (reverse / forward) on the template, therefore a template integrated at another part of the genome won't be amplified.

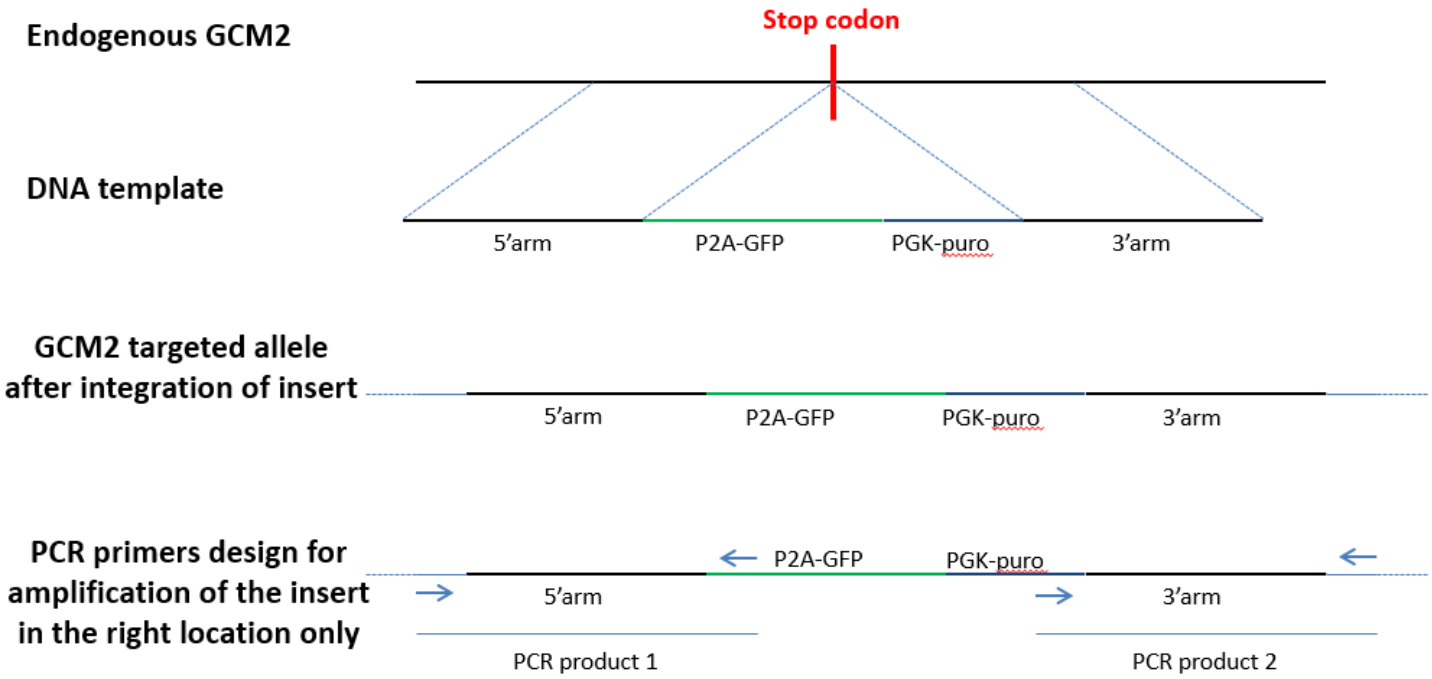


FIGURE 9 SCHEMATIC DEPICTION OF PRIMER DESIGN USED FOR SCREENING FOR TARGETED ALLELE

1.2.8.3. PCR screening for wild type GCM2 and further tests

After selection of clones with at least one targeted allele, the next step is to test the status of the second allele. A set of primers was designed to amplify the region surrounding the stop codon of the GCM2 gene which is the region targeted by the CRISPR/Cas9 system. .This region can be amplified only if there is no insert of the gene reporter therefore a positive PCR indicates that the second allele was not targeted by HR. However, the PCR product needs to be sequenced to be able to detect if there is an indel mutation on the allele as a result of NHEJ. Only clones without mutation on the second allele will be selected for further use. If the PCR is negative, it could indicate that both alleles were targeted with the gene reporter or that the allele has been damaged and cannot be amplified.

1.2.8.4. Activating GCM2 using SAM complex

As mentioned earlier in this report, Cas9 is a endonuclease that can be guided to a specific sequence in a genome to introduce a double-stranded break. However, Cas9 can be converted into a transcriptional activator by inactivating its two catalytic domains and then fusing a transcription activation domains. This has been done in the past with some degree of success, but the magnitude of transcriptional up-regulation achieved by this system has been poor and other challenges arose during usage. [33]

\ Fen Zhang et al. recently developed an engineered CRISP-Cas9 transcriptional activation complex that is robust, specific, can be used for single and multi-gene activation and can facilitate a genome-scale screening with high degree of consistency and validation. They first observed on the crystal structure of Cas9-sgRNA-target DNA ternary complex, that the tetraloop and stem loop 2 of the sgRNA protrude outside of this complex and the distal 4 bp of each of them is completely free of interactions. They used those loops to add a minimal hairpin aptamer that selectively binds dimerized MS2 bacteriophage coat proteins, which greatly increased upregulation of transcription efficiency. Further tests proved that combination of different activation domains act in synergy to also enhance the level of transcription activation - and that addition of HSF1 transactivation domain further increased it. [33]

They designated the system they created as a result of those tests: synergistic activation mediator (SAM). It comprises a dCas9 with VP64 activation domain, single-guide RNA with two MS2 domains and MS2-p65-HSF1 transactivations domains. [33]

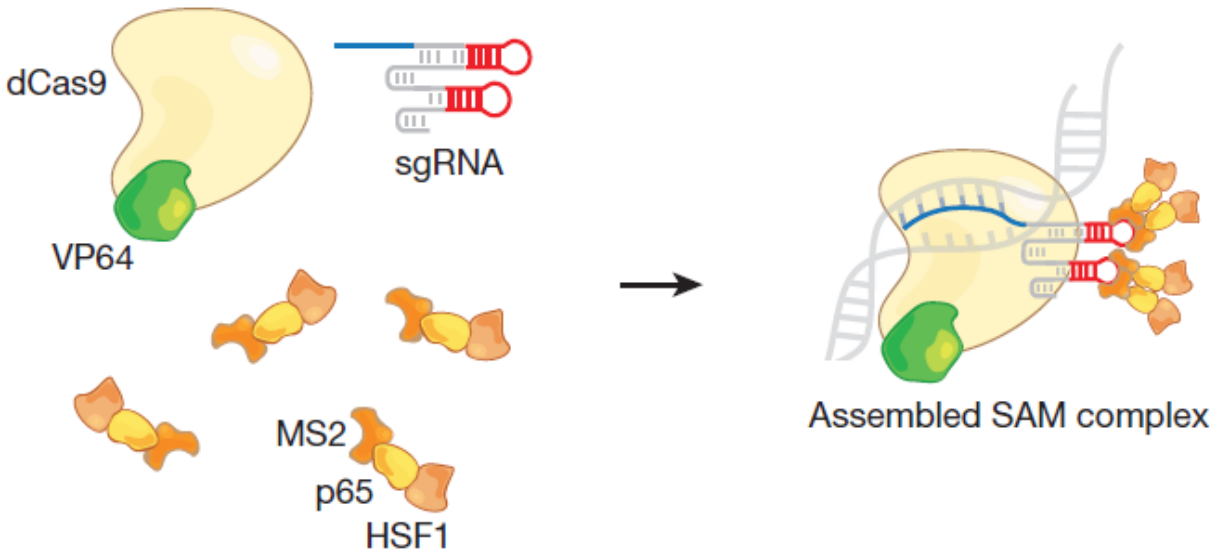


FIGURE 10 SCHEMATIC OF THE THREE-COMPONENT SAM SYSTEM

I used this system to activate the GCM2 gene in the newly created gene reporter cell line to confirm its function. If the gene reporter works as intended there should be a strict correlation between GCM2 expression and GFP presence in the cell. To properly conduct this assay I first tested different guide RNA in HEK 293 cells and singled out one guide that led to the highest upregulation of GCM2. Then I transfected the gene reporter cell line with SAM complex – containing the chosen guide – and sorted

the cells for GFP positive and negative using fluorescence-activated cell sorting. Then these cells were used for a qRT-PCR to measure the difference in expression of GCM2 in GFP positive cells versus negative.

2. Materials and methods

2.1 Generation of induced pluripotent stem cells from fibroblast lines

2.1.1. Reprogramming protocols

2.1.1.1 Retrovirus reprogramming protocol

The following protocol is based on the work of Park I.H, et al. in 2008 [20]. The incubation condition for the cells is 37º C with 5% CO_2. If not stated otherwise, the medium for fibroblast, 293T and MEF cells is DMEM 10% FBS with pen/strep.

On Day 0, 293T cells are plated in a T75 flask and incubated in appropriate conditions. After 72 hours, on Day 3, the flask containing 293T cells should be confluent and the cells can be collected using Trypsin/EDTA for 4-5 minutes. DMEM 10% is then added to inhibit the trypsin and the cells are collected in a 15ml tube. The tube is centrifuged, the supernatant removed and cells are re-suspended in DMEM 10% FBS without pen/strep media. Re-plate the cells into four 10 cm plates with one fourth of the collected cells in each plate.

On the same day, thaw one frozen vial of human fibroblast – those will be the reprogrammed cells. Plate those fibroblast cells into a T25 flask.

After 24 hours – on day four of the protocol – all the 10 cm plates containing 293T cells should be over 90% confluent. These cells will be used for transfection of retroviral plasmids.

Transfect the 293T cells with a transfection method of choice to ensure the production of replication defective viruses carrying the appropriate reprogramming factor Oct 4, Sox2, Klf4 or c-Myc. I used Lipofectamin 2000 reagent from Life Technologies and followed their recommendation.

After 24 hours – on day five of the protocol - remove the medium from 293T cells plates and replace with fresh DMEM 10% with pen/strep. Since these cells can detach easily, be very careful and gentle when adding the medium. The 293T cells are already producing the retrovirus virions, therefore appropriate handling in regards to biosafety recommendation is necessary.

On that same day, collect the fibroblast cells using Trypsin/EDTA. Centrifuge the cells, remove the supernatant, re-suspend the pellet in the media and count the cells. 100 000 fibroblast cells per well are then plated into each well of a 6-well plate. Two of those wells will be used for reprogramming, while four of those to test each virus individually.

After 24 hours – on day six of the protocol – collect the medium from 293T cells plates, which now contains viruses, and filter it through 0,45u filter. After collecting all the medium, add fresh DMEM 10% FBS with pen/strep to 293T cells. Again, the cells can detach easily, so gentle and careful handling is necessary, as following biosafety recommendation, since virus handling occurs.

The medium collected from the 293T cells will be used for the first round of transduction. Start by removing the medium from each well of a 6-well plate containing fibroblasts. Then – for the two wells used for reprogramming - add 1 ml of media from 293T cells producing viruses carrying these reprogramming factors: Oct4, Sox2 and Klf4 and add 200 µl of media containing c-Myc virions. When the virus containing media is in the wells, add 1 ml of DMEM 10% and polybrene, at 2.5 µg/ml concentration (the latest is used to increase the transduction efficiency)

To the 4 wells containing fibroblast for the virus test , add 1 ml of one virus containing media for 1 designated well, so that each of these wells contains only one type of virus, then add 1 ml of DMEM 10% FBS and polybrene, at 2.5 µg/ml concentration.

After 6 to 8 hours of transduction, remove the media from the wells, wash them once with PBS and add 2 ml of fresh DMEM 10% FBS.

On day seven of the protocol – after another 24 hours – a second round of transduction has to be carried out. Follow the same procedure as on the previous day with the difference being: do not add media to the 293T cells plates, as this is the last round of virus collection, so they will be no longer needed, instead add vesphene or bleach to each of the plates. Leave them for at least 20 minutes before trashing the plates.

Again, as the previous day, change the media to DMEM 10% FBS after 6 to 8 hours of transduction and incubate the plates.

After 72 hours – on day ten of the protocol – check the transduction efficiency by observing the fibroblast cells under fluorescent microscope for GFP expression. Take pictures for the record. If the transduction has been successful and highly efficient, meaning that close to 100% of cells express GFP, plate Mouse Embryo Fibroblast (MEF) into two 10 cm plates – one million MEF cells for each plate.

After 24 hours – on day eleven of the protocol – collect the fibroblast from two reprogramming wells using Trypsin/EDTA and re-plate each of this well on a 10 cm plate containing MEF. Use DMEM 10% FBS as the media in the 10 cm plates.

After another 24 hours – on day twelve of the protocol – switch the medium in the 10 cm plates from DMEM 10% FBS to human embryonic stem cell medium. This change will allow the emerging iPS colonies to grow in the appropriate conditions.

Addition of Valproic Acid (VPA) at 50 µM concentration for the next 7 days is optional. VPA is a histone deacetylase inhibitor and was shown to be a reprogramming enhancer, improving efficiency of the OKSM cocktail, by upregulating the transcription of embryonic stem cell-specific genes [21].

Change the hESC medium in the 10 cm plates every other day for the next seven days and afterwards change the medium every day.

Fully reprogrammed iPS colonies should appear on the plates 3 to 4 weeks after transduction. The colonies should be selected based on morphological criteria and GFP silencing. Those picked colonies can be plated on a feeder plate, containing MEF, or on a feeder-free plate, coated with matrigel. After picking Rock inhibitor at 10µM concentration should be used to increase survival of the colonies for the first 24 hours only.

Each of the picked colonies should become one iPS line.

The efficiency of the method, evaluated as the percentage of patient fibroblast that is fully reprogrammed, is estimated at 0,01% [20]. However, only one clone is required for the creation of a new line that can be expanded, virtually infinitely, thus this protocol can be used as a reliable, robust way to derive iPS from fibroblast.

2.1.1.2. Sendai virus reprogramming protocol

For the purpose of reprogramming fibroblast cells with sendai virus I used CytoTune™-iPS 2.0 Sendai Reprogramming Kit (Catalog number: A16517; ThermoFisher Scientific). The following protocol is based on the user guide from the manufacturer [13]. I reprogrammed the cells with the use of feeder culture, however reprogramming in a feeder-free setup is also possible.

The medium referred to in this protocol as fibroblast or MEF medium is a DMEM 10% FBS with addition of 1 ml of 10mM MEM Non-Essential Amino Acids solution and 100 µL of 55mM β-mercaptoethanol.

The medium referred to in this protocol as iPSC medium is DMEM/F-12, 20% KSR with addition of: 1 ml of 10 mM MEM Non-Essential Amino Acids Solution, 1 ml of GlutaMAX™-I, 100 µL of 55mM β-mercaptoethanol, 1 ml of Penicillin-Streptomycin and 40 µL of bFGF (10 ng/mL).

The incubation conditions are 37® Celsius with a humidified atmosphere of 5% CO_2.

On day -2, 48h before transduction, plate fibroblast cells onto a 6-well plate. At least two wells have to be plated for each sample, as one well has to be used for viral volume calculations and one for transduction.. Culture the cells till day 0, split them if they become over confluent (50-80% confluency is recommended on the day of transduction).

After 48h on **Day 0** perform the transduction. Harvest the cells from one well - those cells will not be transduced - by using 0.05% trypsin/EDTA following the manufacturer instructions. Count the harvested cells using available method. Use the live cell count number to calculate the volume of each virus needed to reach the target MOI and titer information on the CoA with the following formula:

$$Volume\ of\ virus\ (\mu L) = \frac{MOI\ \left(\frac{CIU}{cell}\right) \times\ number\ of\ cells}{titer\ of\ virus\ \left(\frac{CIU}{mL}\right) \times\ 10{-}3\ \left(\frac{\mu L}{mL}\right)}$$

Fetch one set of CytoTune 2.0 Sendai tubes from storage and proceed to thaw each tube by placing it shorty (5-10 seconds) in a 37 degrees water bath and then letting it thaw at room temperature. Once all the tubes are fully thawed centrifuge them shortly and place them on ice.

Add the calculated volume from each of the three tubes to 1 ml of pre-warmed (to 37 degrees) fibroblast medium. Mix the medium by pipetting up and down.. Afterwards change the medium in wells with the cells to the prepared virus mixture. Incubate the cells overnight.

On day 1, after 24 hours, replace the virus mixture in the wells with cells with fresh fibroblast medium. For the next six days culture the cells by keeping them incubated and changing the medium to a fresh one every other day.

On day 5 or 6 prepare one 10 cm MEF culture dish per each well of fibroblast cells.

On day 7 transfer the transduced cells onto the MEF culture dishes. Wash the fibroblast cells with D-PBS, harvest the cells using 0.005%trypsin/EDTA, centrifuge the cells at 200 x g for 4 minutes, aspirate the supernatant and re-suspend the pellet in 10 ml of fibroblast medium. Count the cells using a method of choice and plate from 50,000 to 200,000 cells per 10 cm MEF plate with fibroblast medium and incubate them overnight.

On day 8, after 24 hours, change the medium in the 10 cm plates from fibroblast medium to iPSC medium. Afterwards change the medium to a fresh one every day and observe the plates under a microscope for emerging cell clumps, which are indication of cells undergoing reprogramming.

After three to four weeks, depending on the growth of the colonies, manually pick the reprogrammed colonies and transfer them onto fresh MEF plates.

All additional information and troubleshooting information is available on the manufacturer website or the user guide.

2.1.2. Cell line characterization

2.1.2.1. Immunocytochemistry

iPS grown on MEF culture dishes were washed three times with PBS and fixed in 4% paraformaldehyde for 20 min. Then, they were permeablized with PBS/0.1% Triton X-100 for 15 minutes at RT. Next they were washed 3 times with PBS/0.05% Tween 20, blocked with 4% Donkey Serum/PBS at 4 degrees Celsius overnight wrapped in parafilm.

The following day, the cells were washed with PBS and treated with primary antibodies diluted in 4% Donkey Serum/PBS. The primary antibodies were diluted as follow: Oct 4 (1:100); Nanog (1:50); SSEA 3 (1:200); SSEA 4 (1:200); TRA-1-60 (1:200). The cells were incubated for 1 hour at room temperature..

After 1 hour incubation period, the cells were washed 3 times with PBS/0.05% Tween 20 and secondary antibodies, diluted 1:500 in PBS, were added. The secondary antbodies for the corresponding primary antibodies are as follow: Rabbit antibody for Oct 4 and Nanog; Rat IgM for SSEA-3, Mouse IgG for SSEA 4 and Mouse IgM for TRA-1-60. Cover the plate with tinfoil and incubate at room temperature in the dark.

After the 1 hour incubation period, the cells were washed 3 times with PBS/0.05% Tween 20 and stained with DAPI (enough to cover the whole well). Pictures were taken under fluorescence microscope.

2.1.2.2. qPCR

RNA was extracted from harvested cells using RNeasy mini kit from Qiagen (Cat No. 74106) following manufacturer instructions. Complementary cDNA was synthesized from the RNA sample using qScript cDNA SuperMix from Quanta Biosciences (Cat No. 95048-500) following the manufacturer instructions with 1 µg of RNA used in the reaction.

The synthesized genomic cDNA was used to perform qPCR with the use of PerfeCTa® SYBR® Green SuperMix (Cat# 95054-500) following manufacturer instructions in ViiA™ 7 Real-Time PCR System from Applied Biosystems (Cat No. 4453536) with the following run set-up:

Hold stage: 95®C for 20 seconds

PCR Stage (repeated 40 times): 95®C for 1 second and 60®C for 20 seconds.

Melt curve stage: 95®C for 15 seconds, 60®C for 1 minute, 95®C for 15 seconds.

2.1.2.2.1. qPCR for pluripotency markers

For the qPCR for the pluripotency markers the cells were harvested in pluripotent, non-differentiated state. Genomic RNA obtained from fibroblast was used as a negative control and a combination of genomic RNA from three Hues lines (Hues 8; Hues 7 and Hues 9) was used as a positive control. Primers for following pluripotency markers were used: Oct4, Nanog, Sox2, hTERT, Rex1 and Actin primers as a control.

2.1.2.2.2. qPCR for differentiation markers

For the purpose of qPCR for differentiation markers the cells were first harvested and re-plated on ultra-low attachment plates (Costar# 3471) fed with hESC medium without FGF (DMEM/F-12 20% KSR with addition of: 1 ml of 10 mM MEM Non-Essential Amino Acids Solution, 1 ml of GlutaMAX™-I, 100 µL of 55mM β-mercaptoethanol and 1 ml of Penicillin-Streptomycin) every three days.

After one week, the cells were harvested and re-plated on gelatin coated wells in MEF Medium (DMEM 10% FBS with addition of 1 ml of 10mM MEM Non-Essential Amino Acids solution and 100 µL of 55mM β-mercaptoethanol), which was changed every three days. This treatment allowed the cells to form embryonic bodies (EB) and differentiate. After one week the cells were harvested and RNA was extracted.

Genomic RNA extracted from Hues cell line differentiated in the same manner was used as a positive control. Primers for following differentiation markers were used: AFP, Brachyury, Sox17, Pax6, Gata2, MAP2 and primers for Actin as a control.

2.1.2.3. Karyotyping

Chromosomal studies were performed at Cytogenetics laboratory at Tufts Medical Center Department of Pathology and Laboratory Medicine.

2.2. Gene editing in pluripotent stem cells

2.2.1. Testing guide RNA cutting efficiency by TIDE assay

Two days before transfection 250 000 HEK 293T cells per well were plated in 5 wells of a 6-well plate in DMEM/10% FBS without antibiotic. After 48h when they reached 70-80% confluence the medium was changed to a fresh one. I prepared five tubes with each of them containing 100 µl of opti-MEM (Reduced Serum Media from ThermoFisher Scientific; Cat #31985070) and four of them containing 2,5 µg of plasmid containing Cas9 and respective guides (1-4) and one tube without guide RNA (as a negative control). Another tube with 42 µl of Lipofectamin 2000 (ThermoFisher, Cat #11668019) and 600 µl of opti-MEM was prepared. Then,107 µl of dilluted Lipofectamin 2000 was added to each of the tube containing dilluted guide RNA and negative control tube with only opti-MEM. All tubes were incubated at room temperature for 5 minutes, after which 210 µl of solution from each tube was added to a respective well of a 6-well plate. After 24h media was changed to a fresh one and after another 24h, the cells were harvested using 0.05% trypsin. DNA was extracted from the harvested cells using QIAamp DNA Mini Kit (Qiagen, Cat. no. 51306). The DNA was then sent to our collaboratorsat MGH where the PCR, sequencing and TIDE analysis were performed.

2.2.2. Puromycin kill curve

To establish a puromycin kill curve, BJ RiPS iPS cells were split from existing culture to 6-wells of a six-well plate coated with Matrigel and cultured in mTeSR™1 media (STEMCELL Technologies; Catalog #05875). Upon reaching 70-80% confluency the cells were treated with increasing

concentrate ion of puromycin: 0,5; 1; 1,5; 2 and 2,5 µg/ml and one well was cultured in standard conditions as a control. Cells were cultured for the next 7 days with mTeSR™1 media and puromycin with a medium change every day.

2.2.3. Transfection

For thetransfection, BJ RiPS cells were cultured on a 10 cm plate in a feeder-free condition with mTeSR™1 media. Upon reaching 70-80% confluence cells were dissociated using Accutase and counted using Bio-Rad Automated Cell Counter (Cat No. 1450102). Three million cells or one million cells were transferred to a 15ml conical tube and centrifuged at 1200 rpm for 4 minutes at room temperature and washed twice with 2ml of DMEM/F12 medium.

During the washing step,two tubes were prepared with two solutions: one with 3.75 µl of Lipofectamine 3000 (ThermoFisher Scientific; Catalog #L3000015) Reagent in 125 µL Opti-MEM (Reduced Serum Media from ThermoFisher Scientific; Cat #31985070) and a second one, a master mix of DNA, with 3 µg of template insert, 1 µg of guide RNA/Cas9 plamid, 125 µL of Opti-MEM and 5 µL of P300 Reagent (ThermoFisher Scientific; Cat #L3000-015). Then, diluted DNA was added to the tube with diluted Lipofectamine 3000 and incubated for 5 minutes at room temperature.

After the incubation period and washing steps the DNA-lipid complex was added drop-wise to the cell pellet, which was gently loosen by tapping the tube, then mixed gently and incubated at room temperature for 14 minutes. Then the cells were plated on a 10 cm plate coated with matrigel with mTeSR™1 media (STEMCELL Technologies; Catalog #05875). 20 µM of ROCK inhibitor was added to the plate to increase the cell survival after transfection, but was not added to the subsequent feeds.

2.2.4. Puromycin selection

After transfection, the cells were cultured for 4 days with mTeSR™1 media changed to a fresh one every day. After 4 days the selection was started by adding puromycin to the media at 0,5 µg/ml concentration.

After 7 days of selection, the plate was washed with Dulbecco's phosphate-buffered saline (DPBS) (ThermoFisher Scientific, Cat # 14190-250, then – in some cases - 5ml of fresh DPBS was added to the plate and it was incubated at 37®C for 5 minutes. After incubation time individual colonies were picked by scrapping each with a sterile 200-uL pipet tip and transferred into a well of 48-well plate coated with matrigel.

2.2.5. Screening for template insert

After being transferred into 48-well plate, individual clones were cultured with mTeSR™1 media changed daily. When a colony reached 80-90% confluence it was expanded, by 1:3 split. When a clone was expanded to at least three confluent wells of 48-well plate, cells were harvested from two wells using Accutase, while the third one was subject to further expansion. The harvested cells were used to extract DNA using QIAamp DNA Mini Kit (Qiagen, Cat. no. 51306).

This DNA was then used for PCR screening with two sets of primers designed to amplify only template inserted at the right location in the genome. The PCR reaction was run using Taq DNA Polymerase with Standard Taq Buffer (New England BioLabs, Catalog # M0273X) with following thermocycling conditions:

- Initial Denaturation: 95°C for 30 seconds
- 30 Cycles:

    - 95°C for 30 seconds;
    - 58°C for 30 seconds;
    - 68°C for 60 seconds
- Final Extension: 68°C for 5 minutes
- Hold: 4°C

After PCR loading dye was added to the samples, they were loaded onto 1% agarose (VWR, Catalog Number: EM-2125) 1xTAE (ThermoFisher Scientific, Cat #15558-042) and run at 100 V for 40 minutes. The bands were then visualized under UV light.

2.2.6. Screening for wild type allele and further tests

DNA of the samples that screened positive for the presence of the template insert in the right place of the genome was used for screening for wild type allele on the other allele with one pair of primers. The PCR reaction was run using Taq DNA Polymerase with Standard Taq Buffer (New England BioLabs, Catalog # M0273X) with following thermocycling conditions:

- Initial Denaturation: 95°C for 30 seconds
- 30 Cycles:
    - 95°C for 30 seconds;
    - 51°C for 30 seconds;
    - 68°C for 60 seconds
- Final Extension: 68°C for 5 minutes
- Hold: 4°C

The PCR products from the clones that screened positive for the presence of wild type allele were sent to GENEWIZ laboratory in Boston for DNA Sanger DNA Sequencing service.

2.2.7. Activating GCM2 using SAM complex

For the purpose of gene activation using SAM complex, plasmids from addgene – a non-profit plasmid repository – were used, while specific guide RNAs were designed by the MGH lab.

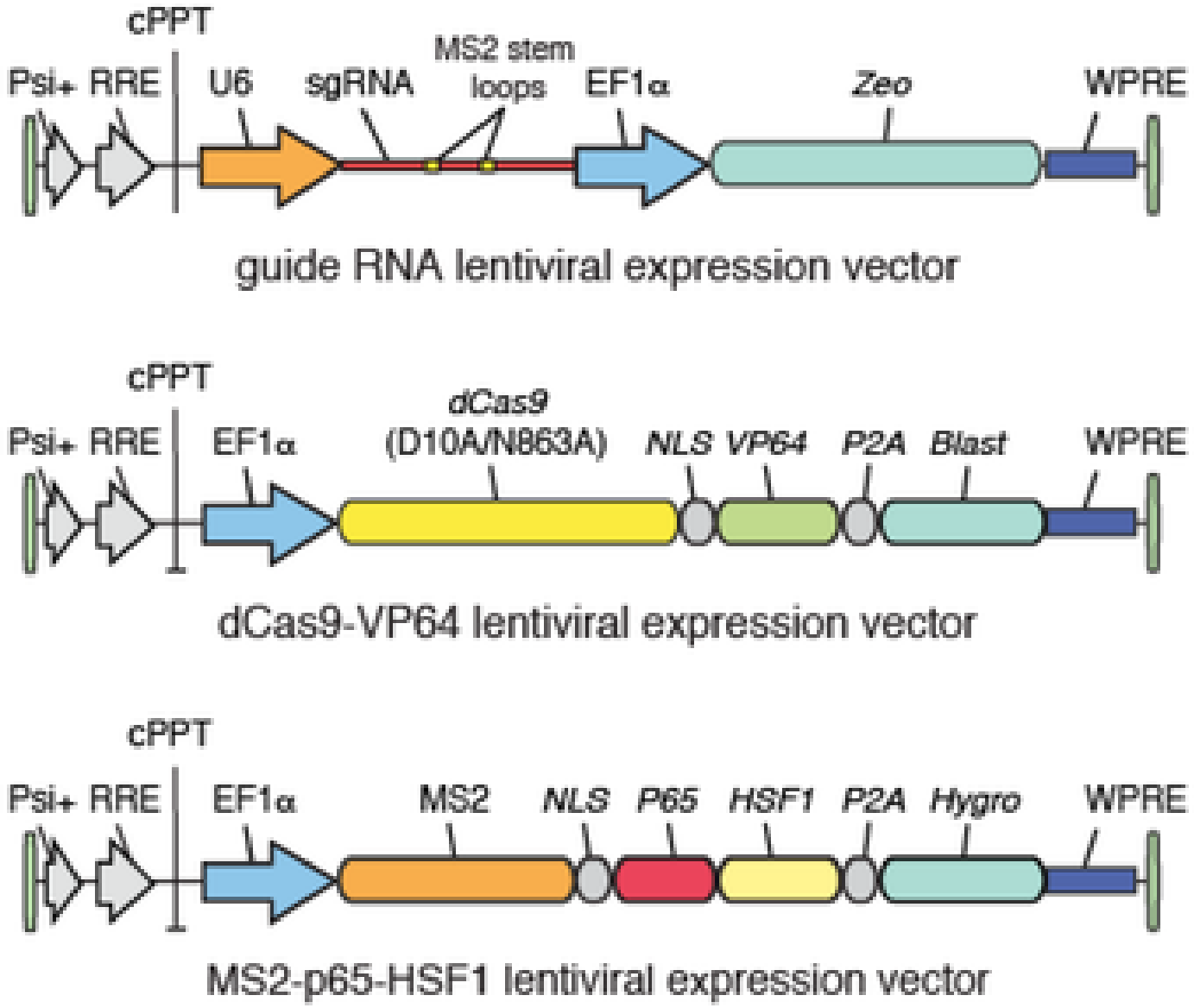


FIGURE 11 DESIGN OF THREE LENTIVIRAL VECTORS FOR EXPRESSING SGRNA, DCAS9-VP64, AND MS2-P65-HSF1. EACH VECTOR CONTAINS A DISTINCT SELECTION MARKER TO ENABLE CO-SELECTION OF CELLS EXPRESSING ALL THREE VECTORS.

2.2.7.1. Testing different guide RNAs in HEK 293 cells

Two days before transfection 250 000 HEK 293T cells per well were plated in 12 wells of two 6-well plate in DMEM/10% FBS without antibiotic. After 48h when they reached 70-80% confluency, the medium was changed. Twelve tubes containing 150 µl of opti-MEM each were prepared (Reduced Serum Media from ThermoFisher Scientific; Cat #31985070). Plasmids were added to the tubes as follow: nine tubes with each 0,7 µg of respective guide (SAM1-2, SAM4-10), one tube with no guide RNA, one with an empty vector from addgene and one with a RNA guide for an unrelated gene (MRKN3) (those three serve as negative controls). Then two other tubes were prepared: one with 63 µl of Lipofectamin 2000 (ThermoFisher, Cat #11668019) and 1885 µl of opti-MEM; second with 8,4

µg of Cas9-VP64, 8,4 µg of MS2-p65-HSF1 and 5,6 µl of Water (to dilute the solution to 1 µg of plasmid per µl). 1,4 ug of the plasmid master mix was then added to 11 out of 12 tubes (the last one serving as a plasmid-free control). The lipofectamin solution was incubated for 5 minutes at room temperature, after which 155 µl of diluted Lipofectamin 2000 was added to all 12 tubes. All tubes were then incubated at room temperature for 15 minutes after which 310 µl of solution from each tube was added to a respective well of a 6-well plate. After 24h, media was changed and after another 24h the cells were harvested using 0.05% trypsin. RNA was extracted from the harvested cells using RNeasy Mini Kit (Qiagen, Cat. no. 74106).

Complementary cDNA was synthesized from the RNA sample using qScript cDNA SuperMix from Quanta Biosciences (Cat No. 95048-500) following the manufacturer instructions with 1 µg of RNA used in the reaction.

The synthesized genomic cDNA was used to perform qPCR with the use of PerfeCTa® SYBR® Green SuperMix (Cat# 95054-500) following manufacturer instructions in ViiA™ 7 Real-Time PCR System from Applied Biosystems (Cat No. 4453536) with the following run set-up:

Hold stage: 95®C for 20 seconds

PCR Stage (reapeated 40 times): 95®C for 1 second and 60®C for 20 seconds.

Melt curve stage: 95®C for 15 seconds, 60®C for 1 minute, 95®C for 15 seconds.

Two set of primers were used for the qPCR: one set for actin, as an endogenous control, and another one for GCM2.

2.2.7.2 Activating GCM2 in gene reporter cell line

The gene reporter cell line was cultured on a 10 cm plate in a feeder-free condition with mTeSR™1 media. Upon reaching 70-80% confluence cells were dissociated using Accutase and counted using Bio-Rad Automated Cell Counter (Cat No. 1450102). 5 million cells, 1 million cells and 0,7 million cells were transferred to a respective 15ml conical tube and centrifuged at 1200 rpm for 4 minutes at room temperature. The 0,7 million cells were re-suspended in mTeSR™1 media and plated on a matrigel coated plate with addition of 20µM ROCK inhibitor to increase survival. Cells from other two tubes were washed twice with 2ml of DMEM/F12 medium.

During the washing step three tubes were prepared: one with 18,75 µl of Lipofectamine 3000 (ThermoFisher Scientific; Catalog #L3000015) Reagent in 625 µL Opti-MEM (Reduced Serum Media from ThermoFisher Scientific; Cat #31985070); second one with 5µg of Cas9-VP64 plasmid, 5µg of MS2-p65-HSF1 plasmid, 5 µg of guide RNA SAM 8, 625 µL of Opti-MEM and 25 µL of P300 Reagent (ThermoFisher Scientific; Cat #L3000-015); third one with 1µg of Cas9-VP64 plasmid, 1µg of MS2-p65-HSF1 plasmid, 1 µg of guide RNA for an unrelated gene (MRKN3), 125 µL of Opti-MEM and 5 µL of P300 Reagent. Then 125 µl of diluted Lipofectamine reagent mixture were added to both DNA containing tubes and the tubes were incubated at room temperature for 5 minutes.

After the incubation period and washing steps the DNA-lipid complex was added drop-wise to the cell pellet – the one with bigger volume to five million cells and the smaller one to one million cells - which was gently loosen by tapping the tube, then mixed gently and incubated at room temperature for 14 minutes. Then the cells were plated on a separate 10 cm plates coated with matrigel with mTeSR™1 media (STEMCELL Technologies; Catalog #05875). 20 µM of ROCK inhibitor was added to the plate

to increase the cell survival after transfection, but was not added to the subsequent feeds. After 48h the cells were harvested using Accutase and used for FACS.

2.2.7.3. Testing GCM2 expression in GFP positive and negative cells

Using FACS sorting, the cells transfected with the GCM2 activation SAM complex were sorted based on their GFP expression, while the non-transfected cells and cells transfected with unrelated guide were tested for the presence of GFP, but not sorted.

After FACS sorting, RNA was extracted from the GFP+ and GFP- cells, complementary cDNA was synthesized from the RNA samples, which was used for a qPCR with two sets of primers: one set for actin, as an endogenous control, and other one for GCM2. The same methods as mentioned before were used in those steps.

3. Results

3.1 Generation of induced pluripotent stem cells from fibroblast cells

3.1.1. Reprogramming by viral methods

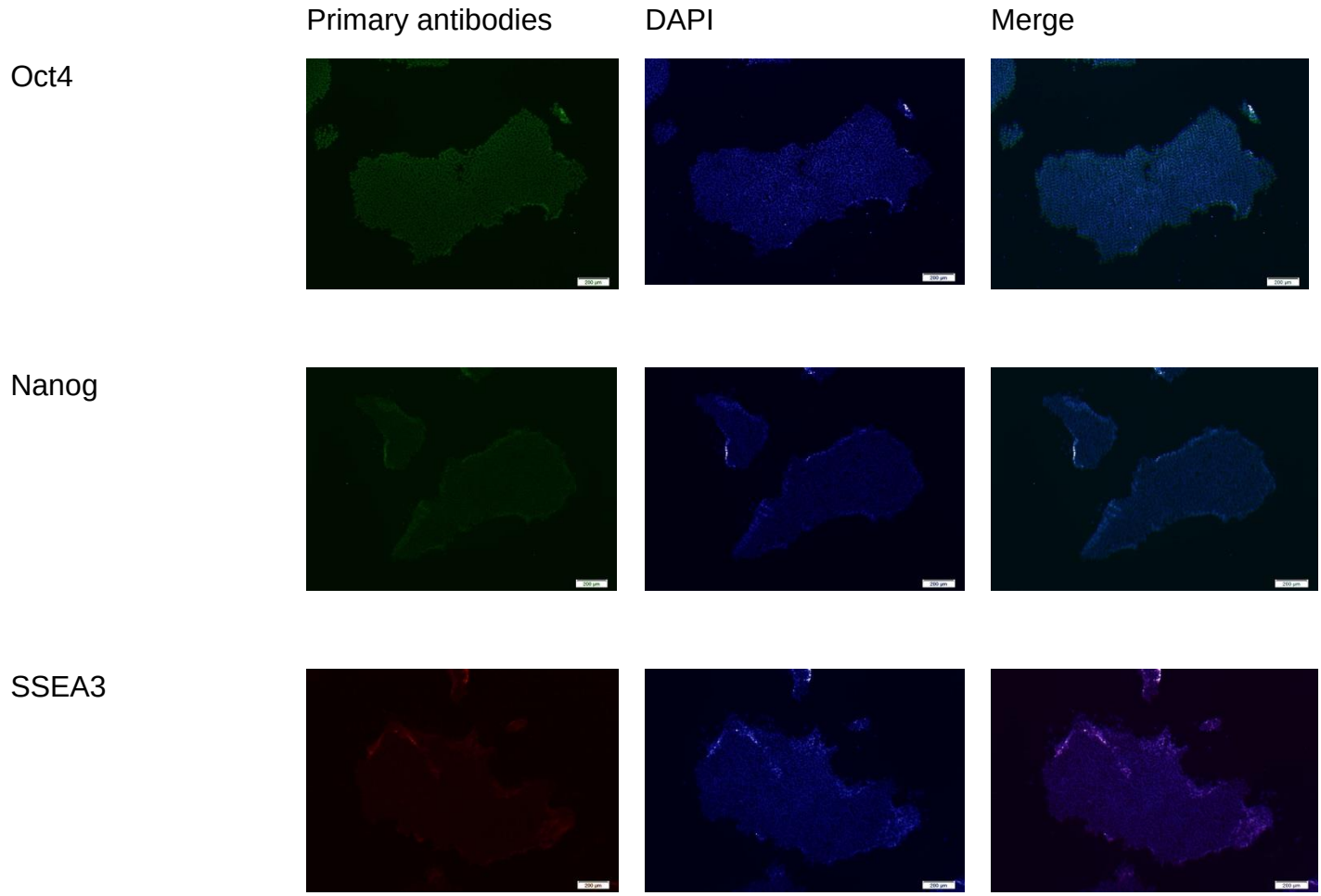
I derived two iPS cell lines from patient fibroblast: one using a retroviral reprogramming method and one using CytoTune®-iPS 2.0 Sendai Reprogramming Kit.

3.1.2. Cell line characterization

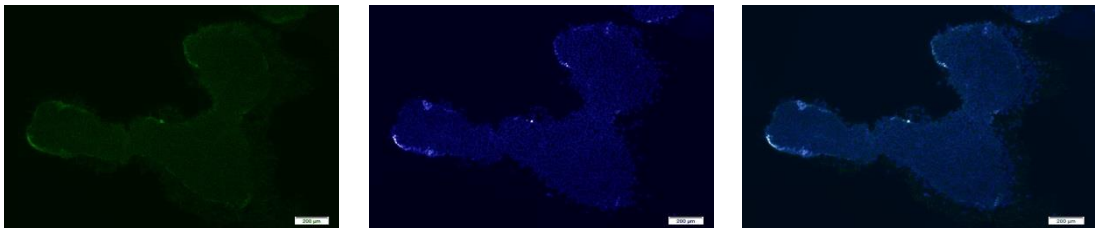
One line that was derived from patient fibroblast with the use of CytoTune®-iPS 2.0 Sendai Reprogramming Kit was characterized. This line will be reffered to as TL Sev-2. To characterize this line, we assessed the expression of pluripotent markers by immunocytochemistry and Q-RT-PCR. We differentiated the line into embryoid bodies for two weeks and checked the expression of differentiation markers for endoderm, mesoderm and ectoderm. We also sent the iPS line for karyotying.

3.1.2.1. Immunocytochemistry

The immunocytochemistry showed that TL Sev-2 line cells were positive for all five tested pluripotency markers (Figure 11).



SSEA4



TRA-1-60

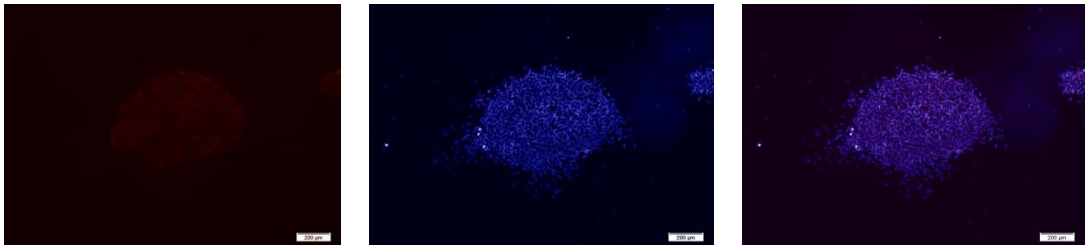


FIGURE 12 IMMUNOCYTOCHEMISTRY IMAGES DONE UNDER FLUORESCENT MICROSCOPE FOR CELLS FROM TL SEV-2 LINE.

3.1.2.2. qPCR

3.1.2.2.1. qPCR for pluripotency markers

For each set of primers (Actin, Oct4, Nanog, Sox2, hTERT, Rex1 and Actin) three separate wells were run.

A gene expression plot presented as relative quantification (comparative CT) [22] with positive control (Hues cell lines) acting as a reference sample and actin primers as endogenous control. The plot shows different levels of relative expression of pluripotency markers: some are expressed poorly (Rex1), some mildly (Sox2 and Nanog), while others highly (Oct4 and hTERT).This qPCR shows that the pluripotent markers are re-activated after reprogramming (Figure 12)

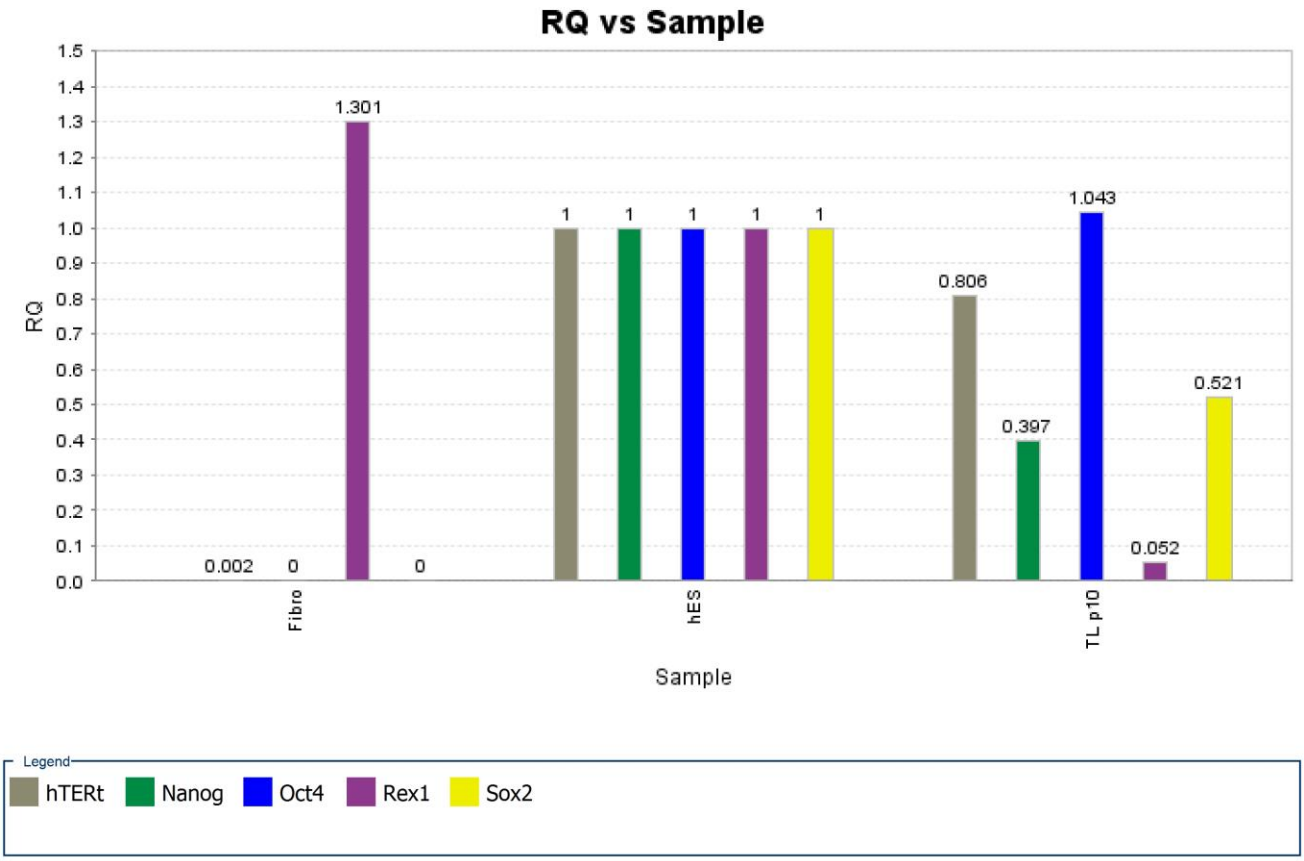


FIGURE 13 RELATIVE QUANTIFICATION OF GENE EXPRESSION PLOT FOR 6 PLURIPOTENCY MARKERS FOR POSITIVE CONTROL (HUES LINES), SAMPLE TL SEV-2 CELL LINE AND FIBROBLAST. GENERATED BY VIIA™ 7 REAL-TIME PCR SYSTEM, FROM APPLIED BIOSYSTEMS (CAT NO. 4453536), SOFTWARE, PROVIDED BY MANUFACTURER.

3.1.2.2.2. qPCR for differentiation markers

A gene expression plot presented as relative quantification (comparative CT) [22] with positive control (Hues line after EB differentiation) acting as a reference sample and actin primers as endogenous control. The plot shows different levels of relative expression of differentiation markers: The TL Sev-2 can differentiate into cells from the three germ layers. (Figure 13)

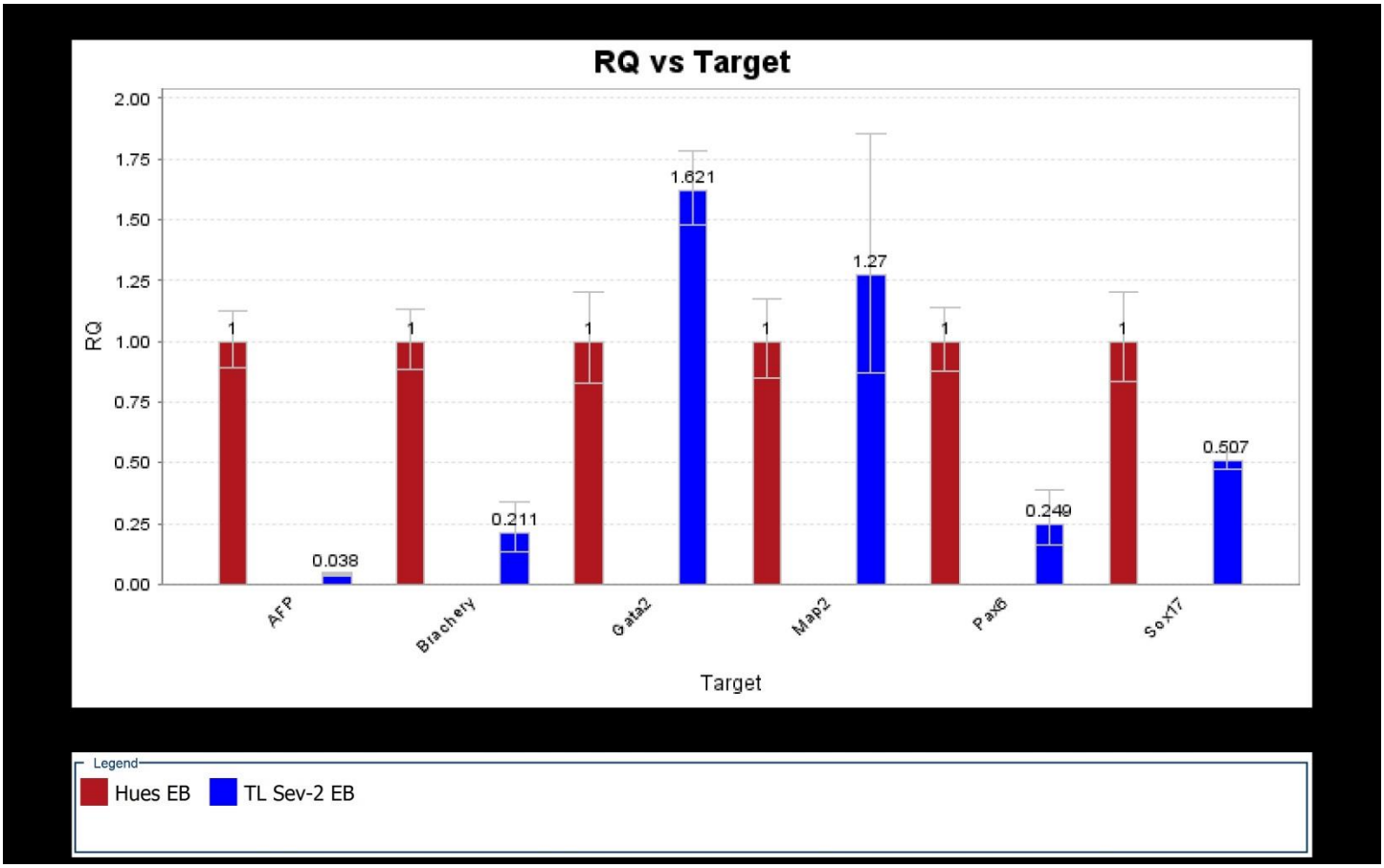


FIGURE 14 RELATIVE QUANTIFICATION OF GENE EXPRESSION PLOT FOR 6 DIFFERENTIATION MARKERS FOR POSITIVE CONTROL (HUES LINE AFTER EB DIFFERENTIATION) AND SAMPLE TL SEV-2 CELL LINE AFTER EB DIFFERENTIATION. GENERATED BY VIIA™ 7 REAL-TIME PCR SYSTEM, FROM APPLIED BIOSYSTEMS SOFTWARE, PROVIDED BY MANUFACTURER.

3.1.2.3. Karyotyping

Karyotyping of the TL SeV-2 iPS line at passage 13 was done by analyzing six metaphase cells and it was found to be a normal male karyotype (Figure 14)

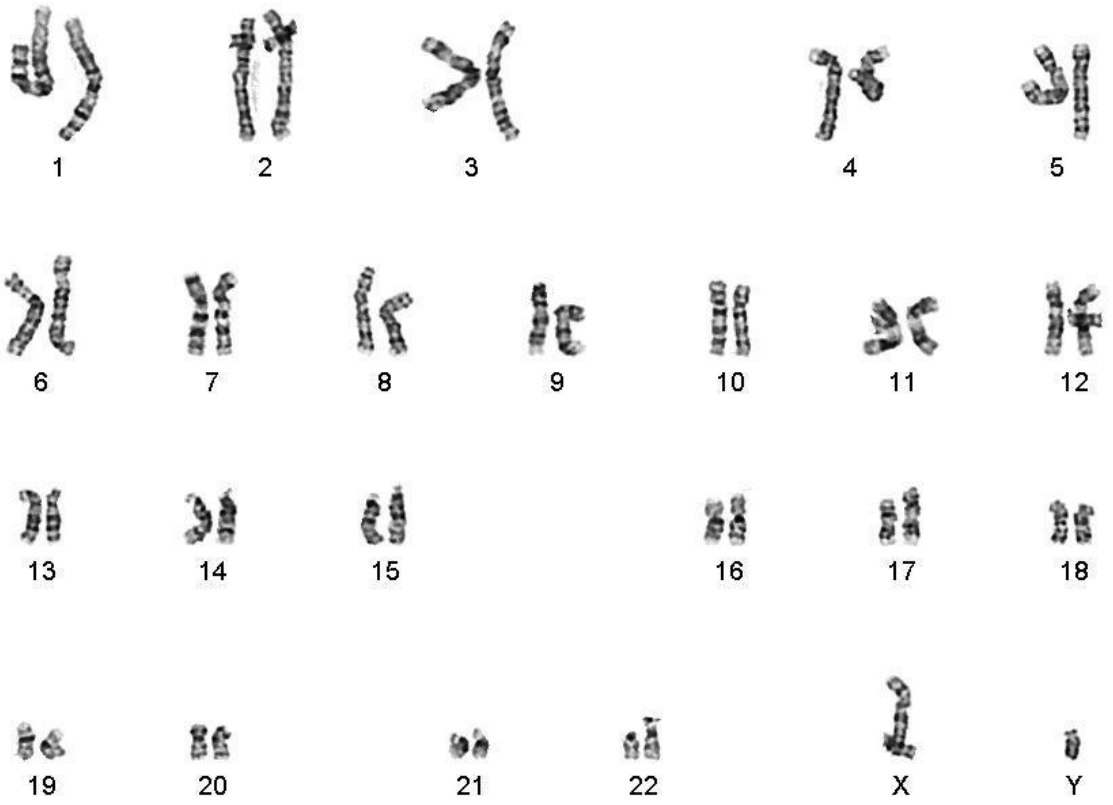


FIGURE 15 RESULTS OF CYTOGENETIC ANALYSIS OF PASSAGE 13 CELLS FROM TL SEV-2 CELL LINE DONE BY CYTOGENETICS LABORATORY OF TUFTS MEDICAL CENTER DEPARTMENT OF PATHOLOGY AND LABORATORY MEDICINE.

The characterization showed that the cell line has normal karyotype, expresses pluripotency markers and is able to differentiate into all three germ layers, establishing it as an induced pluripotent cell line available for research purposes. [1,14]

3.2 Gene editing in pluripotent stem cells

3.2.1. TIDE

We used the TIDE assay to measure the cutting efficiency for each RNA guide. TIDE assay showed the following efficiency: 31,2% for Guide 1, 45,4% for Guide 2 and 33,7% for Guide 3 (Figure 15, 16 and 17). Although Guide 2 is more efficient, we decided to test the three RNA guides in iPS line since they target sequences located at different position from the targeted GCM2 STOP codon.

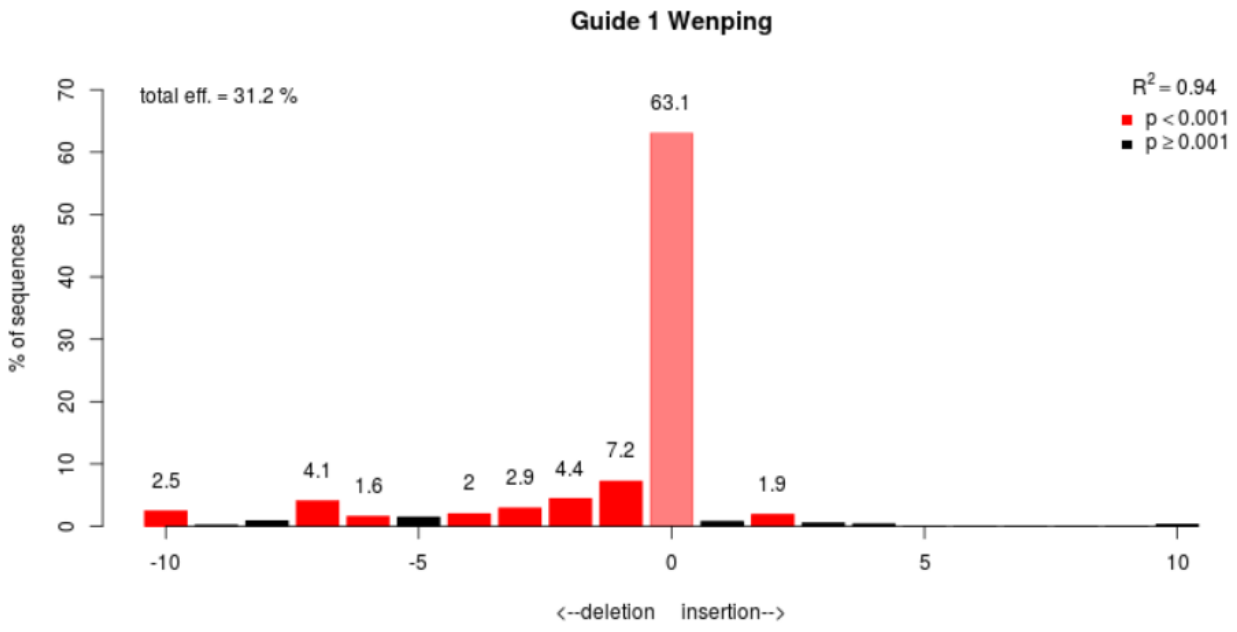


FIGURE 16 TIDE RESULTS FOR GUIDE 1

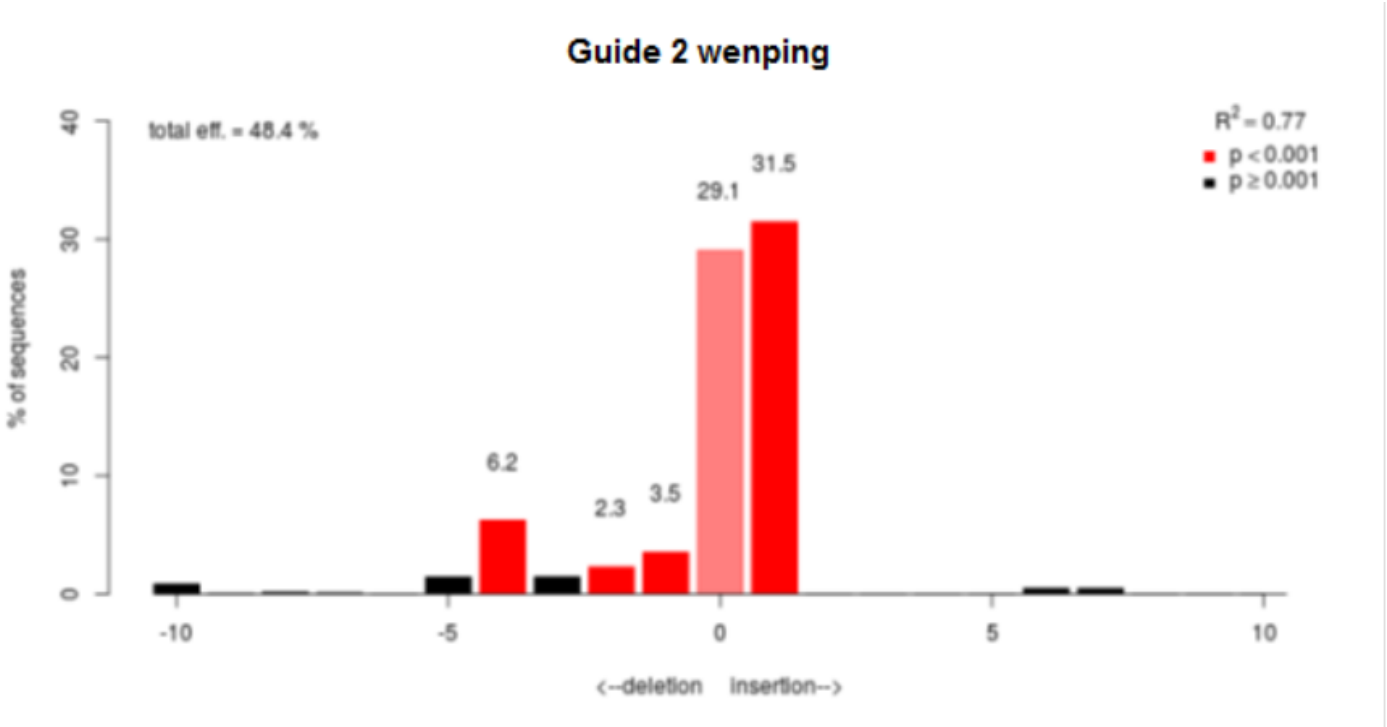


FIGURE 17 TIDE RESULTS FOR GUIDE 2

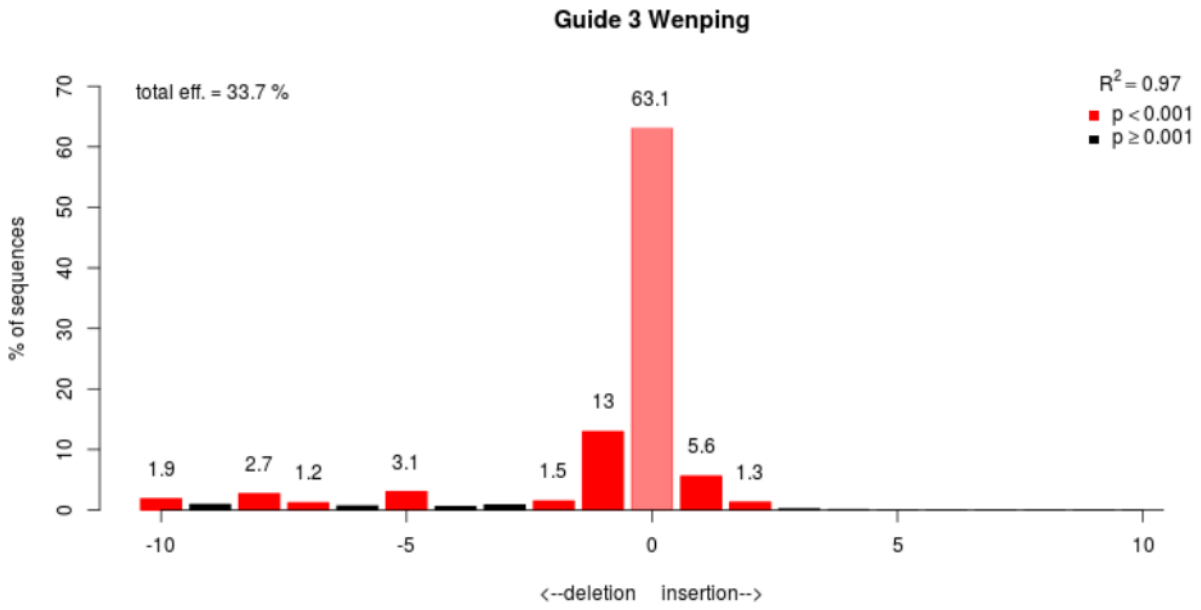


FIGURE 18 TIDE RESULTS FOR GUIDE 3

3.2.2. Kill curve

For the first kill curve five concentrations of puromycin were used: 0,5; 1; 1,5; 2 and 2,5 µg/ml. After 7 days, we observed 100% cell death in all 5 wells that were subjected to puromycin, while the cells in a control well were alive and well.

Therefore, we ran a second kill curve with lower concentrations: 0,1; 0,2; 0,3; 0,4 and 0,5 µg of puromycin per ml. After 7 days of treatment, we observed 100% death rate in the well with the highest concentration and some degree of living cells in the other wells. Hence, it was concluded that 0,5 µg/ml was the right concentration of puromycin for the selection process as it was the lowest concentration that resulted in 100% death rate among non-resistant cells.

3.2.3. Transfection and puromycin selection

Three separate transfections were carried out with different number of cells and different guide RNA. First one was done with three million cells and guide RNA number 1. Second and third one with one million cells and guide 2 or guide 3 respectively.

After the puromycin selection process, 7 clones were successfully picked from thefirst transfection; 13 clones were successfully picked from second transfection and 26 clones from the third.

3.2.4. Screening for template insert

Screening for the insertion of the reporter gene in the right location was done by PCR.It showed that 3 clones from first transfection were positive, 1 clone from second transfection and 9 clones from third transfection. In total, 13 clones from 46 picked were positive for template insert – that gives a relative targeting efficiency (targeted Clones/drug-resistant clones) of 28,2%. PCR results for negative and positive clones are shown in Figure 18

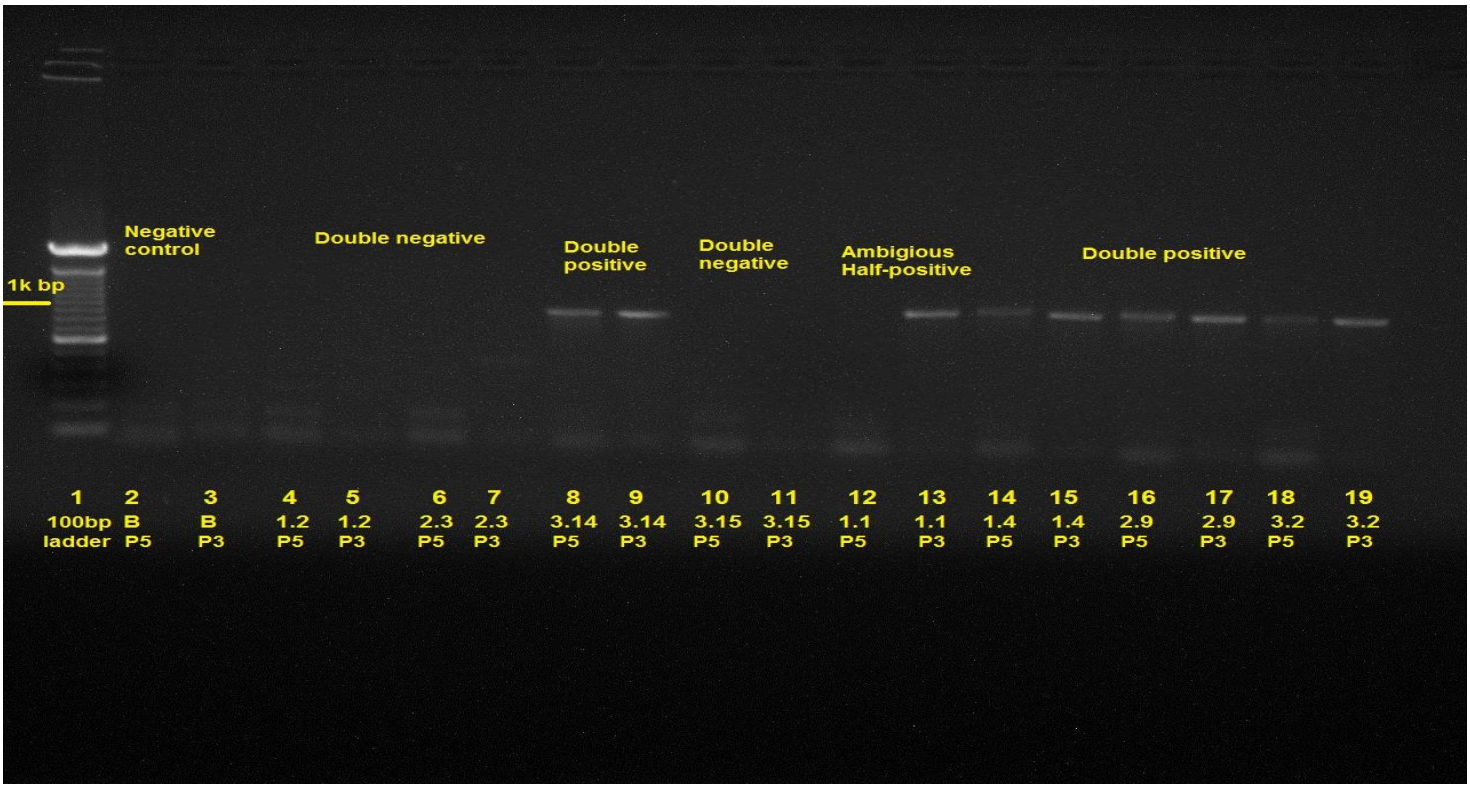


FIGURE 19 ONE OF MANY GEL IMAGES THAT SHOWS RANGE OF THE POSSIBLE OUTCOMES FOR VARIOUS CLONES.

3.2.5. Screening for wild type allele

Next, we wanted to check if the GCM2 gene on the second allele was intact or not in these 13 clones. We performed a PCR using primers located around the STOP codon of GCM2. 8 clones were positive for wild type gene and 5 were negative. (Figure 19)

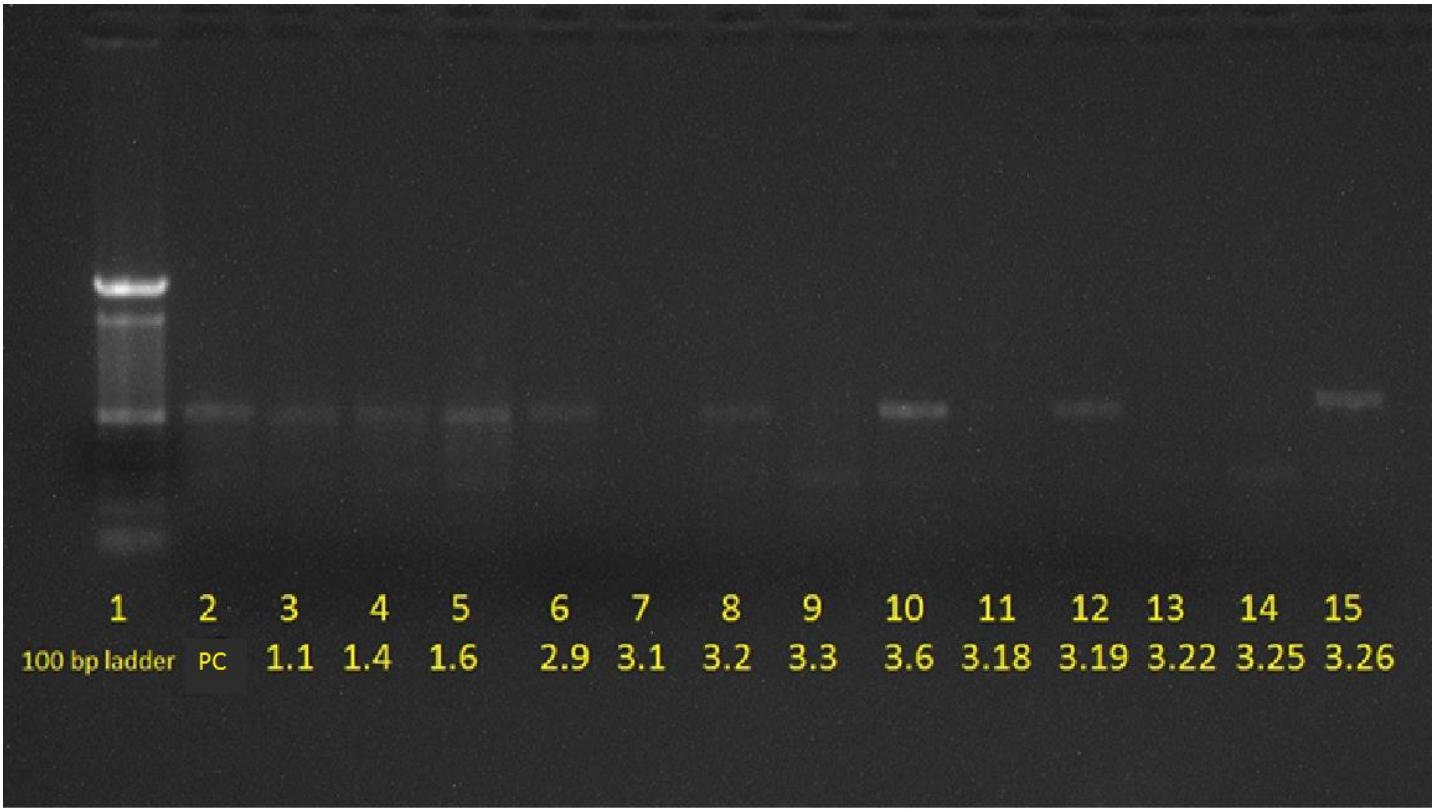


FIGURE 20 GEL IMAGE SHOWING THE RESULTS OF SCREENING FOR WILD TYPE GCM2. A BENT AT THE SAME HEIGHT AS POSITIVE CONTROL (PC) INDICATES A PRESENCE OF WT GENE

Next, the PCR products for the 8 positive clones were sent for sequencing to detect potential indels around the targeting site (Figure 20)

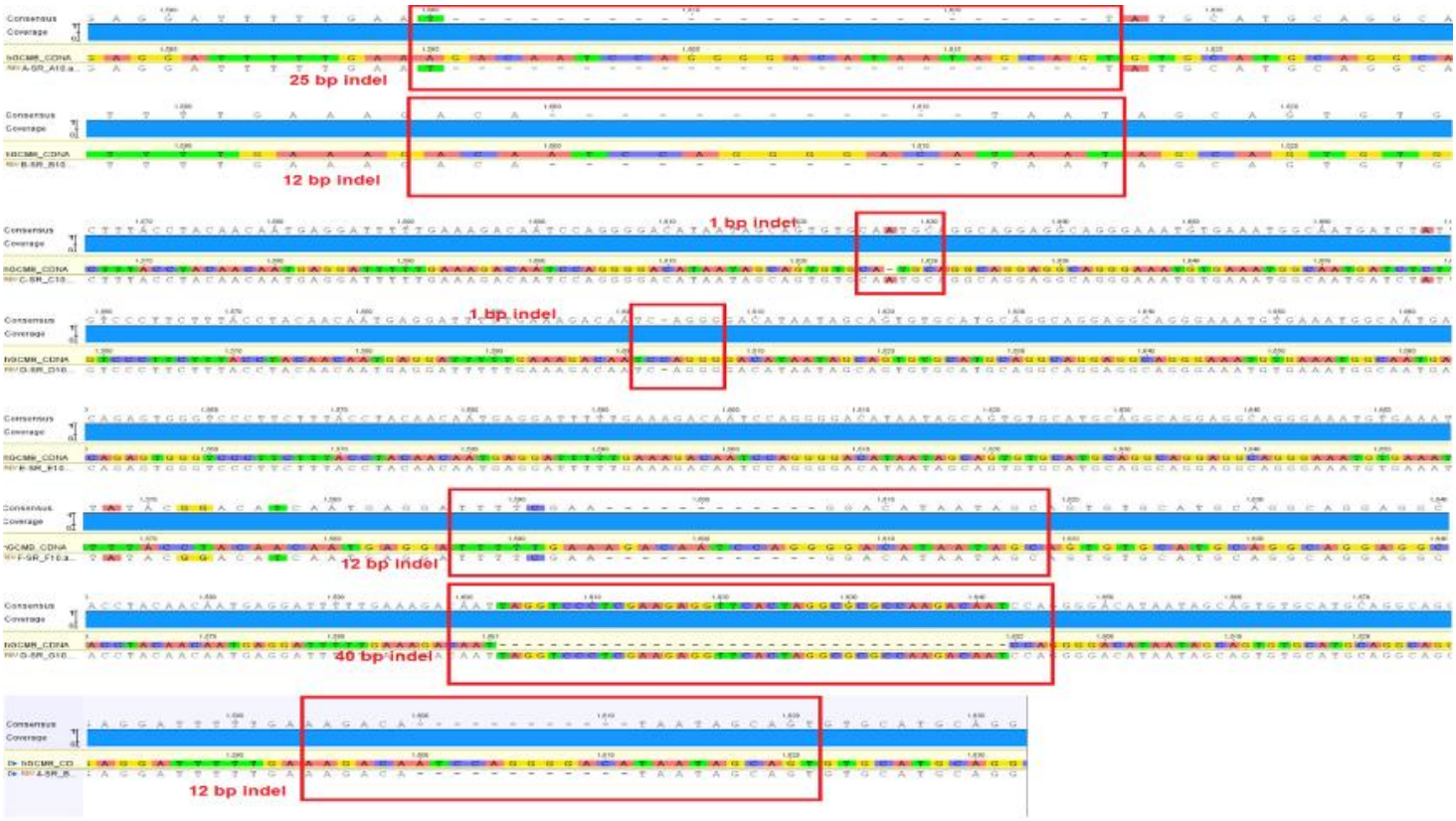


FIGURE 21 SEQUENCING RESULTS SHOWN IN GENEIOUS SOFTWARE AT THE GUIDE CUT LOCATION WITH INDEL MUTATIONS MARKED BY RED SQUARES

The sequencing showed that 7 of them had an indel mutation (ranging from 1 to 40 bp) near the cutting region. However, one clone – 3.6 – had no mutation at the untargeted allele. This clone was chosen as a "perfect" clone and expanded, because it had all the desired features. It was chosen as a gene reporter cell line and will be further referred as that.

For the "perfect" clone – 3.6 - a sequencing of targeted allele was also run and it showed a perfect integration of the insert (except a 1 bp insertion in the 3' arm) at the targeted allele. (Figure 21 and 22)

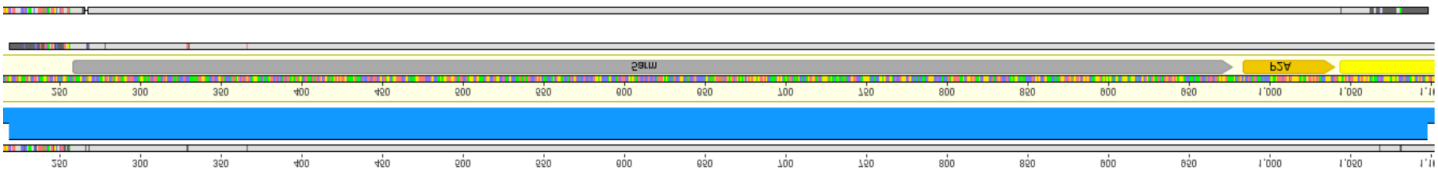


FIGURE 22 SEQUENCING RESULTS OF THE GENE REPORTER CELL LINE AT THE 5' ARM SHOWN IN GENEIOUS SOFTWARE

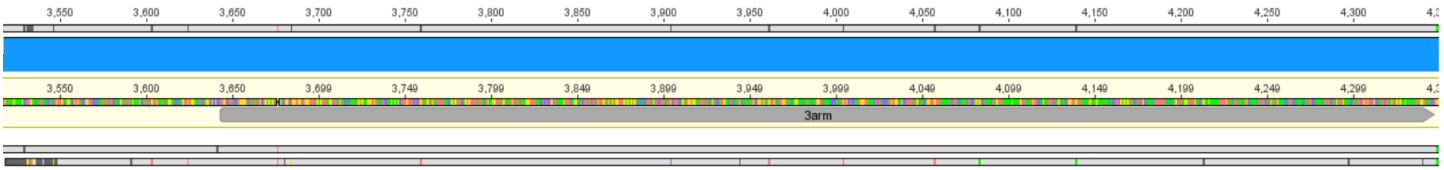


FIGURE 23 SEQUENCING RESULTS OF THE GENE REPORTER CELL LINE AT THE 3' ARM SHOWN IN GENEIOUS SOFTWARE

3.2.6. Confirming gene reporter function

Although clone 3.6 was confirmed to have the desired insertion of the GFP reporter by PCR and sequencing, it is important to verify if this reporter cell line is functional. In this case, GFP should be turned on in parathyroid cells. However, our collaborator at MGH did not have a protocol to derive parathyroid cells from iPS. We decided to use a different strategy to test the reporter line. We used the SAM transcriptional activation system to artificially activate the transcription of GCM2 in iPS lines and checked if GFP is expressed as well.

The testing of different SAM guides in HEK cells showed that the activation of GCM2 using SAM complex is possible even in cells that do not express the gene at all (such as HEK) and that guide 8, which is located 145 bp from TSS, has the most efficiency (Figure 24 and 25). It has to be noted that the expression level was extremely low relatively to the level of expression in parathyroid cells, which can be seen on figure 24. Three negative controls were used, however only one appears on the graph as other did not show any expression at all. The differences in expression level obtained with different guides show how important proper guide design is and why it was crucial to run an assay in HEK cells to choose the best one for testing the gene reporter line. We therefore used guide 8 for the next experiment.

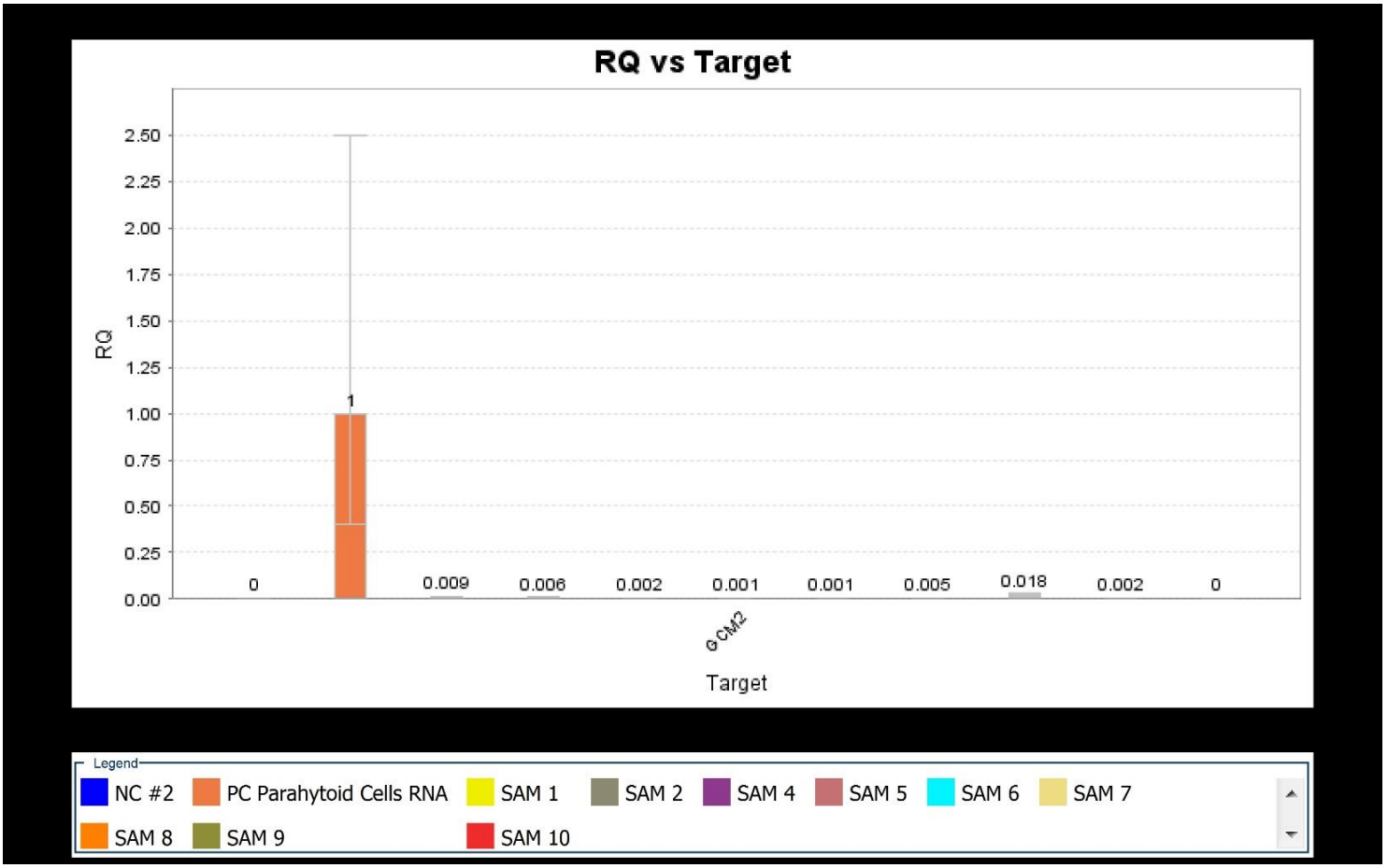


FIGURE 24 RELATIVE QUANTIFICATION OF GENE EXPRESSION PLOT FOR GCM2 IN DIFFERENT SAMPLES WITH PARATHYROID CELLS SERVING AS A REFERENCE SAMPLE. GENERATED BY VIIA™ 7 REAL-TIME PCR SYSTEM, FROM APPLIED BIOSYSTEMS SOFTWARE, PROVIDED BY MANUFACTURER.

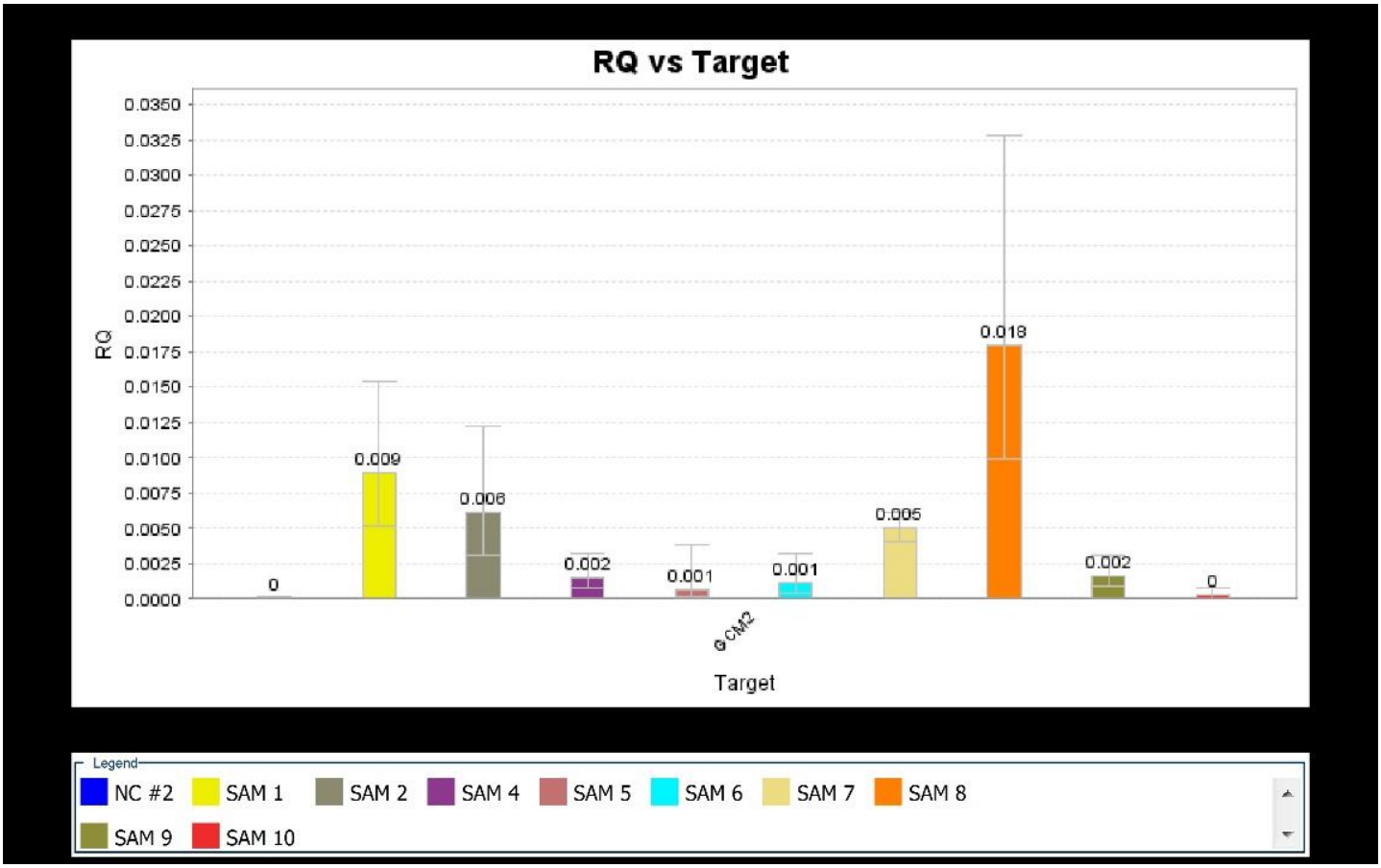
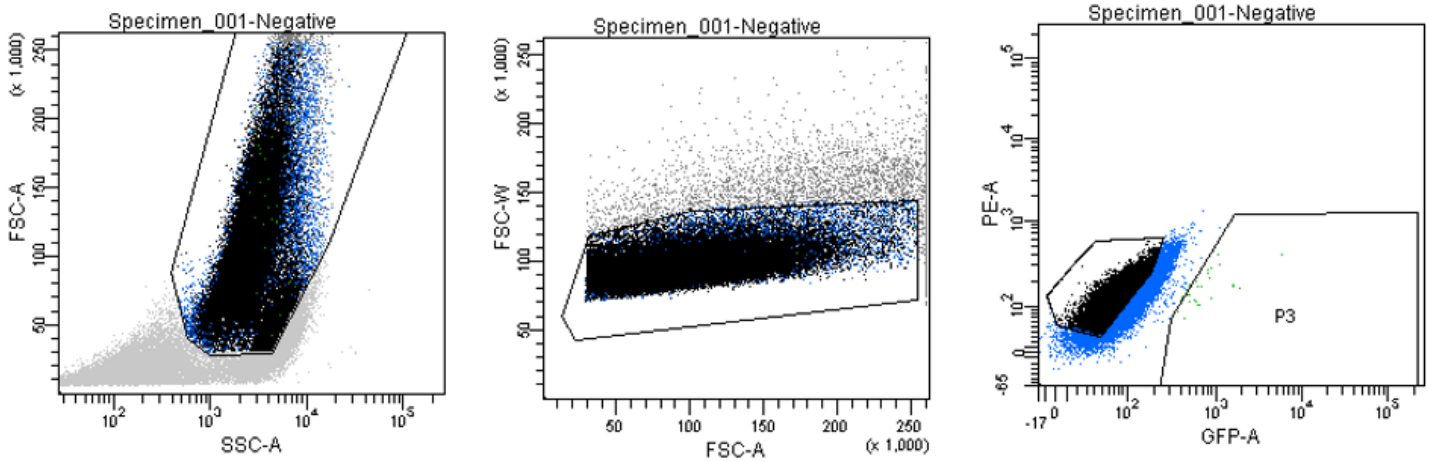


FIGURE 25 RELATIVE QUANTIFICATION OF GENE EXPRESSION PLOT FOR GCM2 IN DIFFERENT SAMPLES WITH PARATHYROID CELLS SERVING AS A REFERENCE SAMPLE, BUT NOT SHOWN ON THE GRAPH. GENERATED BY VIIA™ 7 REAL-TIME PCR SYSTEM, FROM APPLIED BIOSYSTEMS SOFTWARE, PROVIDED BY MANUFACTURER.

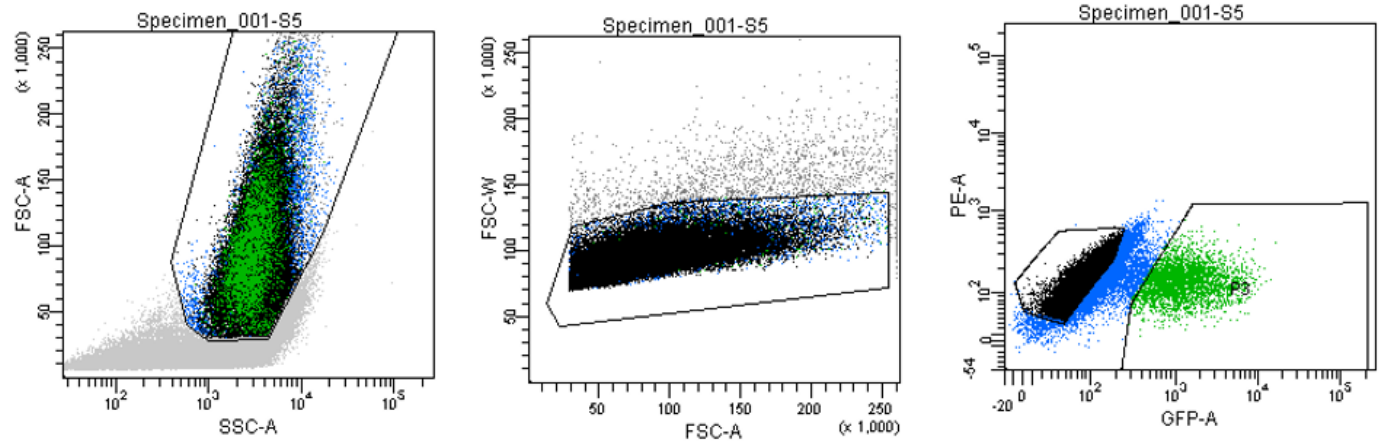
The gene reporter line (clone 3-6) was then transfected with the SAM system using the Guide 8 to activate GCM2 expression. As control, the reporter line was also transfected with the SAM system without guide or with an unrelated guide. 48 hours post-transfection, the cells were analyzed by FACS analysis. As shown in Figure 26, 27 and 28, GFP was undetectable for the two negative control but ~4% of the cells were GFP positive when transfected with guide 8, suggesting a good correlation between GCM2 and GFP expression.



Tube: Negative

Population	#Events	%Parent	%Total
All Events	100,000	####	100.0
P1	58,952	59.0	59.0
P2	54,326	92.2	54.3
P3	32	0.1	0.0
P4	38,586	71.0	38.6

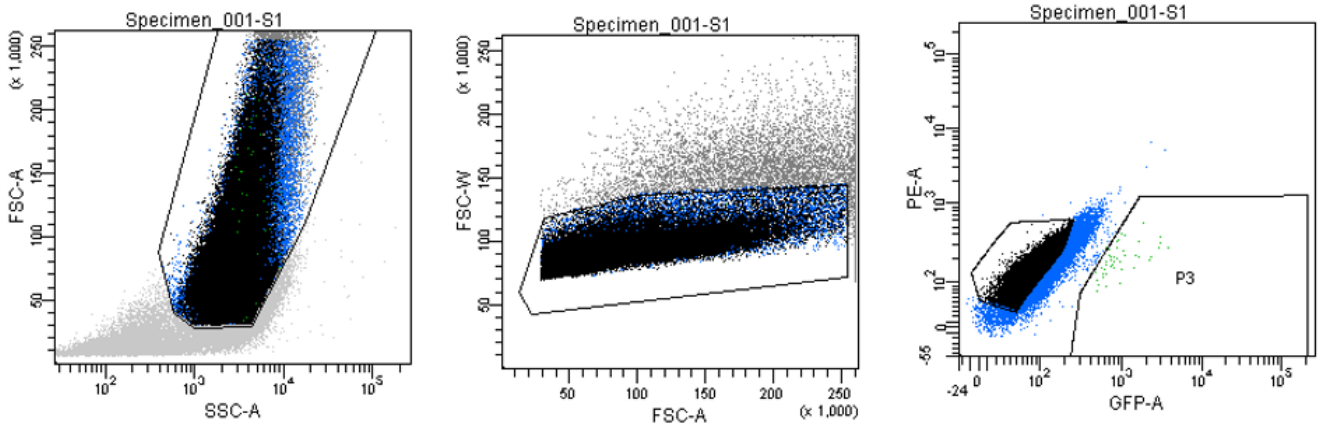
FIGURE 26 RESULTS OF FACS FOR NON-TRANSFECTED GENE REPORTER CELLS



Tube: S5

Population	#Events	%Parent	%Total
All Events	90,000	####	100.0
P1	45,909	51.0	51.0
P2	43,998	95.8	48.9
P3	3,514	8.0	3.9
P4	29,357	66.7	32.6

FIGURE 27 RESULTS OF FACS FOR GENE REPORTER CELLS TRANSFECTED WITH SAM 8 COMPLEX



Tube: S1

Population	#Events	%Parent	%Total
All Events	90,000	####	100.0
P1	68,934	76.6	76.6
P2	55,635	80.7	61.8
P3	50	0.1	0.1
P4	39,939	71.8	44.4

FIGURE 28 RESULTS OF FACS FOR GENE REPORTER CELLS TRANSFECTED WITH A GUIDE OF UNRELATED GENE

Next, we sorted the GFP+ and GFP- that were transfected with Guide 8 and extracted RNA to measure the expression of GCM2 in these 2 populations. qPCR of the sorted cells and the control showed a relatively high level of expression in GFP positive cells (5,8% of parathyroid cells expression) with no expression at all in GFP negative cells or control (Figure 29). These results shows clearly that in the gene reporter cell line the expression of GFP is strictly correlated with that of GCM2, which proves that the gene reporter cell line works as intended.

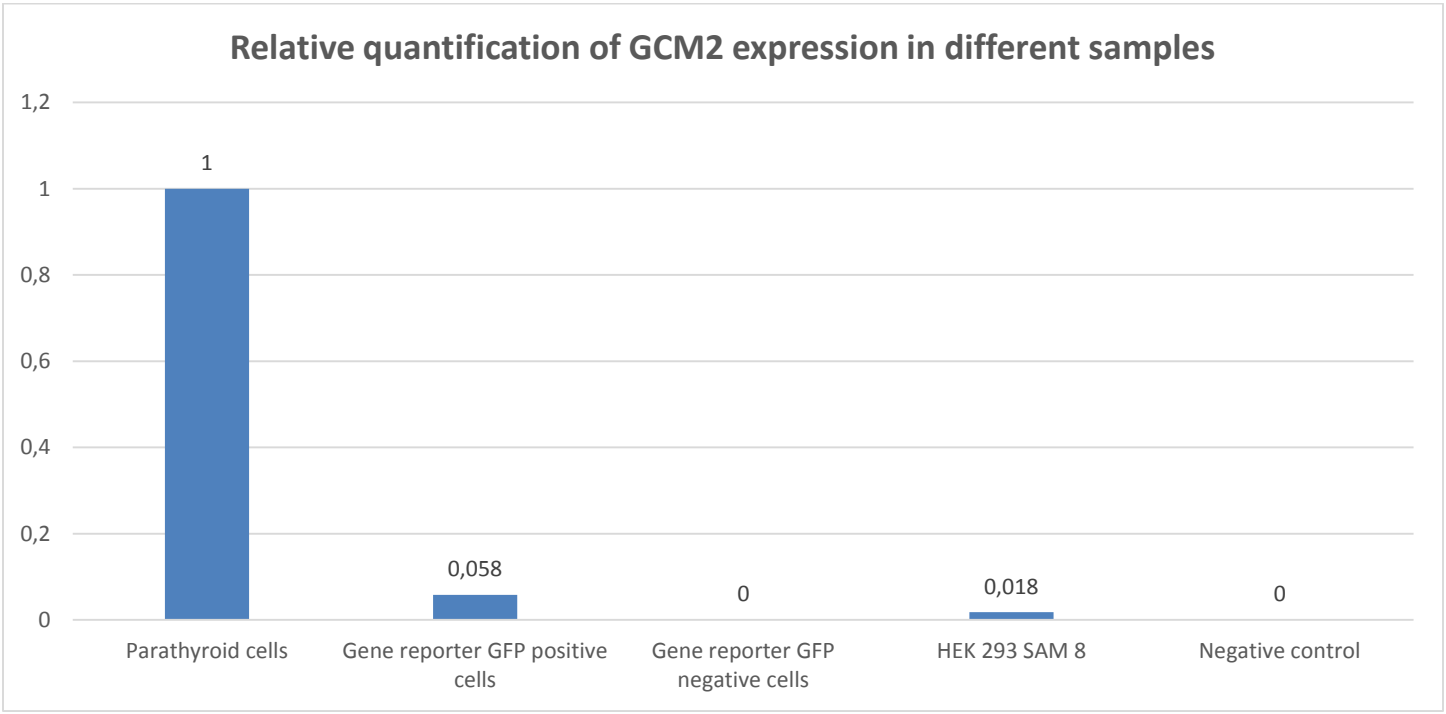


FIGURE 29 RELATIVE QUANTIFICATION PLOT OF GCM2 EXPRESSION IN DIFFERENT SAMPLES WITH PARATHYROID CELLS SERVING AS A REFERENCE. GENERATED IN EXCEL.

4. Discussion

4.1 Generation of induced pluripotent stem cells from fibroblast cells

4.1.1. Others method of reprogramming

Since the 2006 discovery of Takahashi and Yamanaka, many different approaches to generate iPS lines have been tested and established. In 2013, Emilie Bayart and Odile Cohen-Haguenauer published a review of the established methods for reprogramming. It listed 21 delivery methods of the reprogramming factors, with 10 different “cocktails” – combinations of transgenes and 14 different cell types, with variety of sub-types available [23].

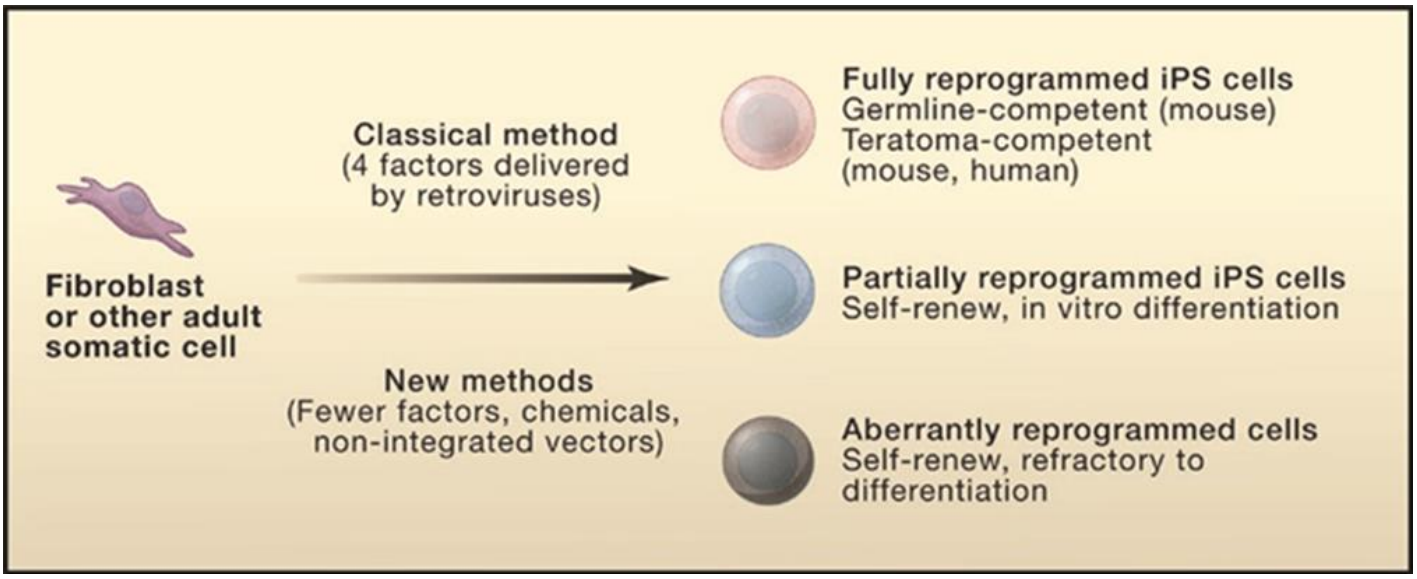


FIGURE 30 OLD AND NEW WAYS TO GENERATE IPS CELLS

This shows how potent and fast-moving the field of reprogramming is, with such a variety of methods being reported in just seven years. The field is still growing and progress is made on regular basis. To stay concise, I decided to restrict this paper to the combination of delivery methods, transcriptional factors cocktail and cell type that I used in my work.

4.1.2. Limitations

The utility of iPS cells is virtually endless, as more and more applications are found with the progress of research, however there are still severe limitations that have to be overcome. One of the most promising but challenging area is the use of cells derived from iPS for cell therapy. Definitely the first and most dreaded complication is the formation of teratomas, which are germ cell tumors containing several cell types. This is especially problematic in cells reprogrammed with the use of retrovirus or lentivirus, since the transgenes are integrated into the genome. Although the retroviruses are silenced in pluripotent stem cells, the transgenes could be reactivated during differentiation which could lead to tumorigenesis. [19].

Another important issue is the quality of each of the individual clones, and the verification that they are fully reprogrammed. An aberrant reprogramming can impact negatively the ability of the cell to differentiate and again increase the chance of tumor formation. A very thorough characterization of the iPS lines would be required to allow their use for clinical applications.

One of the most tragic examples of the use of retrovirus gene therapy that have back-fired is a case, where patients suffering from X-linked severe combined immunodeficiency, where treated with retroviruses, which led to activation of LMO2, a proto-oncogene, and caused leukemia as a result. This example shows how dangerous the use of retroviral derived iPS could be, knowing that each iPS cell clone can have up to 40 retroviral integration sites. [19]

To summarize, the possibility to derive patient and disease specific induced pluripotent stem cells using a robust and reliable method opened up many new avenues in the biomedical sciences. However, to reach the full potential of those cells, especially in the direct treatment of patient, work is still needed to generate robust cells in quantity and quality.

4.1.3. Interpretation of characterization results

The results of the karyotyping and immunocytochemistry are, in my opinion, straightforward and do not require further discussions. This cannot be said about the qPCR results.

4.1.3.1. qPCR for pluripotency markers

In the case of the qPCR for pluripotency markers there are markers that are expressed at much lower levels than in the positive control (Rex1) and slightly or mildly lower (Sox2, Nanog, hTERT) with one marker expressed at higher level in the sample than in the positive control (Oct4).

In my opinion these results still confirm the pluripotency of the cell line, since the reactivation of the markers can be clearly seen in the comparison to the fibroblast. Also Oct4 is one of the hallmarks of the pluripotency and it is the most crucial reprogramming factors, found in all the different reported reprogramming cocktails [4] and is expressed even stronger in the TL Sev-2 cells than in the positive control, which makes a strong case for the pluripotency of the line. The lower level of expression detected for the other genes compared to the hESC control could be due to some partial differentiation in the culture, to the difference in passage number for the iPS line and the hESC or the differences in culture conditions for the iPS and hESC at the time of collection for RNA extraction.

4.1.3.2. qPCR for differentiation markers

In the case of the qPCR for differentiation markers there is almost no expression of one marker (AFP), low relative expression levels of two (Brachyury and Pax6), mild expression of other one (Sox17) and very high relative expression of Map2 and Gata 2.

It is well documented that iPS lines vary in their differentiation potential and level of expression for differentiation markers is different for each line. To ensure that iPS lines are pluripotent, we checked two markers for each germ layer. In this case, at least one marker for each germ layer show high expression confirming that the line is pluripotent,

Both Map2 and Pax6 are ectoderm related markers, therefore a high level expression of one in combination with a low level of the second one indicates that the line should be able to differentiate into ectoderm cells. Sox17 is an endoderm marker and its mild expression might indicate some difficulties with differentiation of the cells in that direction. Gata2 is associated with hematopoietic progenitors and its very high level of expression indicates that the line have predispositions to differentiate in that direction. Last, but not least, Brachyury is a marker associated with mesoderm and its relatively low level of expression might indicate possible difficulties with differentiating the cells in that direction. [34, 35]

It is important to keep in mind that the embryoid body assay is a random differentiaton assay. This iPS line might differentiate properly into endoderm cells if protocol for endoderm differentiation were used.

All in all, I think that the different levels of expression of markers should be taken into account when using this cell line for differentiation and can be used as an advantage if used properly (for instance use this line for blood cell differentiation)).

4.2. Gene editing in pluripotent stem cells

4.2.1. Identity of ambivalent clones

There are some clones whose “identity” as to what is really present on their GCM2 allele remains unknown to me. These are the clones that screened positive for one arm of the template insert, but negative for the other and also clones that screened negative for wild type gene. To identify them was not important since the goal was to obtain at least one clone with the right targeting . What I think might be the result of this unknowns, are cuts of the Cas9 that distorted the alleles in a ways that the methods I used were unable to show.

4.2.2. Possibly useful clones

After sequencing of the wild type allele one clone was chosen because of no mutations at all, however other clones have only relatively small indel mutations in the non-coding region of the gene, which indicates that there is a possibility that those mutations do not have an impact on the expression of the gene, so they would still be very much useful as a gene reporter cell line. Still, similar as above, the topic was not further pursued, because the “perfect” clone was already identified. However, those clones remain as a viable alternative in case of some unforeseen difficulties.

4.2.3. Conclusions from GCM2 activation

The activation of GCM2 using SAM complex showed that the gene reporter line works as expected, therefore sorting of cells using FACS during differentiation into parathyroid cells should be possible. Interestingly, even at the low level of expression obtained with the SAM system, which was only 5,8% of the level of expression in differentiated parathyroid cells, I was able to see GFP under fluorescence microscope with 20x magnification. This would suggest that when the cells are differentiated, the GFP expression should be stronger and be clearly visible under a fluorescence microscopy,. This would facilitate measuring the efficiency of new protocols for parathyroid cell differentiation just by estimating the number of GFP+ cells under the microscope.. Also, since FACS sorting and observing GFP under microscope is possible even at 6% of expression of parathyroid cells it should be possible to sort cells that are not even fully differentiated, which might be helpful during development of differentiation protocol. All in all, I think that this is a cell line that is very much appropriate to fulfill the function it was engineered for.

References:

1. Park IH, Arora N, Huo H, et al. Disease-specific induced pluripotent stem cells. Cell 2008; 134: 877–886.
2. John Ydstie and Joe Palca, "Embryonic Stem Cells Made Without Embryos," *NPR*, November 21, 2007.
3. Takahashi K, Yamanaka S. Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined
4. Kulcenty K, Wróblewska J, et al. Molecular mechanisms of induced pluripotency. Contemp Oncol (Pozn). 2015;19(1A):A22-9. doi: 10.5114/wo.2014.47134.
5. Buganim Y, Faddah DA, Cheng AW, Itskovich E, Markoulaki S, Ganz K, Klemm SL, van Oudenaarden A, Jaenisch R. 2012. Single-cell expression analyses during cellular reprogramming reveal an early stochastic and a late hierarchic phase. Cell 150: 1209–1222.
6. Kurth, Reinhard; Bannert, Norbert, eds. (2010). Retroviruses: Molecular Biology, Genomics and Pathogenesis. Horizon Scientific. ISBN 978-1-904455-55-4.
7. Coffin, John M. (1992). "Structure and Classification of Retroviruses". In Levy, Jay A. The Retroviridae 1 (1st ed.). New York: Plenum. p. 20. ISBN 0-306-44074-1.
8. "Transfection". Protocols and Applications Guide. Promega.
9. Jones, Elizabeth; Hartl, Daniel L. (1998). Genetics: principles and analysis. Boston: Jones and Bartlett Publishers. ISBN 0-7637-0489-X.
10. He TC, Zhou S, da Costa LT, Yu J, Kinzler KW, Vogelstein B (March 1998). "A simplified system for generating recombinant adenoviruses". Proc. Natl. Acad. Sci. U.S.A. 95 (5): 2509–14. doi:10.1073/pnas.95.5.2509. PMC 19394. PMID 9482916.
11. Graham FL, et al. Defective transforming capacity of adenovirus type 5 host-range mutants. Virology 86: 10-21, 1978. PubMed: 664220
12. "Fibroblast". Genetics Home Reference. U.S. National Library of Medicine. 2014-05-05. Retrieved 2014-05-10.
13. Life Technologies Corporation, CytoTune-iPS 2.0 Sendai Reprogramming Kit User Guide. 2013, Publication Number: MAN0009378.
14. The International Stem Cell Initiative, Characterization of human embryonic stem cell lines by the International Stem Cell Initiative, Nature Biotechnology 25, 803 - 816 (2007), doi: 10.1038/nbt1318.
15. Immunocytochemistry Methods and Protocols. Edited by Lorette C. Javois, 2nd edition, 1999. Human Press
16. Joseph Sambrook and David W. Russel (2001). Molecular Cloning: A Laboratory Manual (3rd ed.). Cold Spring Harbor, N.Y.: Cold Spring Harbor Laboratory Press. ISBN 0-879-69576-5. Chapter 8: In vitro Amplification of DNA by the Polymerase Chain Reaction
17. Ponchel F, Toomes C, Bransfield K, Leong F.T, Douglas S.H, Field S.L, Bell S.M, Combaret V, Puisieux A, Mighell A.J (2003). "Real-time PCR based on SYBR-Green I fluorescence: An alternative to the TaqMan assay for a relative quantification of gene rearrangements, gene amplifications and micro gene deletions" (w). BMC Biotechnol 3: 18. doi:10.1186/1472-6750-3-18.
18. White, M.J.D. (1973). The chromosomes (6th ed.). London: Chapman & Hall.
19. Yamanaka S. A fresh look at iPS cells. Cell. 2009;137(1):13–7. doi: 10.1016/j.cell.2009.03.034.
20. Park IH, Lerou PH, Zhao R, Huo H, Daley GQ. Generation of human-induced pluripotent stem cells. Nat Protoc. 2008;3:1180–6.

21. Huangfu D. et al., 2008a. Induction of pluripotent stem cells by defined factors is greatly improved by small-molecule compounds. Nat Biotechnol. 26(7):795-7.
22. Thomas D Schmittgen & Kenneth J Livak, Analyzing real-time PCR data by the comparative CT method, Nature Protocols 3, - 1101 - 1108 (2008), Published online: 5 June 2008 | doi:10.1038/nprot.2008.73
23. Bayart E, Cohen-Haguenauer O. Technological overview of iPS induction from human adult somatic cells. Curr Gene Ther. 2013;13(2):73–92. doi: 10.2174/1566523211313020002.
24. Koo, J.; Kim, Y.; Kim, J.; Yeom, M.; Lee, I. C.; Nam, H. G. (2007). "A GUS/Luciferase Fusion Reporter for Plant Gene Trapping and for Assay of Promoter Activity with Luciferin-Dependent Control of the Reporter Protein Stability". Plant and Cell Physiology 48 (8): 1121–1131. doi:10.1093/pcp/pcm081. PMID 17597079
25. Grigorieva, I. V. and Thakker, R. V. (2011), Transcription factors in parathyroid development: lessons from hypoparathyroid disorders. Annals of the New York Academy of Sciences, 1237: 24–38. doi: 10.1111/j.1749-6632.2011.06221.x
26. Kim JH, Lee S-R, Li L-H, Park H-J, Park J-H, Lee KY, et al. (2011) High Cleavage Efficiency of a 2A Peptide Derived from Porcine Teschovirus-1 in Human Cell Lines, Zebrafish and Mice. PLoS ONE 6(4): e18556. doi:10.1371/journal.pone.0018556
27. Phillips GJ (Oct 2001). "Green fluorescent protein--a bright idea for the study of bacterial protein localization". FEMS Microbiology Letters 204 (1): 9–18. doi:10.1016/S0378-1097(01)00358-5. PMID 11682170
28. Cormack BP, Valdivia RH, Falkow S (1996). "FACS-optimized mutants of the green fluorescent protein (GFP)". Gene 173 (1 Spec No): 33–38. doi:10.1016/0378-1119(95)00685-0. PMID 8707053.
29. Pestka, S. (1971). "Inhibitors of ribosome functions". Annu. Rev. Microbiol. 25: 487–562. doi:10.1146/annurev.mi.25.100171.002415. PMID 4949424.
30. MacDonald C, Piper RC (2015). "Puromycin- and methotrexate-resistance cassettes and optimized Cre-recombinase expression plasmids for use in yeast". Yeast 32 (5): 423–38. doi:10.1002/yea.3069. PMID 25688547.
31. Hsu PD, Lander ES, Zhang F; Development and applications of CRISPR-Cas9 for genome engineering. Cell. 2014 Jun 5;157(6):1262-78. doi: 10.1016/j.cell.2014.05.010.
32. Eva K. Brinkman, Tao Chen, Mario Amendola and Bas van Steensel; Easy quantitative assessment of genome editing by sequence trace decomposition; Nucleic Acids Research, 2014, Vol. 42, No. 22 e168, doi: 10.1093/nar/gku936
33. Konermann, S., Brigham, M. D., Trevino, A. E., Joung, J., Abudayyeh, O. O., Barcena, C., … Zhang, F. (2015). Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature, 517(7536), 583–588. http://doi.org/10.1038/nature14136
34. Tsankov AM, et al. A qPCR ScoreCard quantifies the differentiation potential of human pluripotent stem cells. Nat Biotechnol. 2015 Nov;33(11):1182-92. doi: 10.1038/nbt.3387. Epub 2015 Oct 26.
35. Bock, C., Kiskinis, E., Verstappen, G., Gu, H., Boulting, G., Smith, Z.D., Ziller, M., Croft, G.F., Amoroso, M.W., Oakley, D.H., et al. (2011). Reference Maps of human ES and iPS cell variation enable high-throughput characterization of pluripotent cell lines. Cell 144, 439–452.

Figure references:

Figure 1: Kulcenty K, Wróblewska J, et al. Molecular mechanisms of induced pluripotency. Contemp Oncol (Pozn). 2015;19(1A):A22-9. doi: 10.5114/wo.2014.47134.

Figure 2: Eric P O'Keefe; Nucleic Acid Delivery: Lentiviral and Retroviral Vectors; MATER METHODS 2013;3:174; doi: http://dx.doi.org/10.13070/mm.en.3.174

Figure 3: Jonathan P. Stoye, Studies of endogenous retroviruses reveal a continuing evolutionary saga; Nature Reviews Microbiology 10, 395-406 (June 2012), doi:10.1038/nrmicro2783.

Figure 4: Life Technologies Corporation, CytoTune-iPS 2.0 Sendai Reprogramming Kit User Guide. 2013, Publication Number: MAN0009378.

Figure 5: Yamanaka S. A fresh look at iPS cells. Cell. 2009;137(1):13–7. doi: 10.1016/j.cell.2009.03.034.

Figure 6: Merkle et al., Efficient CRISPR-Cas9-mediated generation of knockin human pluripotent stem cells lacking undesired mutations at the targeted locus; Cell Rep. 2015 May 12;11(6):875-83. doi: 10.1016/j.celrep.2015.04.007. Epub 2015 Apr 30.

Figure 7: Hsu PD, Lander ES, Zhang F; Development and applications of CRISPR-Cas9 for genome engineering. Cell. 2014 Jun 5;157(6):1262-78. doi: 10.1016/j.cell.2014.05.010.

Figure 8: Eva K. Brinkman, Tao Chen, Mario Amendola and Bas van Steensel; Easy quantitative assessment of genome editing by sequence trace decomposition; Nucleic Acids Research, 2014, Vol. 42, No. 22 e168, doi: 10.1093/nar/gku936

Figure 10: Konermann, S., Brigham, M. D., Trevino, A. E., Joung, J., Abudayyeh, O. O., Barcena, C., … Zhang, F. (2015). Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature, 517(7536), 583–588. http://doi.org/10.1038/nature14136

Figure 11: Konermann, S., Brigham, M. D., Trevino, A. E., Joung, J., Abudayyeh, O. O., Barcena, C., … Zhang, F. (2015). Genome-scale transcriptional activation by an engineered CRISPR-Cas9 complex. Nature, 517(7536), 583–588. http://doi.org/10.1038/nature14136

Figure 12: Yamanaka S. A fresh look at iPS cells. Cell. 2009;137(1):13–7. doi: 10.1016/j.cell.2009.03.034.