PROTOCOL OPTIMIZATION OF THE GENERATION OF GENE REPORTER LINES

FINAL RESEARCH PAPER

MASTER THESIS

submitted at the
IMC Fachhochschule Krems
(University of Applied Sciences)

Lisa DORNER

Table of Contents

List of Figures

List of Tables

List of Abbreviations

AICS	Allen Institute for Cell Science
ATCC	American Type Culture Collection
BIR	breakage-induced replication
bp	base pairs
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
ddPCR	Droplet digital PCR
DLX5	distal-less homeobox 5 gene
DSB	double strand break
dsDNA	double stranded DNA
ESC	embryonic stem cells
GFP	green fluorescent protein
GOI	genes of interest
gRNA	guide RNA
HA	Homology arms
HDR	Homology-directed repair
ICM	inner cell mass
INDELs	insertions and deletions
iPSC	induced pluripotent Stem cell
IVT	*in vitro* transcription
NHEJ	non-homologous end-joining
nt	nucleotide
PAM	protospacer adjacent motif
PCR	Polymerase Chain reaction
PSC	Pluripotent stem cells

RNP	Ribonucleoprotein complex
rpm	revolutions per minute
RT	reverse transcription
sgRNA	single chimeric geode RNA
ssDNA	Single stranded DNA
ssODN	single stranded oligo donor
TALEN	Transcription Activator-like Effector Nuclease
WHP	Woodchuck Hepatitis Virus
WPRE	Posttranscriptional Regulatory Element
ZFN	Zinc-Finger Nucleases

1 Introduction

In 1961, two scientists observed that in irradiated mice the intravenous injection of bone marrow cells led to the formation of blood-forming progenitor cells. With this finding they were able to open the clinical use of bone marrow transplantation for hematopoietic disorders (Till & Mc, 1961). Due to the wide range of applications, stem cell research reached extreme importance (Butler et al., 2014). It can help to explain how some functions of the body work, and how they sometimes go wrong. In recent years, some major discoveries have been made in pluripotent stem cells (PSC). PSC are undifferentiated cells which are able to differentiate into any type of cell and have the potential to self-renew (Wobus & Boheler, 2005). PSC can be separated into two types, the embryonic stem cells (ESC) and the induced pluripotent stem cells (iPSC). ESC are taken from the inner cell mass (ICM) of a blastocyst, that is a 4–5 day old embryo, which raises many ethical questions (Vazin & Freed, 2010). A suitable solution is the use of human iPSC. They can be generated by reprogramming adult cells through the introduction of four specific genes which encode transcription factors (Takahashi & Yamanaka, 2006) (Figure 1). Similarly to ESC, iPSC can be expanded indefinitely and they are also able to differentiate into any type of cell. However, to use their full potential in medicine, some lines might need to be modified (gene correction) before usage (X. L. Li et al., 2018).

Genome editing technologies, specifically the Clustered Regularly Interspaced Short Palindromic Repeats (CRISPR)/Cas9 system, allow the generation of knock-out lines, the repair of mutated lines, and the derivation of reporter lines (Adli, 2018). Generating gene reporter lines for transcriptionally silent genes in iPSC can be challenging (Roberts et al., 2019). The HSCI iPS Core Facility is proposing to optimize tools that will simplify the targeting of transcriptionally silent genes to generate gene reporter PSC lines. These tools will allow us to improve protocols to streamline and increase the efficiency of PSC line targeting.

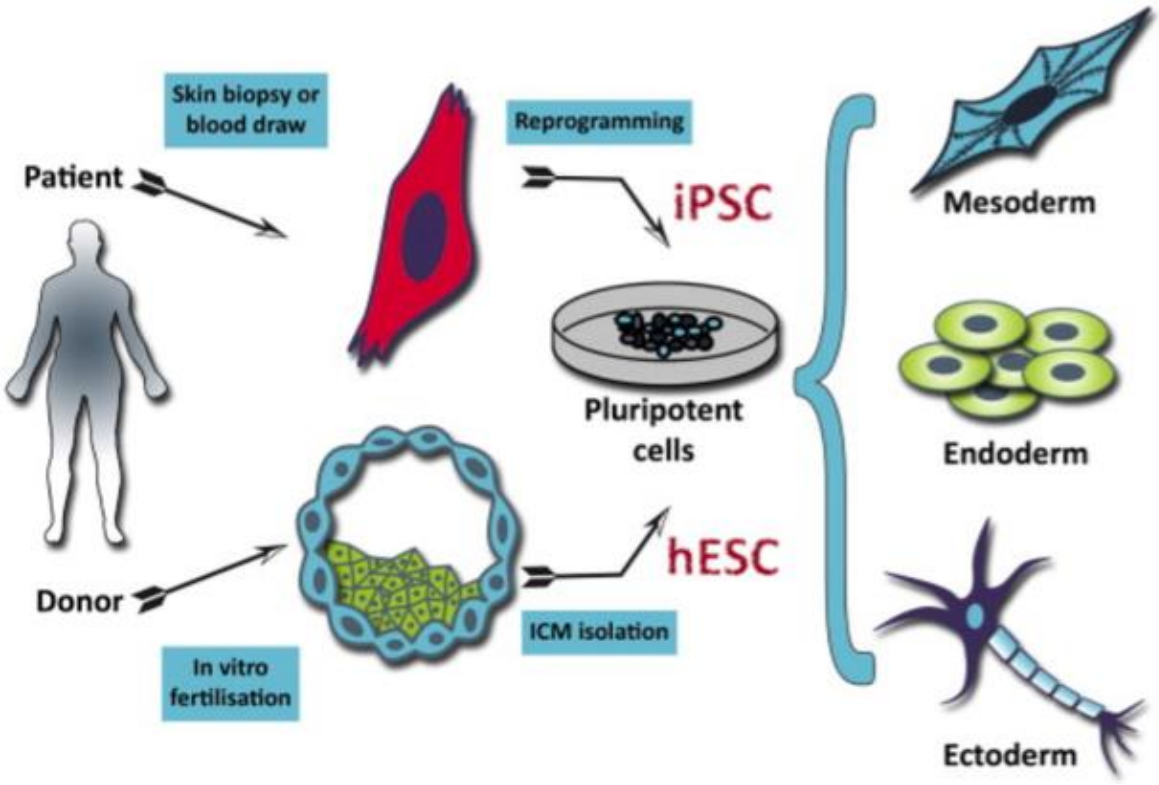

Figure 1: human ESC vs. iPSC generation – Despite their different origins, they can fulfill similar functions. Both are pluripotent and are able to differentiate into all three germ layers (mesoderm, endoderm and ectoderm). (Prowse et al., 2014)

1.1 Genome Editing

In the 1970s, the development of recombinant DNA technology started a new stage of biology. From this point on, scientists had the ability to manipulate DNA molecules and study genes. This knowledge found usage not only in medicine, but also in biotechnology, food and animal models. But just studying DNA was not enough, and scientists developed systems to directly edit or modulate the function of the DNA. This technology enables preventing the function of genes and regulatory elements, and in genome is organized and how genes and biological functions are connected (Hsu, Lander, & Zhang, 2014). Three major gene editing methods were developed in the last 25 years: Zinc Finger Nucleases, TALEN and CRISPR/Cas9. Within all three methods a double strand break (DSB) get introduced.

1.2 DSB repair mechanism

Genome editing is based on the cellular DNA repair machinery, which gets activated by the induced DNA DSB. The non-homologous end-joining (NHEJ) pathway is characterized by the ligation of the broken ends while homology-directed repair (HDR) repairs the DSB by using a donor template (Figure 2) (Sollu et al., 2010). There are also some alternative DSB repair pathways like single-strand annealing (SSA) and breakage-induced replication (BIR). But these are more error prone than the others.

The most common repair carried out by a cell when a DSB occurs is NHEJ. In this method the broken ends get directly ligated (Liu et al., 2018). It often generates random insertions and deletions (INDELs) at the site of repair. Therefore it is a more error prone but robust, predominant and fast pathway with high flexibility (Miyaoka, Mayerl, Chan, & Conklin, 2018). The introduction of indels can produce a frame shift in the coding frame of a gene of interest therefore it is used to knock-out genes. Compared to HDR, which occurs mainly in the S- or G2-phase of the cell cycle, NHEJ repair occurs during the entire cell cycle with a predominance in the G1 phase. The enzymes involved in NHEJ repair include *XRCC6, PARP3, PNKP, CRCC4, NHEJ1* and others. In the HDR pathway, homologous donor DNA sequences from sister chromatids or foreign DNA are used to repair gaps or create specific insertions or modifications. It is a faithful repair mechanism.

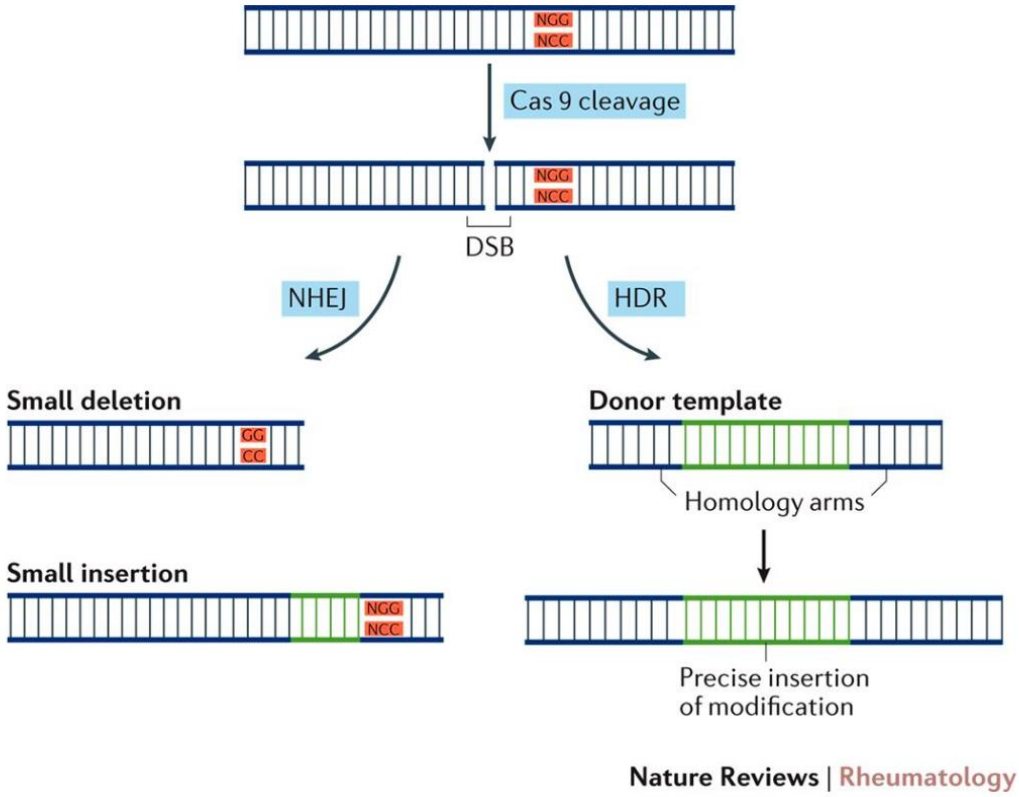

Figure 2: HDR and NHEJ Pathway - When a DSB occurs, the cell wants to repair this break immediately. Depending on the phase of the cell cycle and if a homologous sequence exists, the NHEJ or HDR pathway takes place. In the G1 phase of the cell cycle and the absence of a donor template, the NHEJ pathway occurs. Because of the inaccuracy of this method, small insertions or deletions can happen.

1.3 Gene reporter lines

Genes encoding proteins that are easily distinguished from the background of endogenous proteins are known as reporter genes (Jurgielewicz et al., 2017). Reporter genes are classified into two categories, fluorescent and non-fluorescent. The reporter gene category is chosen depending on the application. (X. L. Li et al., 2018).

As described above, with the CRISPR/Cas9 system a DSB can be introduced at a specific genomic location. The generation of gene reporter lines relies on HDR for the integration of the reporter genes. Exogenous donor DNA sequences with homology arms (HA) surrounding the reporter gene are used to integrate new genetic information to the cleaved genomic target (Jasin & Haber, 2016). Different DNA forms can act as efficient HDR donor templates.

1.3.1 Plasmid as HDR donor

The most common methodology to create gene reporter lines relies on the use of a large DNA template including 5' and 3' HA, typically ranging from 500 to 1000 nucleotides, encompassing the fluorescent marker of interest and a potential selectable marker (such as drug resistance to puromycin) (Baker et al., 2017). Generally, the workflow starts with the generation of this DNA template with cloning steps, bacterial transformation, DNA preparation and verification of the plasmid construct (S. Li et al., 2018). The next step in the workflow is to efficiently introduce this large DNA template into the cells in combination with the Cas9 protein and gRNA. The CRISPR/Cas9 system will trigger a DSB at a specific site and the DNA template will be integrated into the targeted gene site via HDR (Ryu, Hur, & Kim, 2019). However, human iPSC are known to have a low frequency of gene knock-in by HDR, making the generation of reporter lines a difficult process (Skarnes, Pellegrino, & McDonough, 2019). The addition of a drug resistant gene can facilitate the selection of the targeted cells; however, it increases the size of the DNA template. Moreover, the selection cassette needs to be subsequently excised (usually using a loxP/ Cre recombinase system) to avoid any interference with the endogenous expression of the gene of interest. Another obstacle in the creation of gene reporter cell lines is the multiple polymerase chain reaction (PCR) screens needed to select the few clones with the desired targeting (Samuel et al., 2013).

1.3.2 Double stranded DNA (dsDNA) as HDR donor

Recently, Thermo Fisher developed a new kit called the TrueTag DNA donor kit aiming at facilitating the workflow for the generation of reporter lines. The main advantage is the use of short HA (30-35 bases each) to ease the construction of the DNA template. The first step is to design and order oligos including the HA and a short sequence homologous to the TrueTag template. The gene of interest (GOI) specific DNA template is then generated by a simple PCR amplification followed by column purification. With this DNA template it is possible to tag any endogenous GOI with a fluorescent protein such as green fluorescent protein (GFP). Currently, the TrueTag donor templates include different selection markers (puromycin or blasticidin) fused to GFP through a T2A linker. This enables the enrichment of the edited cells using drug selection.

Although the kit offers a simple option for tagging genes in human iPSC, there are some inherent limitations. Currently this kit does not work to target genes that are silenced in iPSC since the drug resistant marker expression is dependent on the expression of the GOI. Additionally, the TrueTag DNA templates are designed to directly fuse the fluorescent protein to the protein of interest, either on the C-terminal or N-terminal end of the protein (Invitrogen, 2019). This results in the production of a large protein and we are concerned that this fusion could affect the function of the protein of interest. To overcome this hurdle, the T2A linker could be used between the GOI and the fluorescent gene. After the integration of the donor template, the T2A self-cleaving peptide will allow the generation of two separate proteins (one protein from the GOI and the second one being GFP). This reduces the possibility of the GFP interfering with the protein of interest.

1.3.3 Single stranded DNA (ssDNA) as HDR template

dsDNA is not the only type of template that can be used for HDR. Recent studies showed the efficient use of ssDNA as donor (Han Li, 2019). In this paper, the authors measure knock-in specificity at the target site using a droplet digital PCR (ddPCR) assay. By comparing the total frequency of GFP integration in the genome with the GFP integration at the target site, they showed that ssDNA is associated with a reduced likelihood of off-target integration. Furthermore, a lower toxicity was observed compared to dsDNA donors. These results were replicated in PSC (Roth et al., 2018). Small ssDNA donors (<200 nt) are easy to synthesize and therefore widely available. Longer ssDNA HDR templates are more difficult to generate.

An efficient and scalable method for the generation of ssDNAs is the reverse-transcription (RT) of an RNA intermediate followed by specific hydrolysis (Leonetti, Sekine, Kamiyama, Weissman, & Huang, 2016) (Murgha, Rouillard, & Gulari, 2014) (Han Li, 2019). This method still does not solve the whole problem because the RT of long sequences is challenging (Leonetti et al., 2016). Different RT enzymes have been tested and it was found that the commercial TGIRT-III worked best. As shown in Figure 3, the first step is an *in vitro* transcription (IVT) of the dsDNA template. Afterwards the RNA gets reverse-transcribed into an RNA/DNA hybrid and hydrolyzed into an ssDNA. By this synthesis, >50 µg of final ssDNA can be generated (Han Li, 2019).

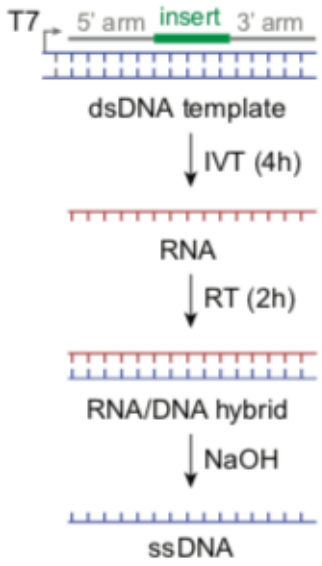

Figure 3: RT mediated ssDNA HDR donor generation – IVT of the dsDNA template leads to the formation of an RNA. Through RT of the RNA intermediate, an RNA/DNA hybrid gets synthesized. The hybrid gets hydrolyzed at high pH, therefore NaOH is used. (Han Li, 2019)

To achieve the best results, some important ssDNA design rules were identified by evaluating the integration efficiency of long ssDNA HDR templates. It was shown that in HEK293T and K562 cells, there is a near-exponential relationship between HA length and knock-in efficiency. Next, the amount of donor was tested and it was shown that a maximum efficiency of 95 % was obtained by using 20 pmol of donor for $2x10^5$ cells. Another important finding was that ssDNA donors have a better repair specificity compared to dsDNA sequences and also the rate of off-target integrations is reduced. In a comparison of donor complementary or non-complementary to the gRNA, no consistent differences were observed in the knock-in efficiency. Overall it was shown that in HEK293T cells, the ssDNA donor integration is highly efficient and also very specific (Han Li, 2019).

1.3.4 Split GFP

A new method that has the potential to revolutionize the field of gene reporter lines is the use of split GFP. The cylindrical structure of GFP consists of 11 strands of beta-sheet with an alpha-helix inside and short helical segments on the ends of the cylinder (Yang, Moss, & Phillips, 1996). In the case of split GFP, the molecule gets split into two parts, GFP11 and GFP1-10. GFP11 corresponds to the 11^{th} ß- strand of the superfolder GFP ß-barrel structure, while GFP1-10 corresponds to the rest of the molecule. Neither is fluorescent by itself (Cabantous et al., 2005). The GFP1-10

fragment has to be constitutively expressed in the cells, while the GFP11 can be tagged to a gene of interest. When this gene is activated, the cells express both GFP fragments which can come together and allow functional GFP tagging (Figure 4). In addition, it was shown that the intensity of the split GFP complex is identical to that of full-length GFP (Kamiyama et al., 2016). The key advantage of this method is the small size of GFP11 (16 amino acids), which allows ssDNA templates to be used as HDR donors (Leonetti et al., 2016).

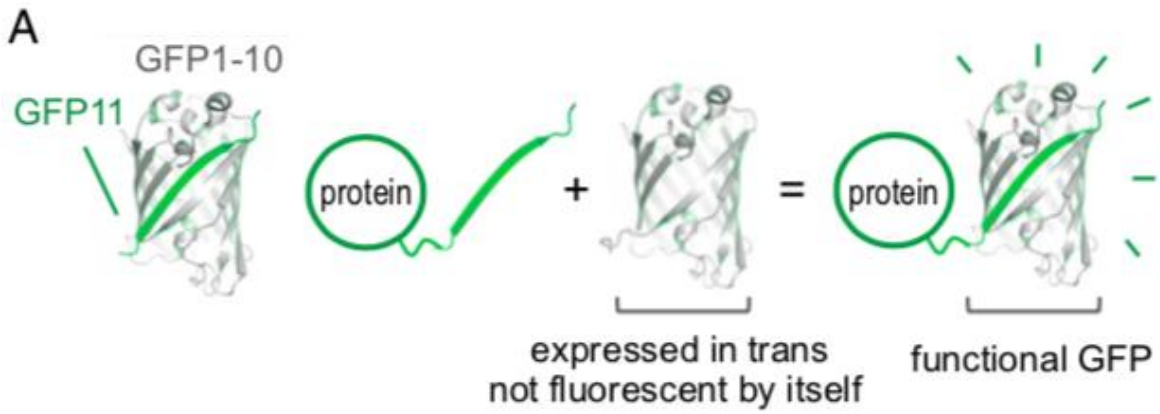

Figure 4: Principle of split GFP – The GFP1-10 fragment is constitutively expressed in the cells. The GFP11 fragment is fused to a protein of interest and recruits its GFP1-10 interaction partner. Through GFP completion, fluorescent tagging takes place. (Leonetti et al., 2016)

Together with a three amino acid linker, it is only 57nt long. Therefore, it is possible to synthesize ssDNA of <200 nt with HAs ~70nt long each. After assembly with the pre-mixed Cas9/gRNA (RNP), the donor template gets introduced into GFP1-10 expressing cells by electroporation. This method enables a high-efficiency fluorescent tagging of human proteins at their endogenous loci with minimal preparation (Figure 5) (Leonetti et al., 2016).

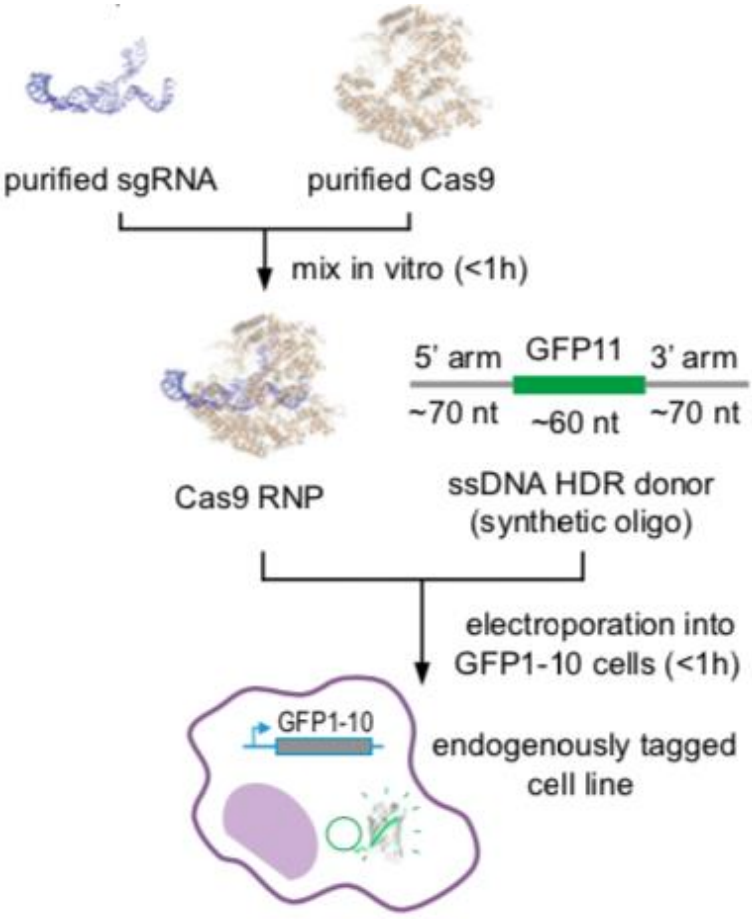

Figure 5: Workflow of split GFP method – The ssDNA HDR donor consists of GFP11 with a 3 amino acid linker sequence and two HAs (5´and 3´) of ~70nt each. The HAs are important to enable HDR. The purified gRNA and Cas9 get mixed to form the RNP. Next, the RNP and the ssDNA HDR donor get introduced into GFP1-10 expressing cells via electroporation. (Leonetti et al., 2016)

It was shown that low protein expression can affect the efficiency of GFP tagging. The reason might simply be the fact that the fluorescence signal of the target is below the detection limit. To overcome the hurdle of low protein abundance, GFP11 sequence repeats can be used. However, GFP11 repeats exceed the current size limitation of ssDNA HDR donors. In this case, the strategy described in the previous section (including IVT, reverse transcription) should be applied to synthesize the ssDNA donor. It was shown that the GFP11 cassette was integrated with similar efficiency and led to a corresponding increase in fluorescence intensity (Figure 6) (Kamiyama et al., 2016).

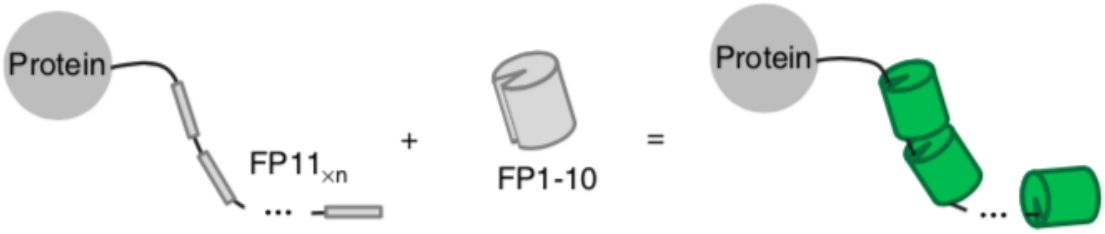

Figure 6: Schematic diagram of GFP11 repeats – GFP11 repeats lead to a proportional fluorescence increase. (Kamiyama et al., 2016)

2 Materials and Methods

2.1 Cell Lines

Human iPS cells (PGP-1 iPS line) were obtained from the George Church Lab (Harvard Medical School). The work with this human cell line was carried out in accordance to national and international guidelines.

Fibroblasts (HS27 Fibro) were obtained from American Type Culture Collection (ATCC).

2.2 Maintenance of iPSC

All steps were carefully carried out under sterile working conditions. Media were prepared according to the manual. The cells were cultured on StemFlex™ Medium (Thermo Fisher Scientific, Waltham, USA). Cultrex® (Trevigen, Gaithersburg, USA) matrix-coated culture plates (Greiner Bio-One, Kremsmünster, Austria) at a concentration of 120 – 180 µg/ml were used to culture iPSCs. Cells cultured on Stemflex were fed every 48 hours.

The cells were passaged when reaching a confluency of 70-80 % at a ratio of 1:4 to 1:6. Before passaging, cells exhibiting spontaneous differentiation were removed using a pipette tip and a dissection microscope (SMZ1500 Stereoscopic microscope, Nikon, Minato, Japan) under sterile conditions. For passaging, the cells were treated with Gentle Cell Dissociation Reagent (Stemcell Technologies, Vancouver, Canada) for four minutes at room temperature. After cells appeared to begin separating from each other, the solution was aspirated and media was added. The cells were detached utilizing a cell lifter (Corning Inc., Corning, USA). Afterwards the cell suspension was replated or used for NEON® transfection.

2.3 Maintenance of Fibroblasts

All steps were carefully carried out under sterile working conditions. Media were prepared according to the respective manual. The cells were cultured on DMEM

(Thermo Fisher Scientific, Waltham, USA) with 10 % Fetal Bovine Serum (FBS) (VWR, Radnor, USA) and the media were changed every 72 hours.

The cells were passaged at a confluency of 70-80 % at a ratio of 1:4 to 1:6. For passaging, the cells were treated with Trypsin-EDTA (0,05 %) (Invitrogen, Carlsbad, USA) for four minutes at 37 °C. After cells appeared to detach media with FBS was added to inhibit the Trypsin-EDTA. The cell suspension was replated or used for NEON® transfection.

2.4 Viable cell count

To distinguish viable from nonviable cells, a Trypan Blue Stain exclusion test was performed. Upon collection, 20 µl of cell suspension was mixed with 20 µl Trypan Blue Stain (Thermo Fisher scientific, Waltham, USA) and analyzed with the automated cell counter (TC10, BioRAD, Hercules, USA).

2.5 Plasmid for targeting *DLX5*

The plasmid vector HR120:eGFP:TK (Table 1) was used for targeting the transcriptional silent gene, *DLX5* Figure 7.

Table 1: Plasmid vector to target *DLX5*

Name	Purpose	Origin
HR120:eGFP:TK	Targeting of *DLX5*	Gift from Dr. Lee Rubin Lab, Harvard University

The key components of the plasmid are the HA, which are necessary to allow the HDR. To avoid a direct fusion of the GFP to the gene, a T2A linker was used. The Woodchuck Hepatitis Virus (WHP) Posttranscriptional Regulatory Element (WPRE) is a DNA sequence which is used to increase the expression of the selection cassette (Klein et al., 2006). A negative selection marker (HTLV- TK) is placed adjacent to the WRPE (Hall, Limaye, & Kulkarni, 2009). Puromycin drug resistance is also included to allow positive selection. A SV40pA is included to signal the end of the selection unit (S. P. Li, Feng, Wang, Wang, & Lv, 2012). Furthermore, LoxP sites are located on each site of the selection cassette to allow a later excision by *Cre* recombinase.

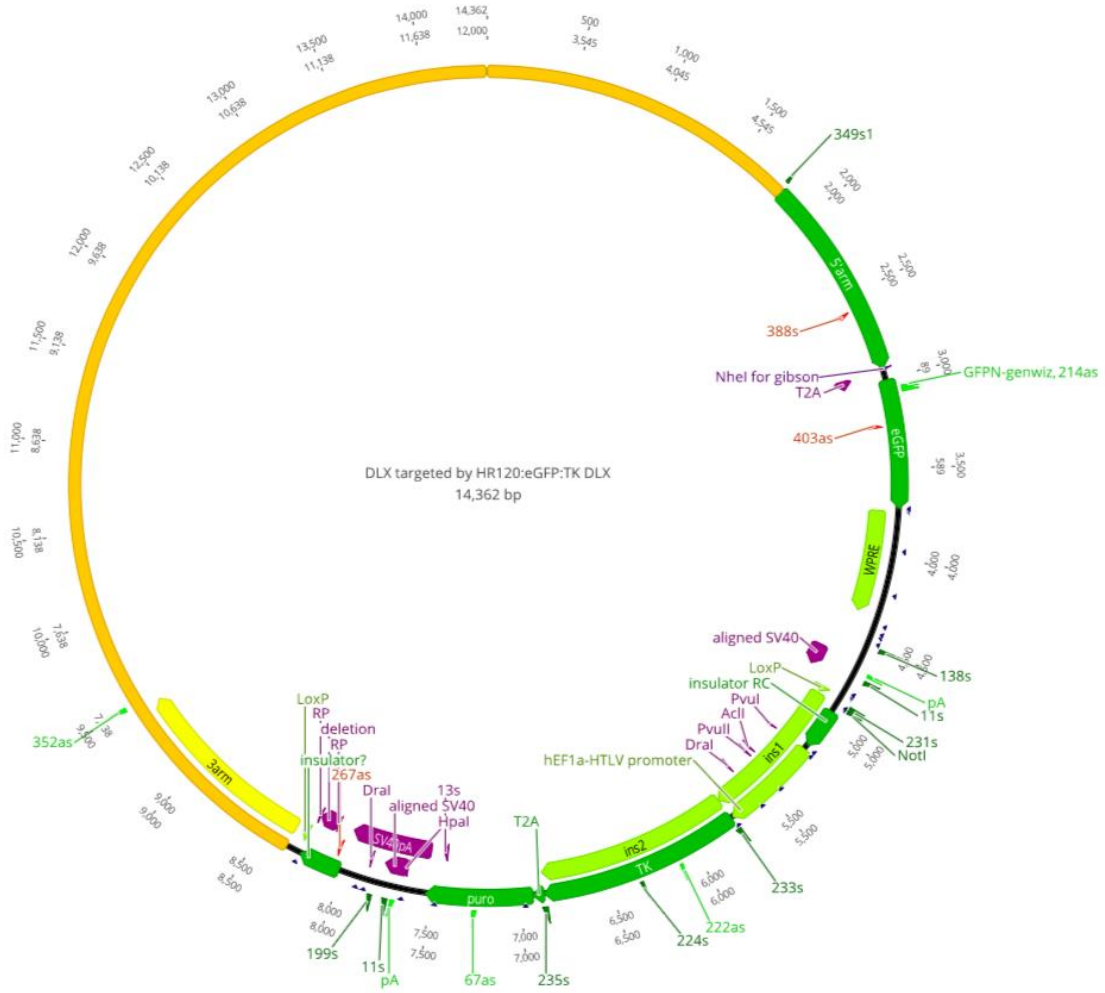

Figure 7: *DLX5* targeting vector - The two HA flank the *DLX5* gene to allow HDR. A T2A linker is included to prevent direct fusion of the eGFP to the *DLX5* gene. WRPE is placed adjacent to the GPF, in front of the negative selection marker (HTLV- TK). For positive selection, puromycin is placed next to the HTLV-TK. A SV40pA is included to signal the end of the transcriptional unit. LoxP sites are located on each site of the selection cassette to allow a later excision by *Cre* recombinase. (Klein et al., 2006) (Hall et al., 2009) (Li et al., 2012)

2.6 N-terminal donor for targeting *ACTB*

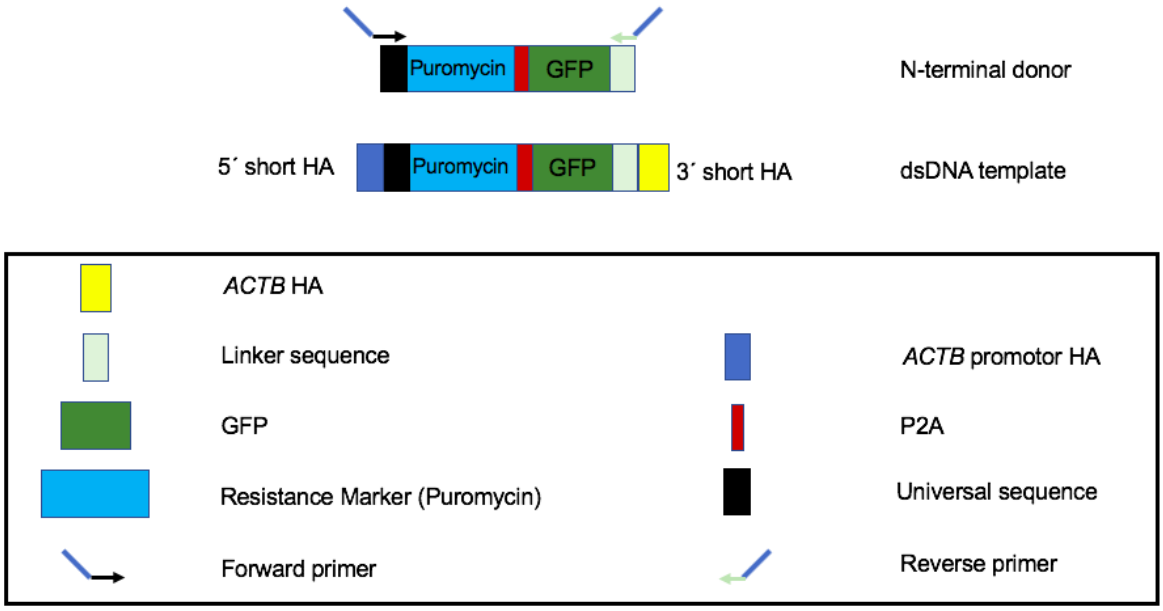

Figure 8: N-terminal donors for generation of dsDNA template to target *ACTB* – The primers are used for the generation of a short dsDNA template without an unnecessary backbone. Puromycin was included as a resistance marker.

The Truetag Donor DNA kit from Thermo Fisher was first tested by using a control provided with the kit. This control includes a N-terminal donor template and primers specific for the Actin B gene (Table 2). It was used for the generation of a dsDNA donor for targeting ACTB. (Figure 8) The forward and reverse primers are used to generate the HA which are needed for the HDR. The addition of a universal and linker sequence facilitates the design to target any gene of interest. By performing a PCR, dsDNA template can be prepared in one easy step. A puromycin selection marker was included to enables an easy selection. With a P2A linker, the direct fusion of the GFP reporter is impeded.

Table 2: N-terminal donor to target *ACTB*

Name	Purpose	Catalog #	Company
N-terminal donor *ACTB* control	Targeting of *ACTB*	A42992	Thermo Fisher Scientific, Waltham, USA

2.7 gRNAs for targeting *DLX5* and *ACTB*

All the gRNAs used for targeting *DLX5* and *ACTB* are listed as 5´-3´ Sequences in Table 3. The gRNAs were selected using the Geneious Prime Software and ordered from Synthego. All gRNAs were approximately 19 nt long and upstream of an NGG PAM site.

Table 3: gRNAs for targeting *DLX5* and *ACTB*

NO.	NAME	SEQUENCE	COMPANY
1	Mod EZ sgRNA-LL591DLXg2	5´-TCCGGGACACTCTATTAGAT-3´	Synthego, Redwood City, USA
2	Mod EZ sgRNA – LL844ACTBg1	5´-GCGGCGATATCATCATCCA-3´	Synthego, Redwood City, USA

2.8 Design gRNA for targeting *TBXT* and *FOXA2*

Specific gRNA for targeting *TBXT* and *FOXA2* were designed using the online tool CRISPOR. The tool is capable of identifying gRNA in an input sequence and ranks them according to different scores. For selection of the best fitting gRNA, the specificity score, the predicted efficiency and the off targets in the genome of interest were compared. The two gRNAs that best satisfied the criteria were selected for each gene (Table 4).

Table 4: gRNAs for targeting *TBXT* and *FOXA2*

NO.	NAME	SEQUENCE	COMPANY
1	gRNA-*TBXT*_1	5´-AACTCCTCTTAAGAAGACGA-3´	Synthego, Redwood City, USA
2	gRNA-*TBXT*_2	5´-CATGTGAAGCAGCAAGGCCC-3´	Synthego, Redwood City, USA
2	gRNA-*FOXA2*_1	5´-AACTCCTCTTAAGAAGACGA-3´	Synthego, Redwood City, USA
3	gRNA-*FOXA2*_2	5´-GAAGCCGTCGTCTTCTTAAG-3´	Synthego, Redwood City, USA

2.9 Design of ssODN for targeting *TBXT* and *FOXA2*

A single stranded oligo donor (ssODN) for targeting *TBXT* and another ssODN for targeting *FOXA2* were designed by using the Geneious Prime software. Compared to a three amino acid linker mentioned in the introduction, our collaborator choose a ten amino acid linker. The use of a longer linker can eliminate the change of interaction between GFP and the protein. In Table 5 the ssODNs are shown for both chosen gRNAs for each gene. Green indicates the GFP11 sequence, blue indicates the linker and red indicates a mutation within the PAM region. Compared to a three amino acid linker mentioned in the introduction, our collaborator choose a ten amino acid linker. The use of a longer linker can eliminate the change of interaction between the GFP and the protein.

Table 5: ssODN for TBXT and FOXA2

Name	ssODN sequence
ssODN *TBXT* (gRNA 1)	GAC GCC GCA GCC CAA GGC CGC CTC ATA GCC TCA TGG ACA CCT GTG TCG CCA CCT TCC ATG GGT GGT GGT GGT TCT GGT GGT GGT GGT TCT ACC GAG CTC AAC TTC AAG GAG TGG CAA AAG GCC TTT ACC GAT ATG ATG TGA AGC AGC AAA GCC CAG GTC CCG AAA GAT GCA GTG ACT TTT TGT CGT GGC AGC CAG TGG
ssODN *TBXT* (gRNA 2)	GAC GCC GCA GCC CAA GGC CGC CTC ATA GCC TCA TGG ACA CCT GTG TCG CCA CCT TCC ATG GGT GGT GGT GGT TCT GGT GGT GGT GGT TCT ACC GAG CTC AAC TTC AAG GAG TGG CAA AAG GCC TTT ACC GAT ATG ATG TGA AGC AGC AAG GCC CAA GTC CCG AAA GAT GCA GTG ACT TTT TGT CGT GGC AGC CAG TGG
ssODN *FOXA2* (gRNA 1)	CTG GCC GCA GAT ACC TCC TAC TAC CAG GGG GTG TAC TCC CGG CCC ATT ATG AAC TCC TCT GGT GGT GGT GGT TCT GGT GGT GGT GGT TCT ACC GAG CTC AAC TTC AAG GAG TGG CAA AAG GCC TTT ACC GAT ATG ATG TAA GAA GAC GAC AGC TTC AGG CCC GGC TAA CTC TGG CAC CCC GGA TCG AGG ACA AGT GAG AGA
ssODN *FOXA2* (gRNA 2)	CTG GCC GCA GAT ACC TCC TAC TAC CAG GGG GTG TAC TCC CGG CCC ATT ATG AAC TCA TCT GGT GGT GGT GGT TCT GGT GGT GGT GGT TCT ACC GAG CTC AAC TTC AAG GAG TGG CAA AAG GCC TTT ACC GAT ATG ATG TAA GAA GAC GAC GGC TTC AGG CCC GGC TAA CTC TGG CAC CCC GGA TCG AGG ACA AGT GAG AGA

2.10 NEON® Transfection

Electroporation of Cas9/gRNA RNP was used due to its high efficiency and specificity. The RNPs, with the appropriate gRNA and TrueCut™ Cas9 Protein v2. (Thermo Fisher Scientific, Waltham, USA), were mixed *in vitro*. The HDR template containing the tag

sequence and HA to the target locus was supplied either as a plasmid or as a dsDNA Template.

2.10.1 Testing of gRNA

For testing the efficiency of the gRNA, different conditions were taken into consideration. Different operation set-ups were used for the electroporation with the NEON® system (Table 6). One zap of transfection includes 10 µl with 100,000 cells.

Table 6: Parameter set-up for Electroporation with NEON® for gRNA testing

NO.	ELECTRIC POTENTIAL [V]	WIDTH [MS]	NO. OF PULSES	NO. OF ZAPS
1	1000	20	1	2
2	1100	20	1	2
3	1200	20	1	2
4	1300	20	1	2

The RNP were prepared after the set-up of the NEON® system according to the manufacturer's manual.

was also performed with a lower concentration of Cas9 and gRNA (Table 8). For the formation of the RNP, the gRNA (Table 3, No. 1), Cas9 and the Resuspension Buffer R (Invitrogen, Carlsbad, USA) were prepared, mixed and incubated at room temperature for 15-20 minutes (Table 7). For comparison, the experiment

Table 7: high concentration RNP formation for gRNA test

Reagents	Mole usage per zap (mass)	Stock concentration	Volume per zap [µl]
Cas9	6,25 pmol	5 µg/µl	0,2
gRNA (Table 3 No.1)	12,5 pmol	50 pmol/µl	0,25
Buffer R	-	-	4,55

Table 8: low concentration RNP formation for gRNA test

Compound	Mole usage per zap (mass)	Stock concentration	Volume per zap [µl]
Cas9	2 pmol	5 µg/µl	0,07
gRNA (Table 3 No. 1)	4 pmol	50 pmol/µl	0,08
Buffer R	-	-	4,85

During incubation, the PGP-1 cells were dissociated and prepared for transfection. For detaching the cells, Accutase (Stemcell Technologies, Vancouver, Canada) was added to the cells and incubated for ten minutes. Pellets of 100,000 cells were resuspended in 4,75 µl of Resuspension Buffer R. The cell suspension was added to the RNP (Table 7 or Table 8) and mixed gently. The electroporation of the cells was performed according to the manufacturer's manual with two zaps per well. After transfection, the cells were plated on Cultrex® matrix-coated plates with media with rock inhibitor. The cells were cultured in the incubator at 37 °C and 5% CO_2 (Thermo Fisher Scientific, Waltham, USA) until analysis.

2.10.2 PGP-1 transfection with Plasmid to target DLX5

For GFP targeting of the *DLX5* gene, PGP-1 cells were transfected with the plasmid (Figure 7). Different operation set-ups were used for the electroporation with the NEON® system (Table 9). One zap of transfection includes 10 µl with 100,000 cells.

Table 9: Parameter for Electroporation with NEON® to target *DLX5*

NO.	ELECTRIC POTENTIAL [V]	WIDTH [MS]	NO. OF PULSES	NO. OF ZAPS
1	1000	20	1	3
2	1100	20	1	3
3	1200	20	1	3

After the set-up of the NEON® system according to the manufacturer's manual, the RNPs (Table 10) were prepared and incubated at room temperature for 15–20 minutes.

Table 10: RNP formation for targeting *DLX5*

Compound	Mole usage per zap (mass)	Stock concentration	Volume per zap [µl]
High concentrated RNP			
Cas9	6,25 pmol	5 µg/µl	0,2
gRNA (Table 3, No. 1)	12,5 pmol	50 pmol/µl	0,25
Buffer R	-	-	4,55
Low concentrated RNP			
Cas9	2,0 pmol	5 µg/µl	0,07
gRNA (Table 3 No.1)	4 pmol	50 pmol/µl	0,08
Buffer R	-	-	1,52

During incubation of the RNP, the cells were prepared, the plasmid was added to a concentration of 500 ng/zap and all steps were continued as described in 2.10.1.

2.10.3 PGP-1 transfection with dsDNA to target *ACTB* as a control

Different operation set-ups were used for the electroporation with the NEON® system for targeting *ACTB* in PGP-1 cells (Table 11).

Table 11: Parameter for Electroporation with NEON® to target *ACTB* in iPSC

NO.	ELECTRIC POTENTIAL [V]	WIDTH [MS]	NO. OF PULSES	NO. OF ZAPS
1	1000	20	1	2
2	1100	20	1	2
3	1200	20	1	2

After the set-up of the NEON® system according to the manufacturer's manual, the RNPs were prepared (Table 12) and incubated at room temperature for 15-20 minutes.

Table 12: RNP formation for targeting *ACTB*

Compound	Mole usage per zap (mass)	Stock concentration	Volume per zap [µl]
Cas9	2,0 pmol	5 µg/µl	0,07
gRNA (Table 3 No. 2)	4 pmol	50 pmol/µl	0,08
Buffer R	-	-	1,52

During incubation of the RNP, the cells were prepared and the TrueTag Template was added to a concentration of 60 ng/zap. All steps were continued as described in 2.10.1.

2.10.4 Fibroblast transfection with dsDNA to target *ACTB* as a control

Different operation set-ups were used for the electroporation with the NEON® system for targeting *ACTB* in fibroblasts (Table 13).

Table 13: Parameter for Electroporation with NEON® to target *ACTB* in Fibroblast

NO.	ELECTRIC POTENTIAL [V]	WIDTH [MS]	NO. OF PULSES	NO. OF ZAPS
1	1400	20	1	3
2	1650	20	1	3

After the set-up of the NEON® system according to the manufacturer's manual, the RNPs were prepared (Table 14) and incubated at room temperature for 15-20 minutes.

Table 14: RNP formation for targeting *ACTB*

Compound	Mole usage per zap (mass)	Stock concentration	Volume per zap [µl]
Cas9	2,0pmol	5µg/µl	0,07
gRNA (Table 3 No. 2)	4pmol	50pmol/µl	0,08
Buffer R	-	-	1,52

During incubation of the RNP, the cells were prepared, the TrueTag Template was added to a concentration of 60 ng/zap. During incubation, the HS27 fibroblast cells were dissociated and prepared for transfection. For detaching of the cells, trypsin 0.05% EDTA was added to the cells and incubated for four minutes at 37 °C. Pellets of 100,000 cells were resuspended in 4,75 µl of Resuspension Buffer R. The cell suspension was added to the RNP (Table 14) and the TrueTag Template was added to a concentration of 60 ng/zap. After a gentle mixing of the cells, the electroporation was performed according to the manufacturer's manual with two zaps per well. After transfection, the cells were plated on Cultrex® matrix-coated plates with media with Rock inhibitor. The cells were cultured in the incubator at 37 °C and 5% CO_2 until analysis.

2.11 DNA Extraction

DNA extraction was performed using the QIAmp® DNA Mini Kit (Qiagen, Hilden, Germany). The DNA concentration was measured with the NanoVue Plus (Biochrom Ltd, St. Albans, UK).

2.12 PCR

The primer mix (5 µM final) were prepared with a ratio of primary pair mix of 1 (sense primer): 1 (antisense primer): 18 (distilled water). The DNA template was diluted to an average concentration of 30 ng/µl.

2.12.1 Testing of gRNA

PCR was used to test the function of the gRNA. The primer pair is shown in Table 15.

Table 15: Primer Sequences for gRNA test

Name	Direction	Sequence	Product Size	Company
LL-DLX5	sense	5´-GTAACGTGTGCCTTTGTTCC-3´	567 bp	Integrated DNA Technologies, Coralville, USA
LL-DLX5	antisense	5´-TTGAAAGTGTCCACAGTTGC-3´		

Table 16: PCR Master Mix for gRNA test

Compound	Amount [µl] / Reaction	Company
Primer Mix	5	Integrated DNA Technologies, Coralville, USA
5X Phusion Green Buffer	10	Thermo Fisher scientific, Waltham, USA
dNTP mix	2	Thermo Fisher scientific, Waltham, USA
Phusion Green Hot Start II High fidelity DNA Pol	0,5	Thermo Fisher scientific, Waltham, USA
UltraPure™ DNase/RNase-Free Distilled Water	30,5	Thermo Fisher scientific, Waltham, USA
Sample DNA [30 ng]	2	-

The components and amounts for the PCR master mix necessary for the amplification of the targeted region are shown (Table 16).

Table 17: PCR reaction conditions for gRNA test

Step	Cycles	Temerature [ºC]	Time [s]
Denature	1	98	10
Amplification	30	98	10
		62	20
		72	15
Final extension	1	72	5
Final hold	∞	4	∞

The PCR Reaction Conditions for testing of the gRNA are shown in Table 17.

2.12.2 dsDNA generation for targeting *ACTB*

The True Tag dsDNA to target *ACTB* with GFP was obtained through amplification by PCR. The primer pair is shown in Table 18.

Table 18: Primer Sequences for the generation of dsDNA for targeting *ACTB*

Name	Direction	Sequence	Product Size	Company
Actin TrueTag5	sense	5´-CACAGCGCACCCGGGTATTCTCGCAGCTCACCATGGGAGGTAAGCCCTTGCATTCG-3´	1486 bp	Integrated DNA Technologies, USA
Actin TrueTag5	Anti-sense	5´-CCGTTGTCGACGACGAGCGCGGCGATATCATCATCACCGCTTCCACTACCTGAACC-3´		

The PCR master mix was prepared and the PCR reaction conditions were set according to the TrueTag™ DNA Donor Kit manual (Invitrogen, Carlsbad, USA). As donor template the plasmid TrueTag, GFP-N-PURO-GFP (Invitrogen, Carlsbad, USA) was used. After PCR reaction the TrueTag DNA was cleaned by the TrueTag Cleanup Kit (Thermo Fisher Scientific, Waltham, USA). The Template was eluted once in the included Elution Buffer and once in Buffer R.

Depending on the recovery, 3-4 days post transfection PGP-1 cells were treated with 0,5 µg/ml puromycin (Thermo Fisher Scientific, Waltham, USA). The HS27 fibroblast cells were treated with 1,25 µg/ml puromycin. After three days of puromycin selection, the media were switched back to culture media without drug selection and feeding the cells proceeded accordingly.

2.13 Agarose Gel Electrophoresis

A 1 % agarose gel was prepared with 50 ml 1x TAE buffer (Thermo Fisher Scientific, Waltham, USA) and 0.5 g agarose (Bio-Rad, Hercules, USA). The solution was carefully microwaved until agarose was completely dissolved and 0,2 µg/ml ethidium bromide (Sigma-Aldrich, St. Louis, USA) was added. The dissolved agarose solution was carefully poured into the casting form and allowed to solidify at room temperature before the samples were loaded. Samples were run at 150 V and 18 mA.

2.14 Sequencing

A purification of the PCR product was performed according to the QUIAquick PCR Purification Kit manual. The Geneious Prime Software was used to select a suitable primer. The extracted DNA and the primer were sent to Genewiz® for sequencing.

2.15 Fluorescence Microscopy

To ensure the correct GFP tagging of the HS27 fibroblasts, the presence of GFP was checked with a fluorescence microscope (Olympus-). Both 4x and 10x magnification were used.

3 Results

3.1 Test of gRNA for targeting *DLX5*

The genome editing performance is highly dependent on the gRNA (Bortesi et al., 2016). The HSCI iPS core had a gRNA for DLX5 that had been ordered few months before and therefore has passed its expiration date. First, I decided to test it to check if it could still cut efficiently. There are a wide variety of transfection methods available for mammalian cells but all are not ideal for stem cells (Manzini, Viiri, Marttila, & Aalto-Setala, 2015). The iPS Core Facility found the NEON® transfection system to be the optimal system to transfect human PSC and was therefore used in all of the experiments.

3.1.1 Old gRNA

A NEON® transfection of PGP-1 iPSC with gRNA and Cas9 was performed. To investigate the best cutting efficiencies, this experiment was performed at different voltages. The cell viability and cell recovery is highly dependent on the voltages. Higher voltage conditions lead to higher cell death. (Figure 9) Similar results were shown in the experiment with the lower RNP concentration (data not shown). After a recovery of four days, the DNA were extracted and PCR was performed. To confirm the size of the PCR product, gel electrophoresis was performed. Then, all the samples were sent for sequencing. As shown in Figure 10, the percentage of INDEL is really low in all four conditions. The highest efficiency obtained was 14 % in the 1100 V test.

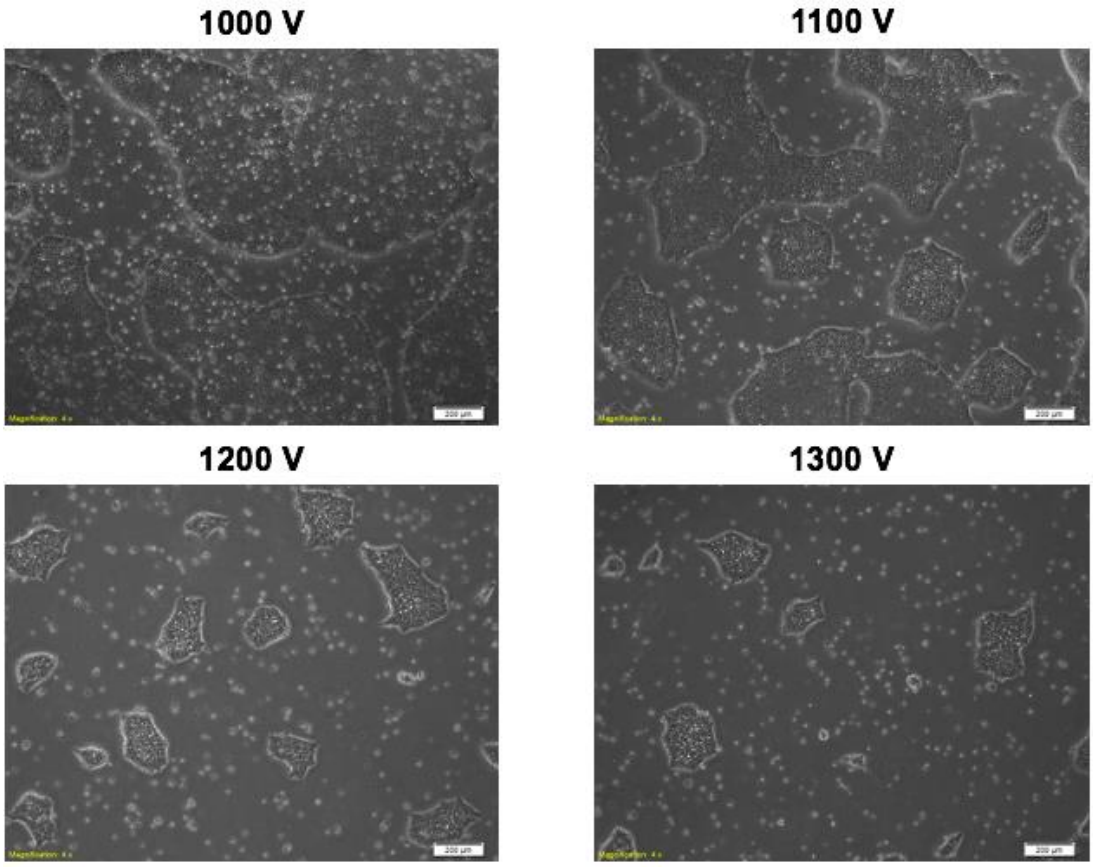

Figure 9: NEON® transfected PGP-1 iPSC with high concentrated RNP - gRNA and Cas9 transfected cells at different voltage conditions. The culture confluence is higher at both 1000 V and 1100 V compared to 1200 V and 1300 V. Scale bar represents 1000 µm.

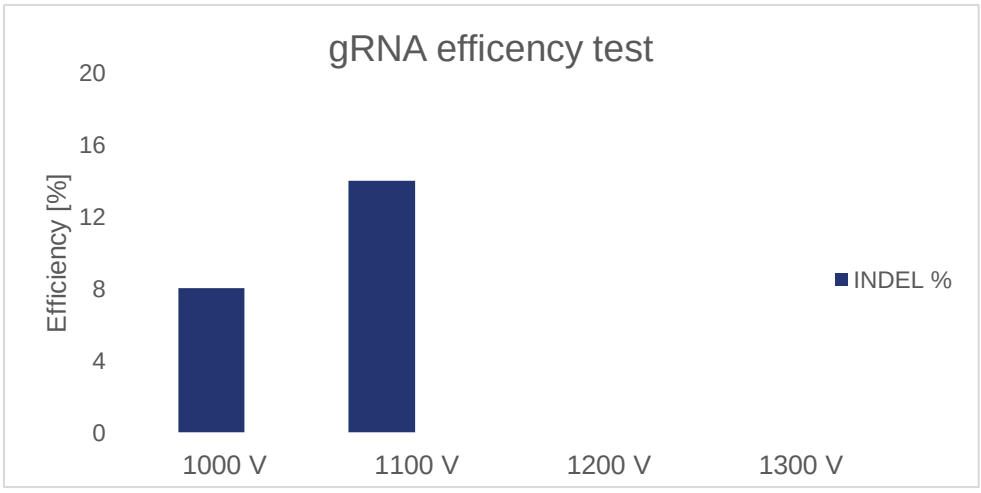

Figure 10: gRNA efficiency test - The efficiency of the gRNA is shown over various transfection conditions (1000 V, 1100 V, 1200 V and 1300 V).

3.2 Optimization of NEON® transfection conditions to target *DLX5* with the donor plasmid

To find the perfect conditions for NEON® transfection, 100,000 iPSC per reaction were transfected with the plasmid (Figure 11) at different instrument settings (Table 9). For this experiment, a commercially available Cas9 protein and the old gRNA were used. In addition to the different settings, the concentration of the Cas9 and gRNA were optimized. The conditions with 6 pmol Cas9 and 12 pmol gRNA at 1000 V/1100 V/1200 V led to high cells death (data not shown). The cells transfected with 2 pmol Cas9 and 4 pmol gRNA (at different voltage settings) showed a much better cell viability compared to the high concentrated RNP amount (Figure 11). Therefore, the subsequent experiments were only performed with these conditions. As expected, the cell viability went down when the voltage was increased.

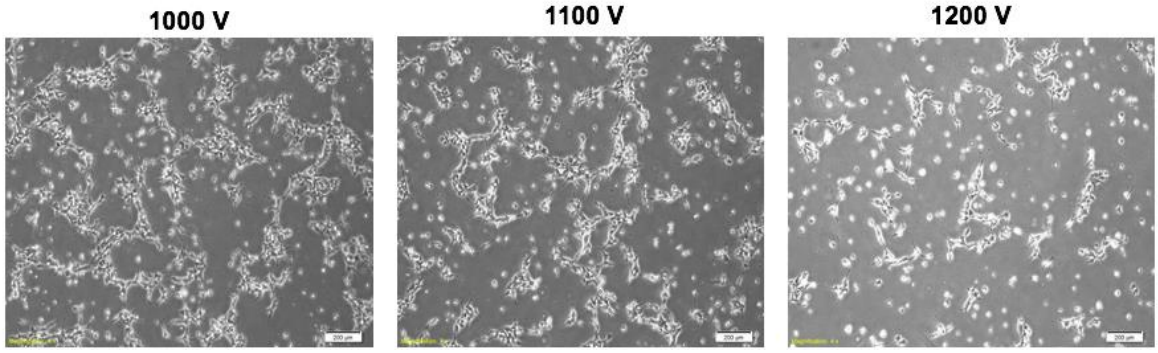

Figure 11: Cell viability after NEON® transfection of PGP-1 iPSC – gRNA, Cas9 and plasmid transfected cells at different voltage conditions. There is an indirect proportion between voltage and cell viability. Higher voltage leads to lower cell viability. Bar represents 200 µm.

Three days after transfection, puromycin selection was performed. All the cells were dead following selection (data not shown). It is likely that the gRNA efficiency was too low to get a successful targeting. Therefore, a new gRNA was ordered from Synthego and tested again.

3.3 Targeting of *ACTB*

To investigate if the TrueTag system works appropriately, the control gene *ACTB* was targeted in fibroblasts and iPSC. For the generation of the dsDNA HDR donor

template, a PCR was performed (Figure 8). In addition, the construction for targeting *ACTB* is shown in Figure 12.

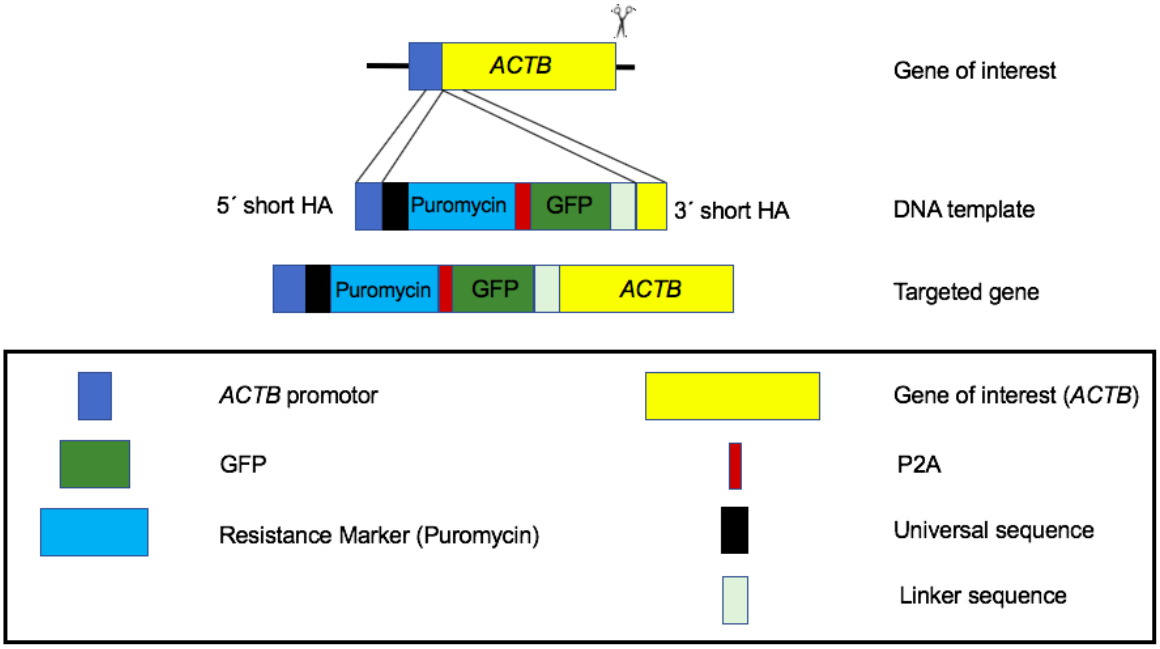

Figure 12: Construction of *ACTB* gene targeting - Creation of the double stranded donor template which contains the GFP as a reporter, puromycin as a selection marker and P2A for cleavage. The HA were added by PCR to a premade tagging template.

The transfection was performed with either 100,000 iPSC or 100,000 fibroblasts and the dsDNA HDR donor template. Different settings were also tested (Table 11). As shown in Figure 11, high PGP-1 cell death occurs at the 1200 V condition, therefore only 1000 V and 1100 V were compared. For fibroblasts, higher voltage settings are tolerated.

Four days after transfection, similar cell recovery was observed in both cell types (Figure 13). The green fluorescence indicates that the transfection worked and the Actin gene was tagged.

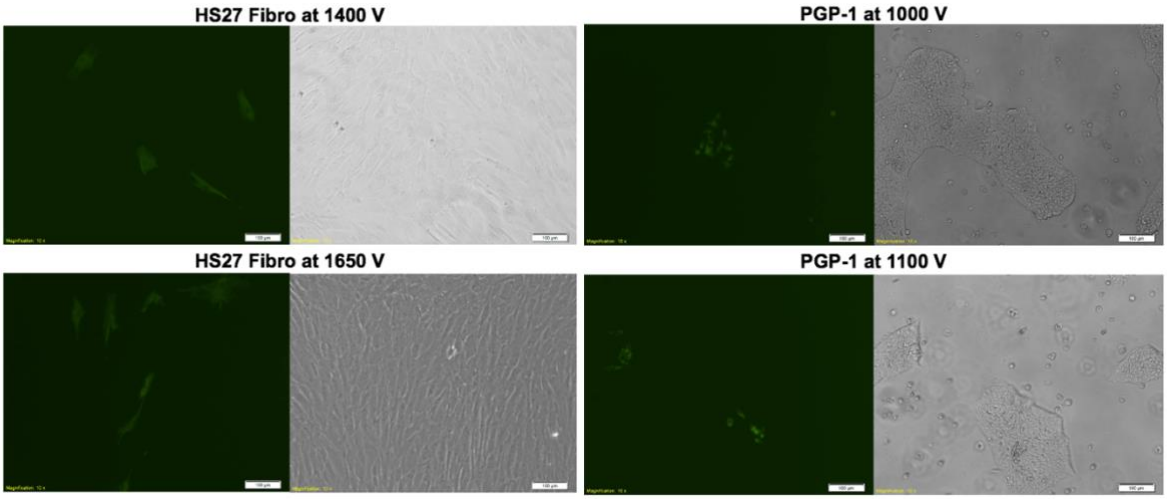

Figure 13: NEON® transfected HS27 Fibroblasts and PGP-1 iPSC – Both cell lines show good recovery after transfection with different voltages. The few green cells indicate that the transfection worked within these cells and the Actin was tagged with GFP. Scale bar represents 1000 µm.

After cell recovery, an appropriate amount of puromycin was added to kill cells without the fluorescence tag. As shown in Figure 14, a large amount of the untagged cells were already dead after few days with puromycin and only the green fluorescent cells remained. These results indicate that the targeting with the TrueTag system worked well for Actin. The fibroblasts, which were no longer needed, were frozen.

For the PGP-1 cells, several colonies were picked and expanded for further characterization. For most of the colonies, attachment problems occurred and further expansion was not possible. Since the number of colonies was low, the experiment needs to be repeated.

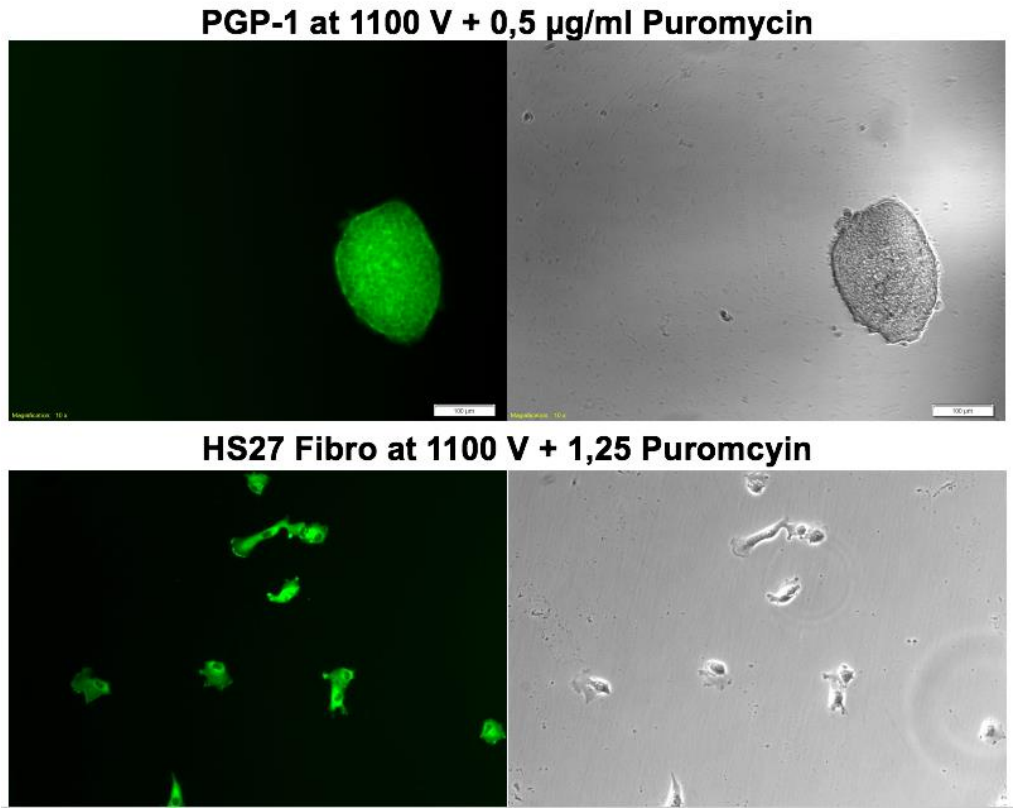

Figure 14: NEON® transfected HS27 Fibroblasts and PGP-1 - iPSC after puromycin selection (0,5 µg/ml for PGP-1 and 1,25 µg/ml for HS27 Fibro). Scale bar represents 1000 µm.

3.4 Experimental design for targeting *DLX5* with a dsDNA template

Figure 15 shows the construction for the targeting of the *DLX5* gene. The first steps will be to create a double stranded donor template which contains the GFP as a reporter and T2A for cleavage. This will be done with amplification by PCR of a region of the large plasmid template used for the targeting of *DLX5*. The designed forward and reverse primers with short HA will allow an easy construction of the donor molecule, where the HA are added by PCR to a pre-made tagging template. The use of the PCR product without the unnecessary vector backbone minimizes the size of the donor, eliminates the risk of toxic *E.coli* elements being integrated into the cell and should result in higher editing efficiency. The correct integration of the donor template to the genomic target is accomplished by HDR using the matching HA on each side of the tag. A DSB will be generated close to the DLX5 STOP codon using Cas9/gRNA and then the DSB will be repaired by the natural HDR pathway. Together with the optimized NEON® transfection conditions, a transfection of the PGP-1 cells (70%

confluency) with both the RNP and the generated dsDNA HDR template can be performed.

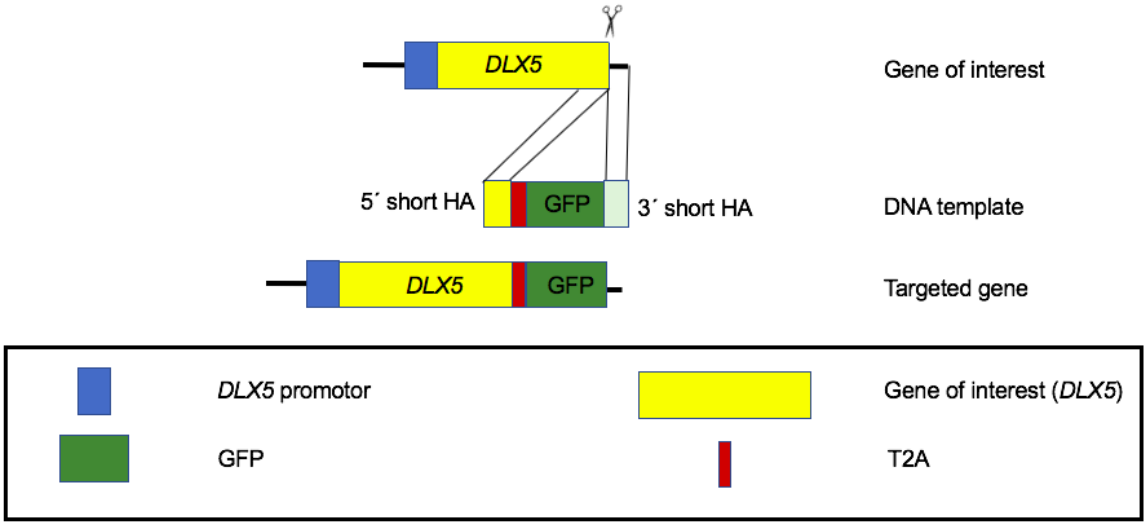

Figure 15: Construction of the *DLX5* gene targeting - Creation of the double stranded donor template which contains the GFP as a reporter and T2A for cleavage. The HA were added by PCR to a pre-made tagging template.

3.5 Split GFP

Good results in the production of gene reporter lines have been shown with the use of split GFP (Yang et al., 1996).The GFP1-10 part needs to be constitutively expressed in iPSC. Our collaborator for this project, introduced the GFP1-10 into the AAVS1 locus of a PSC line. The GFP11 part can then be introduced as a tag for the gene of interest using a HDR donor template. When both GFP parts are expressed, green fluorescent light can be observed (Leonetti et al., 2016). For this project, we are planning to target two genes: *TBXT* and *FOXA2*. For both genes, we need to design a precise gRNA for cutting and a ssODN for targeting.

3.5.1 gRNA design

To enable CRISPR cutting, the gRNA was designed with the use of the CRISPOR web tool. Two gRNAs for each gene were designed, with the second gRNA serving as a backup in the event that the first gRNA does not work well. (Table 19 and Table 20) The specificity score, the predicted efficiency and the off targets in the genome of interest were compared. The specificity score ranges from 0-100 and measures the

uniqueness of the guide in the genome. Higher specificity scores are desirable because this indicates a low number of off-target effects. In addition to high specificity scores, high efficiency scores are recommended as gRNAs with higher scores are more likely to cleave at the specific position.

Table 19: gRNAs for *TBXT*

Strand	gRNA	PAM	MIT Specificity score	Efficiency score	Off targets for 0-1-2-3-4 mismatches
forward	Table 4 No. 1	AGG	65	75	0-0-7-23-212
forward	Table 4 No. 2	AGG	52	52	0-0-4-29-248

For *TBXT,* gRNA 5´-ACCTTCCATGTGAAGCAGCA-3´ was chosen. In Figure 16A, the gRNA, the STOP codon and the PAM site are marked within the sequence. The expected cleavage position is located -3bp 5´of the PAM site. This allows the expected cleavage to occur after the stop codon, avoiding disruption of the coding gene (if one allele is targeted with GFP and the other has indels). With an MIT specificity score of 65 and an efficiency score of 75, both values are in a good range for the chosen gRNA. The number of off-targets with two mismatches is higher compared to the alternative gRNA but other good values made this gRNA the first choice (Table 19). In the event that the real cutting efficiency of this gRNA is lower than the expected efficiency, another gRNA was designed as a backup. As an alternative, gRNA 5´-CATGTGAAGCAGCAAGGCCC-3´ was chosen. As shown in Figure 16B, the cut is 10bp after the stop codon which is not optimal as we want to cut as close as possible to the STOP codon to increase the HDR efficiency. Both the MIT specificity score and the efficiency score are lower compared to the other gRNA (Table 19). Overall, this gRNA is not ideal but could serve as a good alternative if the selected gRNA does not work as expected.

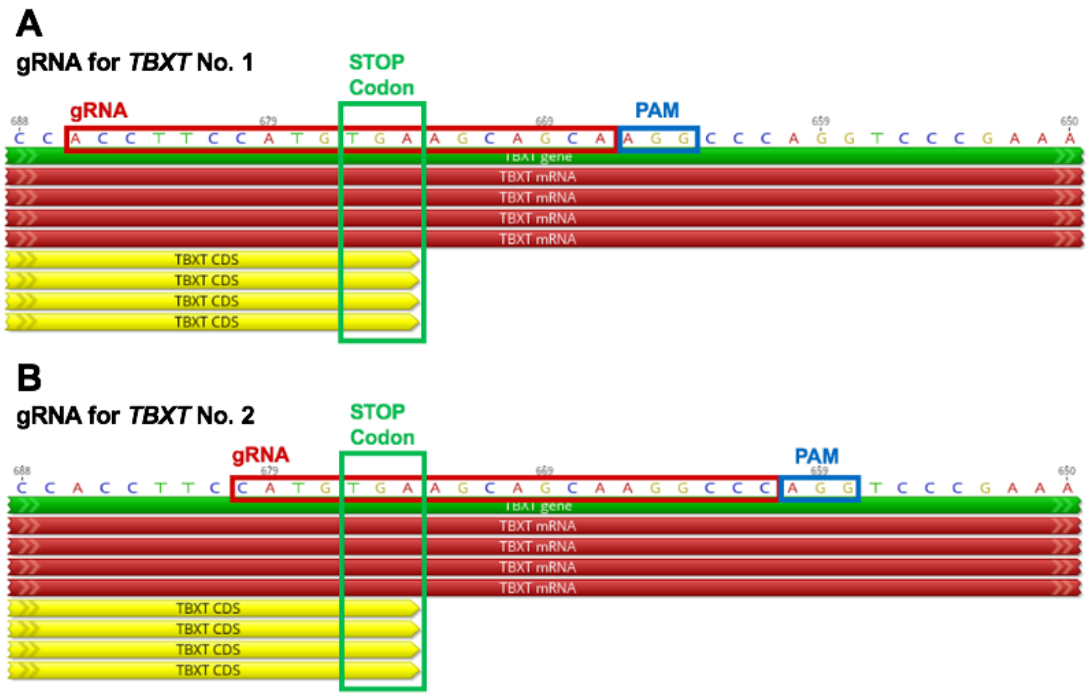

Figure 16: gRNA comparison for *TBXT* – Part of the *TBXT* gene with (A) first chosen gRNA and (B) an alternative gRNA. The red column indicates the corresponding mRNA and the yellow column the protein. The gRNA, stop codon and PAM region are marked with colored boxes (gRNA: red, STOP Codon: green and PAM: blue)

Table 20: gRNAs for *FOXA2*

Strand	gRNA	PAM	MIT Specificity score	Efficiency score	Off targets for 0-1-2-3-4 mismatches
forward	Table 4 No. 3	CGG	82	72	0-0-0-5-93
reverse	Table 4 No. 4	AGG	92	55	0-0-1-2-32

For *FOXA2,* gRNA 5´-AACTCCTCTTAAGAAGACGA-3´ was chosen (Figure 17A). As an alternative, gRNA 5´-GAAGCCGTCGTCTTCTTAAG-3´ was selected (Figure 17B). As shown in Figure 17A, the DSB occurs outside the coding region for this gRNA. In comparison, the gRNA in Figure 17B shows the DSB introduced within the stop codon which could lead to wrong transcription. This gRNA was still used as an alternative because all other potential gRNAs are not usable for this application. Some of them had a very low efficiency score or a high number of off target mismatches, while others

are just too far away from the target sequence. The MIT specificity score is very good and the efficiency score is tolerable. The gRNA with zero off-targets with two mismatches is preferred over the other gRNA where one off-target was observed because less off targets are always better. (Table 20) However, when all criteria are considered, the gRNA located on the forward strand would be the better choice if the cutting efficiency is as expected.

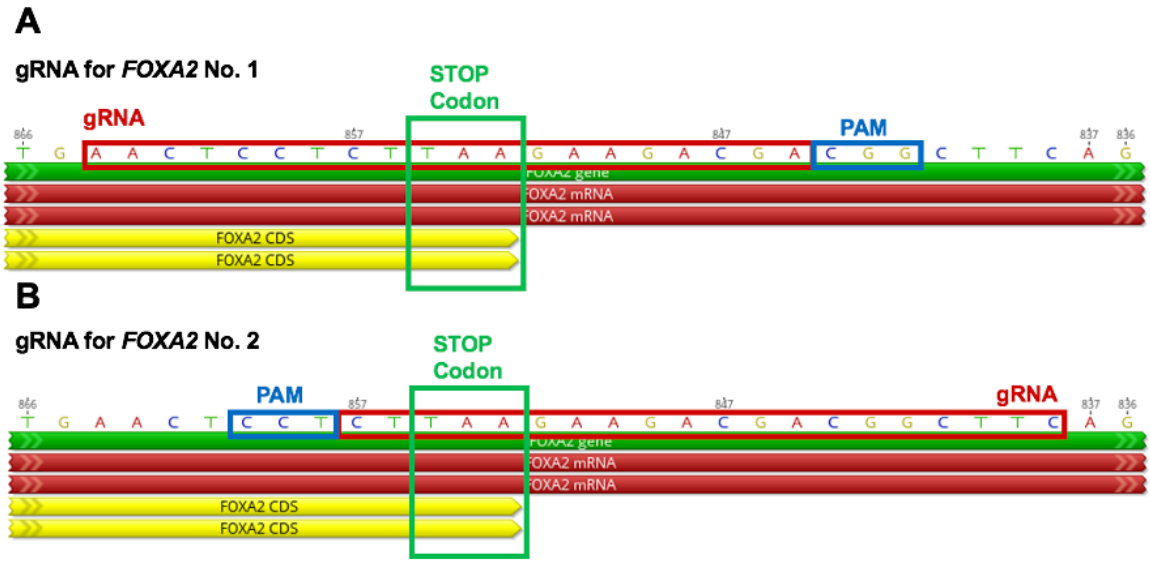

Figure 17: gRNA comparison for *FOXA2* - Part of the *FOXA2* gene with (A) first chosen gRNA and (B) an alternative gRNA. The red column indicates the corresponding mRNA and the yellow column the protein. The gRNA, stop codon and PAM region are marked with colored boxes (gRNA: red, STOP Codon: green and PAM: blue)

3.5.2 ssODN design

For targeting of *TBXT* and *FOXA2,* ssODN were designed. For this project, the goal is to fuse GFP to the C-terminal part of the proteins of interest. To allow HDR, a HA ~ 60 nt long was needed on each side of the STOP codon. For targeting, the 16 amino acid encoding GFP11 sequence (5´ ACCGAGCTCAACTTCAAGGAGTGGCAAAAGGCCTTTACCGATATGATG 3´) was included between the HAs. It has been shown that a lack of GFP detection could originate from the lack of physical accessibility to the GFP11 tag for complementation with GFP1–10 (Leonetti et al., 2016). To overcome this, an 10 amino acid linker (GGT GGT GGT GGT TCT GGT GGT GGT GGT TCT) was included between the targeted protein and the GFP11 sequence.

In Figure 18, the cutting site of the chosen gene is shown together with the linker - GFP11 sequence. The linker sequence is indicated with a blue color and the GFP11 sequence with green. To avoid a second binding and cutting of the gRNA, a mutation was introduced into the PAM (NGG) region. The mutation is indicated in red. For the two gRNAs for the *TBXT* gene and the first gRNA for the *FOXA2* gene only one guanine nucleotide is changed. For the second gRNA for the *FOXA2* gene, since the DSB occurs in the coding region, we need to introduce a silent mutation. The codon TCC can be replaced by TCA.

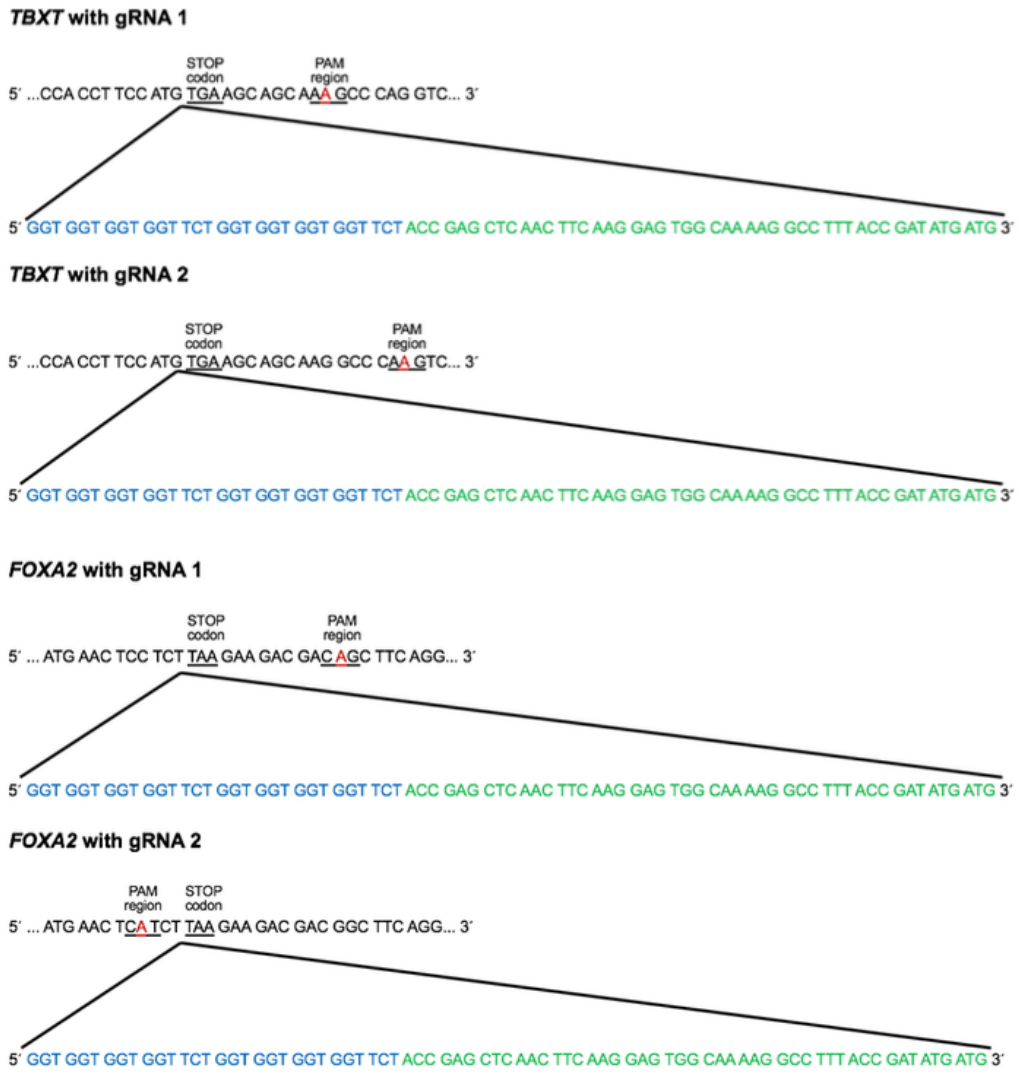

Figure 18: Cutting site of *TBXT* and *FOXA2* - The cleavage position of the chosen genes is shown with different gRNAs. The expected cleavage position is located -3bp 5´of the PAM site. The linker sequence (blue) and the GFP11 sequence (green) get introduced at this position. To avoid a second cutting, a silent mutation get introduced into the PAM region (red).

4 Discussion

4.1 gRNA efficiency and optimized transfection efficiency

From the data shown in Figure 11, higher voltages lead to higher cell death. This toxicity could be due to the NEON® electroporation itself or could be induced in response to the DNA repair pathway activation. The results indicate that the old gRNA has a low efficiency. (Figure 10) Despite the low efficiency, the old gRNA was used for the optimization of the NEON® transfection condition. Through testing of different conditions, it was shown that lower concentrations of Cas9 (2 pmol) and gRNA (4 pmol) were optimal for cell viability. However, no clones were obtained after puromycin selection, suggesting that the cutting efficiency had to be improved. To increase the cutting efficiency, a new gRNA need to be ordered and tested.

4.2 Test of the dsDNA HDR donor system

ACTB was targeted in fibroblasts and iPSCs to show that the TrueTag (dsDNA HDR donor template) system worked as expected. As shown in Figure 13, the transfection worked well in both cell types (HS27 Fibro and PGP-1 iPSC). The tagged cells were then selected with puromycin. After selection of the fibroblasts, a pool of Actin-GFP tagged cells were obtained. The filamentous actin structure could easily be visualized under the fluorescent microscope at higher magnification. For PGP-1, the method was not efficient. The few green colonies obtained after puromycin selection were picked and further expanded. However, these cells were difficult to expand as they displayed attachment problem. This problem was reported previously when the iPS Core Facility first attempted this experiment. The attachment issue might be due to the fusion of GFP to the actin gene. GFP might interfere with the normal function of the Actin protein. To overcome this hurdle, further investigations are needed and the overall experiment needs to be repeated.

4.3 Targeting *DLX5* with a dsDNA HDR donor

The design for targeting the DLX5 gene with a dsDNA HDR donor was performed. (Figure 12) The next step for this project will be the creation of the HDR donor template. To increase HDR, a DSB is generated near the START codon (for N-terminal targeting)

or STOP codon (for C-terminal targeting) with the Cas9 nuclease directed to the desired sequence by the gRNA. The natural HDR pathway will then repair the genomic DNA at the site of the DSB. The NEON® transfection system seems to be the best method to efficiently introduce the Cas9/gRNA/DNA template into the iPSC. This method is generally used for maximum efficiency in difficult-to-transfect cell types (Brees & Fransen, 2014). The day post transfection, the survival must be evaluated by imaging. This will give more information about the toxicity of the template.

Next, the targeting efficiency needs to be tested. Since there is no drug selection cassette in the donor template, the targeted cells cannot be positively selected. An estimation of the efficiency can be first tested in the pool of cells by PCR. If we can detect cells with the correct targeting in the pool, we will proceed with single cell cloning. Single colonies will then be picked to verify that the integration occurred in the correct site. This will be done by PCR using primers in the tag and outside the HA. The region encompassing the 5' HA and 3' HA should be amplified. Only the colonies with the correct targeting will be maintained.

For the next step, we would like to verify if one or two alleles were targeted. This can be achieved by amplification of the region encompassing the STOP codon of *DLX5*. If this region cannot be amplified, most likely both alleles were targeted signifying a homozygous gene reporter line. If it can be amplified, the result would indicate a heterozygous gene reporter line. The PCR product will also be sequenced to detect any INDELS at the site of the DSB.

Finally, it would be necessary to check for any off-target insertions. A ddPCR to assess the copy number of the GFP gene will be performed. This number should be one for heterozygous targeting or two for homozygous targeting. Colonies with additional copies of GFP, most likely due to off target integration, will be discarded (Zhang, Tee, Wang, Huang, & Yang, 2015).

After evaluation of all the important parameters, this system can be compared with other targeting strategies for example using a plasmid as HDR donor. This method is expected to achieve higher efficiencies since the DNA template is small and should be easy to transfect. Moreover, this system offers a very simple way to generate DNA template without any cloning steps. It should also allow the reduction of toxicity

because of the small size of the template. Altogether, because of the facts mentioned above this method could remodel the current method.

4.4 Split GFP

To overcome the hurdle of the long ssDNA donors, a new idea involving split GFP was introduced. The key advantage of this method is the small size of the GFP11 (16 amino acids) which allows ssDNA templates to be used as HDR donors (Leonetti et al., 2016). The method has already been tested with a set of 48 human genes in 293T$_{GFP1-10}$ cells. All of the proteins had a specific subcellular localization to allow an easy analysis of GFP$_+$ cells with a microscope. In 30 (63 %) of the targeted genes, a clear population of GFP$_+$ cells was shown. An important fact which needs to be considered is that the targeted proteins spanned a wide range of native expression level and therefore influence the intensity of GFP fluorescence. Through ribosome profiling, it was shown that GFP fluorescence and protein abundance are well correlated (Leonetti et al., 2016). The two genes we are going to target, *TBXT* and *FOXA2* are activated during the early stage of development. *TBXT* is specific of mesodermal differentiation and *FOXA2*, specific for endoderm differentiation. Although these genes are not expressed in undifferentiated pluripotent stem cells, their activation occurs quickly upon differentiation. Therefore, the functionality of the targeting can be tested within 48h of differentiation.

The next steps in the overall experiment will be the targeting of *TBXT* and *FOXA2* in PGP-1$_{GFP1-10}$. Two potential gRNAs were designed for each gene with the online tool CRISPOR. In addition, ssODNs were designed with the use of Geneious Prime software. The ssODNs were designed for each gRNA. They include a ten amino acid linker sequence, the GFP11 sequence and a HA of 60 nt on each side.

The gRNAs will be ordered and tested in H1 hESC lines which constitutively express GFP1-10. After establishment of an efficient cutting efficiency, a NEON® transfection with the selected ssODN will be performed. Through expression of both GFP parts, functional GFP tagging occurs (Figure 4). The intensity of the split GFP complex should be identical to the full-length GFP (Kamiyama et al., 2016). Compared to the dsDNA HDR template, the targeting efficiency, off-target integration and the number of targeted alleles needs to be tested to ensure correct targeting.

The generation of gene reporter lines utilizing this new method would help the iPS Core Facility improve their current gene editing workflow and would help to improve different gene editing methods in general.

4.5 Gene Reporter Lines

The discovery of human iPSC revolutionized the field of stem cell-based regenerative medicine (Rao & Malik, 2012). New tools, including the CRISPR/Cas9 technology, have greatly facilitated the genome modification of these human iPS lines. Consequently, many human iPSC gene reporter lines have been created in recent years. This is done by stably integrating a fluorescent gene to allow monitoring of specific genes. The Allen Institute for Cell Science (AICS) has established a large set of gene reporter lines in human induced PSC. (Table 21). The list includes the name of the gene, the structure, the fluorescent protein, the chromosomal location and the passage number of the cell line. All these lines are available for distribution. They provide a unique tool to study cell biology. There are also many additional gene reporter lines described in the literature. Examples for endodermal (*NKX6.1*), ectodermal (*NEUROG2*) and mesodermal (*MYH6*) lineage markers are shown in Table 22. (Gupta et al., 2018) (S. Li et al., 2015) (Tsai et al., 2020)

Table 21: Partial List of Gene Reporter Lines from the AICS

Release date	**Line #**	**Gene Name**	**Structure**	**FP tag**	**Chromosome location (transgene)**	**Passage**
27.03.17	AICS-0016-184	*ACTB*	Actin filaments	mEGFP	7p22.1	37
08.06.18	AICS-0037-172	*TNNI1*	Sarcomeric thin filament	mEGFP	1q32.1	34
31.08.18	AICS-0052-003	*MYL7*	Sarcomeric thick filaments	mEGFP	7p13	39

10.10.18	AICS-0060-027	*MYL2*	Sarcomeric thick filaments	mEGFP	3q27.3	37
30.04.19	AICS-0074-026	*SOX2*	Transcription factor	mEGFP	3q26.33	29

Table 22: Partial List of Gene Reporter Lines in literature

Gene Name	Structure	FP tag	Chromosome location (transgene)
NKX6.1	Transcription factor	mEGFP	4q21.23
MYH6	thick filaments of the sarcomere	mCherry	14q11.2
NEUROG2	Transcription factor	mCherry	4q25

Here is the list of gene reporter lines currently under construction at the HSCI iPS core facility. (Table 23)

Table 23: New Gene Reporter Lines from iPS core facility

Gene Name	Structure	FP tag	Chromosome location (transgene)	Passage
DLX5	Transcription factor	mEGFP	7q21.3	Unknown yet
TBXT	Transcription factor	mEGFP	6q27	Unknown yet
FOXA2	Transcription factor	mEGFP	20p11.21	Unknown yet

List of References

Adli, M. (2018). The CRISPR tool kit for genome editing and beyond. *Nat Commun, 9*(1), 1911. doi: 10.1038/s41467-018-04252-2

Baker, O., Tsurkan, S., Fu, J., Klink, B., Rump, A., Obst, M., . . . Stewart, A. F. (2017). The contribution of homology arms to nuclease-assisted genome engineering. *Nucleic Acids Res, 45*(13), 8105-8115. doi: 10.1093/nar/gkx497

Bortesi, L., Zhu, C., Zischewski, J., Perez, L., Bassie, L., Nadi, R., . . . Schillberg, S. (2016). Patterns of CRISPR/Cas9 activity in plants, animals and microbes. *Plant Biotechnol J, 14*(12), 2203-2216. doi: 10.1111/pbi.12634

Brees, C., & Fransen, M. (2014). A cost-effective approach to microporate mammalian cells with the Neon Transfection System. *Anal Biochem, 466*, 49-50. doi: 10.1016/j.ab.2014.08.017

Butler, J., Fonarow, G. C., Zile, M. R., Lam, C. S., Roessig, L., Schelbert, E. B., . . . Gheorghiade, M. (2014). Developing therapies for heart failure with preserved ejection fraction: current state and future directions. *JACC Heart Fail, 2*(2), 97-112. doi: 10.1016/j.jchf.2013.10.006

Cabantous, S., Terwilliger, T. C., & Waldo, G. S. (2005). Protein tagging and detection with engineered self-assembling fragments of green fluorescent protein. *Nat Biotechnol, 23*(1), 102-107. doi: 10.1038/nbt1044

Gupta, S. K., Wesolowska-Andersen, A., Ringgaard, A. K., Jaiswal, H., Song, L., Hastoy, B., . . . Honore, C. (2018). NKX6.1 induced pluripotent stem cell reporter lines for isolation and analysis of functionally relevant neuronal and pancreas populations. *Stem Cell Res, 29*, 220-231. doi: 10.1016/j.scr.2018.04.010

Hall, B., Limaye, A., & Kulkarni, A. B. (2009). Overview: generation of gene knockout mice. *Curr Protoc Cell Biol, Chapter 19*, Unit 19 12 19 12 11-17. doi: 10.1002/0471143030.cb1912s44

Han Li, K. A. B., Veronica Pessino, Bo Huang, Jonathan S. Weissman, Manuel D. Leonetti. (2019). Design and specificity of long ssDNA donor for CRISPR-based knock-in. doi:https://doi.org/10.1101/178905

Hsu, P. D., Lander, E. S., & Zhang, F. (2014). Development and applications of CRISPR-Cas9 for genome engineering. *Cell, 157*(6), 1262-1278. doi: 10.1016/j.cell.2014.05.010

Invitrogen. (2019). *TrueTag™ DNA Donor Kit. User Guide*. (MAN0018537).

Jasin, M., & Haber, J. E. (2016). The democratization of gene editing: Insights from site-specific cleavage and double-strand break repair. *DNA Repair (Amst), 44*, 6-16. doi: 10.1016/j.dnarep.2016.05.001

Jurgielewicz, P., Harmsen, S., Wei, E., Bachmann, M. H., Ting, R., & Aras, O. (2017). New imaging probes to track cell fate: reporter genes in stem cell research. *Cell Mol Life Sci, 74*(24), 4455-4469. doi: 10.1007/s00018-017-2584-z

Kamiyama, D., Sekine, S., Barsi-Rhyne, B., Hu, J., Chen, B., Gilbert, L. A., . . . Huang, B. (2016). Versatile protein tagging in cells with split fluorescent protein. *Nat Commun, 7*, 11046. doi: 10.1038/ncomms11046

Klein, R., Ruttkowski, B., Knapp, E., Salmons, B., Gunzburg, W. H., & Hohenadl, C. (2006). WPRE-mediated enhancement of gene expression is promoter and cell line specific. *Gene, 372*, 153-161. doi: 10.1016/j.gene.2005.12.018

Leonetti, M. D., Sekine, S., Kamiyama, D., Weissman, J. S., & Huang, B. (2016). A scalable strategy for high-throughput GFP tagging of endogenous human proteins. *Proc Natl Acad Sci U S A, 113*(25), E3501-3508. doi: 10.1073/pnas.1606731113

Li, S., Chen, L. X., Peng, X. H., Wang, C., Qin, B. Y., Tan, D., . . . Zhou, X. H. (2018). Overview of the reporter genes and reporter mouse models. *Animal Model Exp Med, 1*(1), 29-35. doi: 10.1002/ame2.12008

Li, S. P., Feng, J. J., Wang, H. G., Wang, X. F., & Lv, Z. J. (2012). [The effects of SV40 PolyA sequence and its AATAAA signal on upstream GFP gene expression and transcription termination]. *Yi Chuan, 34*(1), 113-119. doi: 10.3724/sp.j.1005.2012.00113

Li, X. L., Li, G. H., Fu, J., Fu, Y. W., Zhang, L., Chen, W., . . . Zhang, X. B. (2018). Highly efficient genome editing via CRISPR-Cas9 in human pluripotent stem cells is achieved by transient BCL-XL overexpression. *Nucleic Acids Res, 46*(19), 10195-10215. doi: 10.1093/nar/gky804

Liu, M., Rehman, S., Tang, X., Gu, K., Fan, Q., Chen, D., & Ma, W. (2018). Methodologies for Improving HDR Efficiency. *Front Genet, 9*, 691. doi: 10.3389/fgene.2018.00691

Manzini, S., Viiri, L. E., Marttila, S., & Aalto-Setala, K. (2015). A Comparative View on Easy to Deploy non-Integrating Methods for Patient-Specific iPSC Production. *Stem Cell Rev Rep, 11*(6), 900-908. doi: 10.1007/s12015-015-9619-3

Miyaoka, Y., Mayerl, S. J., Chan, A. H., & Conklin, B. R. (2018). Detection and Quantification of HDR and NHEJ Induced by Genome Editing at Endogenous Gene Loci Using Droplet Digital PCR. *Methods Mol Biol, 1768*, 349-362. doi: 10.1007/978-1-4939-7778-9_20

Murgha, Y. E., Rouillard, J. M., & Gulari, E. (2014). Methods for the preparation of large quantities of complex single-stranded oligonucleotide libraries. *PLoS One, 9*(4), e94752. doi: 10.1371/journal.pone.0094752

Prowse, A. B., Timmins, N. E., Yau, T. M., Li, R. K., Weisel, R. D., Keller, G., & Zandstra, P. W. (2014). Transforming the promise of pluripotent stem cell-

derived cardiomyocytes to a therapy: challenges and solutions for clinical trials. *Can J Cardiol, 30*(11), 1335-1349. doi: 10.1016/j.cjca.2014.08.005

Rao, M. S., & Malik, N. (2012). Assessing iPSC reprogramming methods for their suitability in translational medicine. *J Cell Biochem, 113*(10), 3061-3068. doi: 10.1002/jcb.24183

Roberts, B., Hendershott, M. C., Arakaki, J., Gerbin, K. A., Malik, H., Nelson, A., . . . Gunawardane, R. N. (2019). Fluorescent Gene Tagging of Transcriptionally Silent Genes in hiPSCs. *Stem Cell Reports, 12*(5), 1145-1158. doi: 10.1016/j.stemcr.2019.03.001

Roth, T. L., Puig-Saus, C., Yu, R., Shifrut, E., Carnevale, J., Li, P. J., . . . Marson, A. (2018). Reprogramming human T cell function and specificity with non-viral genome targeting. *Nature, 559*(7714), 405-409. doi: 10.1038/s41586-018-0326-5

Ryu, S. M., Hur, J. W., & Kim, K. (2019). Evolution of CRISPR towards accurate and efficient mammal genome engineering. *BMB Rep, 52*(8), 475-481.

Samuel, R., Daheron, L., Liao, S., Vardam, T., Kamoun, W. S., Batista, A., . . . Jain, R. K. (2013). Generation of functionally competent and durable engineered blood vessels from human induced pluripotent stem cells. *Proc Natl Acad Sci U S A, 110*(31), 12774-12779. doi: 10.1073/pnas.1310675110

Skarnes, W. C., Pellegrino, E., & McDonough, J. A. (2019). Improving homology-directed repair efficiency in human stem cells. *Methods, 164-165*, 18-28. doi: 10.1016/j.ymeth.2019.06.016

Sollu, C., Pars, K., Cornu, T. I., Thibodeau-Beganny, S., Maeder, M. L., Joung, J. K., . . . Cathomen, T. (2010). Autonomous zinc-finger nuclease pairs for targeted chromosomal deletion. *Nucleic Acids Res, 38*(22), 8269-8276. doi: 10.1093/nar/gkq720

Takahashi, K., & Yamanaka, S. (2006). Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors. *Cell, 126*(4), 663-676. doi: 10.1016/j.cell.2006.07.024

Till, J. E., & Mc, C. E. (1961). A direct measurement of the radiation sensitivity of normal mouse bone marrow cells. *Radiat Res, 14*, 213-222.

Tsai, S. Y., Ghazizadeh, Z., Wang, H. J., Amin, S., Ortega, F. A., Badieyan, Z. S., . . . Chen, S. (2020). A human embryonic stem cell reporter line for monitoring chemical-induced cardiotoxicity. *Cardiovasc Res, 116*(3), 658-670. doi: 10.1093/cvr/cvz148

Vazin, T., & Freed, W. J. (2010). Human embryonic stem cells: derivation, culture, and differentiation: a review. *Restor Neurol Neurosci, 28*(4), 589-603. doi: 10.3233/RNN-2010-0543

Wobus, A. M., & Boheler, K. R. (2005). Embryonic stem cells: prospects for developmental biology and cell therapy. *Physiol Rev, 85*(2), 635-678. doi: 10.1152/physrev.00054.2003

Yang, F., Moss, L. G., & Phillips, G. N., Jr. (1996). The molecular structure of green fluorescent protein. *Nat Biotechnol, 14*(10), 1246-1251. doi: 10.1038/nbt1096-1246

Zhang, X. H., Tee, L. Y., Wang, X. G., Huang, Q. S., & Yang, S. H. (2015). Off-target Effects in CRISPR/Cas9-mediated Genome Engineering. *Mol Ther Nucleic Acids, 4*, e264. doi: 10.1038/mtna.2015.37