Applied Stem Cell Technologies - AST

University of Twente

Bachelor assignment

Active at AST
19 April 2021 - 27 June 2021, 10 weeks

CRISPR-Cas genome engineering explained from A to T

Sem A. Sterckel (s1916270)

Understanding CRISPR genome engineering via a rainbow human embryonic stem cell reporter line to identify pacemaker cells and a MEF2c construct production

Commission
Chairman - Prof. Dr. Robert Passier
Daily Supervisor - Dr. Verena Schwach
Outer Member - Dr. Jeroen Leijten
Additional Member - Rolf Slaats MSc.

Version of June 28, 2021

Abstract

The CRISPR/Cas research domain is a fast expanding domain with lots of potential to open up many doors in other research fields. Using a guideRNA (gRNA) and a Cas9 endonuclease precise double stranded breaks (DSB) can be made in the DNA of cells. The repair mechanism, homology directed repair (HDR), is able to then insert sequences with a size of multiple kilobases (kb). In this thesis a HDR vector targeting the *AAVS1* gene for the insert of a MEF2C DNA binding construct with the ability to promote transcription of a mRuby sequence was produced. This construct, which would be able to detect and indicate functional MEF2C proteins, was made with the use of a sticky end PCR technique. Also a human embryonic stem cell (hESC) dual reporter line, expressing mCherry from the *COUP-TFII* gene and GFP from the *NKX2.5* gene, was targeted for the fusion of the fluorescent protein mTAGBFP with the transcription factor SHOX2 via precise genome editing. With SHOX2 being a perfect reporter for functional pacemaker cells. The MEF2C detecting construct will in future experiments be used to discover underlying mechanisms of MEF2C related diseases like cardiac hypertrophy and congenital heart disease. And the cells targeted with the mTAGBFP insert are to be made into a stable rainbow reporter line allowing for standardisation of pacemaker cell differentiation, selection of pacemaker cells, and the possible development of deployable biological pacemakers.

Abstract

Het CRISPR/Cas onderzoeksgebied is een snel groeiend terrein met veel potentie om deuren te openen in andere vakgebieden. Met het gebruik van een guideRNA en een Cas9 endonuclease kunnen exacte dubbelstrengsbreuken gemaakt worden in het DNA van cellen. Het reparatie mechanisme, homologie geleide reparatie (HGR), maakt het mogelijk om vervolgens een sequentie van meerdere kilobasen in te voegen. In dit proefschrift was een HGR vector gemaakt, welke de toevoeging van een MEF2C DNA bindende constructie - met het vermogen om de transcriptie van een mRuby sequentie te promoten - aan het *AAVS1* gen mogelijk maakt. Deze constructie, welke het vermogen heeft om functionele MEF2C eiwitten te detecteren en aan te tonen, was gemaakt door middel van een sticky end PCR techniek. Ook was een menselijke embryonale stamcellen dubbele reporter cellijn, met de expressie van mCherry vanaf het *COUP-TFII* gen en de expressie van GFP vanaf het *NKX2.5* gen, getarget voor de fusie van het fluorescente eiwit mTAGBFP met de transcriptie factor SHOX2 door middel van exacte genetische manipulatie. Met SHOX2 als perfecte reporter voor functionele pacemaker cellen. Het MEF2C detecterende construct zal in toekomstige experimenten gebruikt worden om de onderliggende mechanismen van MEF2C gerelateerde ziekten zoals cardiale hypertrofie en aangeboren hartziekte. En de cellen, die getarget zijn met de mTAGBFP invulling, zullen worden gemaakt tot een stabiele regenboog reporter cellijn, wat de standarisatie van pacemaker cel differentiatie, de selectie van pacemaker cellen, en de mogelijke ontwikkeling van inzetbare biologische pacemakers mogelijk kan maken.

Acknowledgements

I would like to thank Verena Schwach and Rolf Slaats for their guidance in my process through this thesis. Also their constructive feedback helped me improve the substantive depth of this thesis. Furthermore, I would like to thank Jeroen Leijten en Robert Passier for their advice during the interim presentation. Carla C. Fabres for her help with sorting and imaging the cells. And Gerben ten Hag for his advice in the smaller technical details and for learning me the tricks of the trade.

Contents

1 Introduction

1.1 Genome editing

When a double stranded break (DSB) occurs in the DNA of a cell it has two main ways of damage repair [1]: the first is via nonhomologous end joining (NHEJ) and the second is via homology-directed repair (HDR) (as can be seen in figure 1. The latter needs a repair template and has a high fidelity outcome. And the former can happen without a template and is prone for errors. The DSB will be re-ligated via NHEJ, however small insertions or deletions (indels) can take place during this process [2]. These indels in exons can cause frameshift mutations, which can cause incorrect translation and premature stop codons. The gene will be knocked out if a premature stop codon occurs.. With multiple DSBs large deletions in the genome, like the deletion of an exon, can even take place, due to NHEJ leaving out the middle part [3].

For the cell to have less chance on repairing with possible indels via NHEJ the cell also uses HDR [4]. A DNA template, either in double-stranded DNA (dsDNA) form or single-stranded DNA (ssDNA), is used as a template to repair the DNA with a DSB. This system can be used to make single bp changes in the DNA with the use of exogenously introduced DNA containing of up and downstream homologous arms of at around 50-60 bp per arm [5]. Oftentimes ssDNA oligonucleotides (ssODNs) are used for inducing point mutations via HDR [6]. When using this technique, the efficiency of Cas9/HDR can be increased by using ssODNS complementary to the non-target strand, and also homology arms asymmetric in length can improve efficiency [6].

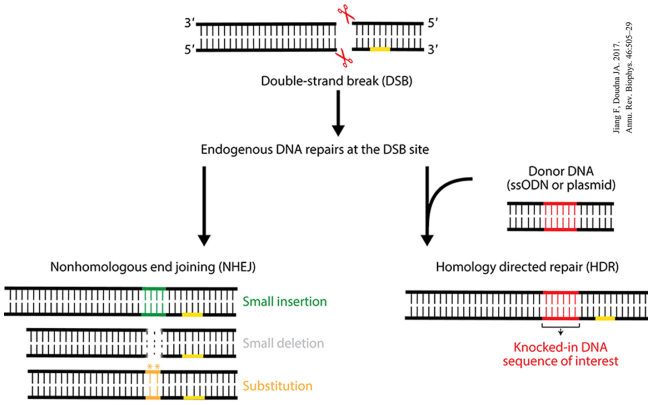


Figure 1: Schematically the two main DSB damage repair ways in cells [1].

With HDR even larger insertions can be made, making it possible to insert new genes or add onto already existing genes. Inserts into the genome as large as 7.4 kb have been reported [7], but these larger inserts oftentimes lead to lower efficiencies [8]. For the insertion of small 100 bp to medium 700 bp fragments via Cas9/HDR homology arms of 50 [5] to 300 bp [9] respectively are sufficient. Extending the length of the homology arms can compensate for the lower efficiency when inserting larger fragments [8], however there is a limit on the enhancing effects of longer arms on the efficiency [7].

The latest gene editing techniques for site specific DSBs in the DNA were zinc finger nucleases (ZFN) [10] and transcription activator-like effector nucleases (TALENs) [11]. Both techniques used a series of peptides, which would recognise and bind to specific sequences of DNA, with a FokI endonuclease on the end of the series [2]. The cost of production and the difficulties in construction and validation however remained a hinder for the widespread usage of these endonuclease techniques [4], [12]. The novel RNA guided DNA endonuclease, CRISPR/Cas9 technique, showed a clear path in more available gene editing, overcoming the difficulties of the previous techniques.

1.2 CRISPR/Cas mechanism

Clustered regularly interspaced short palindromic repeats, also known as CRISPR, were first found in the bacterial genome of Escherichia coli in 1987 [12]. At this CRISPR locus an array consisting of a multiple of a repeating sequence (direct repeats) were detected with random sequences (spacers) in between [13]. These spacers however oftentimes correspond with foreign genetic elements (protospacers), like for example the DNA of bacteriophages. Genes surrounding the CRISPR arrays, which were appropriately called CRISPR associated (Cas) genes, translate into proteins [4]. Three different types of CRISPR loci with multiple Cas genes were known in 2014 and the function of these loci is to form an adaptive immunity for the bacteria by degradation of the protospacer [12]. Bacteria can adapt to viruses and store a copy of the viral DNA in their own DNA as a spacer sequence if it survives a

first attack of this virus (see figure 2). The lineage of this bacterium will then also have this spacer sequence and thus will be more resistant against this viral infection.

The type II CRISPR/Cas was found to cause double stranded breaks (DSBs) in target DNA and later it was discovered that only the Cas9 gene is needed for this defense function [1]. The workings of the type II CRIRSPR/Cas can be explained with only 3 main parts [13]: the Cas operon, the CRISPR array, and the trans-activating CRISPR RNA (tracrRNA). The basics of the mechanism behind CRISPR/Cas can be seen in figure 2. The Cas operon will translate into Cas proteins of which only the Cas9 enzyme is needed for making DSBs [4]. The CRISPR array will translate into RNA consisting of the spacer sequences and the direct repeat sequences. The tracrRNA, which will have certain functional hairpin structures, binds to the direct repeat sequences of the CRISPR array RNA forming a pre-crRNA. After maturation, consisting of RNase III processing and crRNA 5' trimming, the functional crRNA will be ready for usage. For artificial reproduction of crRNA a direct chimeric fusion of these components can be made, called a single-guide RNA (sgRNA) [12].

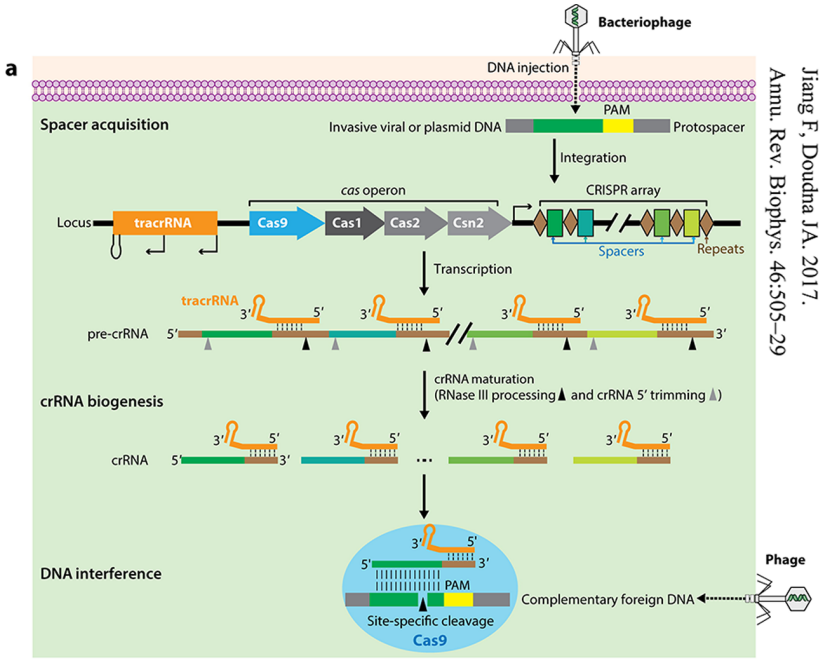


Figure 2: The basic principles of CRISPR/Cas in a schemetical overview [1].

The tracrRNA hairpin part of the guide RNA is for the binding of the Cas9 to the RNA and the spacer sequence is a 20 nucleotide (nt) long sequence for the binding of the RNA to corresponding DNA strands [1]. After the tracrRNA, also called the gRNA scaffold, has bound to the Cas9, the complex will read over the DNA searching for a protospacer adjacent motif (PAM) [14]. This PAM sequence is a Cas9 specific binding spot directly next to the protospacer, meaning that without the correct PAM sequence the complex won't bind to the DNA. Because the PAM sequence is not present in the direct repeat part of the bacterial DNA, the Cas9 enzyme won't cut the domestic DNA [13]. For the canonical PAM of the Cas9 from the Streptococcus pyogenes bacteria this sequence is 5'-NGG-3', where the "N" can be any nucleotide and the "G" should be a guanine nucleotide [12].

So after recognition of the PAM site, local melting of the DNA duplex takes place [14]. With the DNA strand slightly unwound the gRNA stabilises this conformation if the target DNA strand is complementary [1]. An R-loop will be formed with the DNA:gRNA hybrid and the non-targeted strand [15]. In steps, modulated by possible mismatches, the DNA will further be unwound and the strand will be further interrogated by the possible stabilising effect of the gRNA [16]. The further away mismatches are from the PAM site, the less the unwinding and hybridisation will be hindered by these mismatches. The annealing of the gRNA and the target DNA will cause for conformational changes within the Cas9 enzyme [1]. After close to complete annealing, it causes the catalytic HNH domain, the functional domain for cleaving the target strand, to fold over the DNA/RNA hybrid and eventually to make the cleavage [16]. Also due to the unwinding and conformational changes, the

non-target strand will move to the catalytic domain, RuvC, were it will also be cleaved [17]. These two cuts will lead to the DSB, which is three base pairs (bps) upstream of the PAM sequence.

These conformational changes can still take place with target strands which are not 100% complementary, or so called off-target sequences [1]. Up to 4 mismatches in the 20 nt long sequence can be present and the cleavage is still possible [18]. However the exact place of these mismatches and the type of mismatch matter for the gravity of the impact it will have on the energetic favourability of unwinding and thus on the likelihood of the cleavage [16]. Due to the energetic landscape of the R-loop formation within this complex it can be opted for a truncated 17 nt gRNA. These shorter gRNAs are more sensitive in detecting mismatches, but the specificity of the target and also the overall affinity of the binding process will be reduced [18].

1.3 Delivery of Cas9

CRISPR/Cas genome editing thus requires the sgRNA, the Cas9 enzyme and optionally the HDR template in the nucleus of the cell. This could be done by transfecting these items into the cell in their final form, but a more conventional way is by transfecting vectors which transcribe or contain these required items [1]. For example, a vector like pSpCas9-2A-Puro contains the following [5]: a U6 promotor and sgRNA template, which promote transcription of the sgRNA; a CMV or CBh promotor and Cas9 gene, which promotes transcription followed by translation of the Cas9 enzyme; a 2A-linker right after the Cas9 gene and an anti-puromycin gene. Since the 2A-linker, a self-cleaving peptide sequence, is right after the Cas9 gene, transcription and translation will continue and make the puromycin resistance protein [19]. The puromycin resistance will help with the selection of cells which have been successfully transfected and in which the vector has come to expression [5]. Other genes used for selection are fluorescent proteins like mRuby. The Cas9 still needs to enter the nucleus, so an amino acid nuclear localisation sequence (NLS) has been designed to be attached to the enzyme. This sequence will bind to the importin of the nucleus which will help move the enzyme through the nuclear pore [20]. The addition of small sequences and tags to the endonuclease could also be used to improve the genome editing. One study shows that connecting the HDR template as a oligonucleotide to Cas9 can improve the chances of HDR 30 fold compared to Cas9/HDR without the covalent tethering [21].

Moreover, the HDR template can be transfected into the cells in many different forms, as discussed in section 1.1. One of the more conventional forms is having the HDR template as dsDNA in a plasmid [5], since this allows for easy production in large quantities, which will be discussed later. This artificial plasmid is called a vector and will make its way into the nucleus either in relatively small quantities via the nuclear pores or during the reassembly of the nucleus during the mitosis process [22].

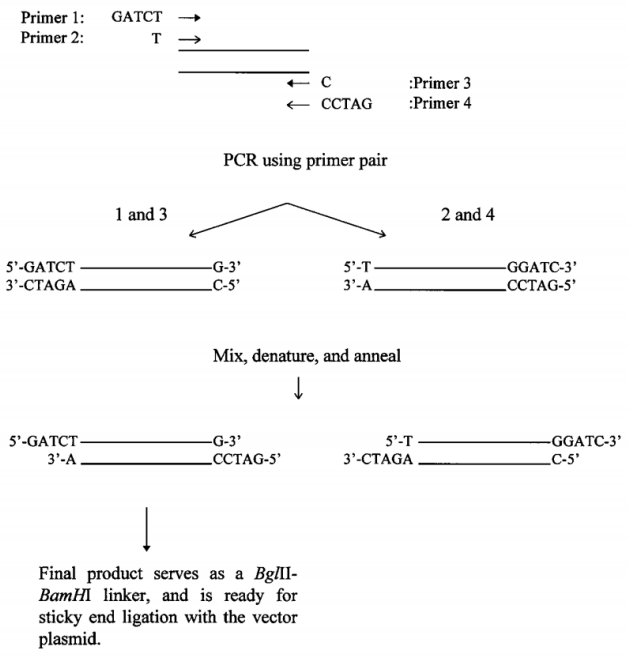


Figure 3: Principles of sticky end PCR.

1.4 Production of vectors

These complicated vectors originated from simple bacterial backbone vectors in which transgene inserts like HDR templates or Cas9 genes have been fused. By taking advantage of bacteria's quick proliferation and tendency to amplify plasmids and vectors, large amounts can be acquired [23]. For the cloning, transformation of the plasmid into the bacterium needs to take place and the selection of the transformed bacteria [24]. An anti-biotic resistance gene in the backbone vector

is a conventional tool to minimise the growth of untransformed cells [23]. To transform the bacteria, the temperature of the bacteria's environment is quickly changed, creating a pressure difference between inside and outside the cell [24]. This pressure difference will induce small pores through which exogenous plasmid can enter.

A fast and easy way to amplify and prepare a transgene construct for insertion into a backbone vector is via sticky end PCR [25]. This technique uses the conventional PCR technique, but with two different PCR primer couples (see figure 3). One primer of each duo has an extension at the 5' end matching the sequence of a restriction enzyme cleavage. After PCR with the primer couples in separate tubes, two products of the wanted construct will be made with the extensions for the primer couple. If these two products are mixed and denaturated and randomly anneal with each other, there is a 25% change that a product is created with the overhanging sequences that will be compatible with the restriction enzyme cleavages. Cutting the backbone vector with these restriction enzymes will create these cleavages and DNA ligase can ligate these compatible overhangs.

1.5 SHOX2 in sinoatrial node cells

As seen in Elliot et al. 2011 [26], and Schwach et al. 2017 [27], the expression of genes can be visualised by inserting a gene for a fluorescent protein into the gene of interest. The gene *NKX2.5* encodes a homeobox-containing transcription factor, which is prominent for the development of the human heart, and was the target for such an insertion [26]. The NKX2.5/GFP reporter line made it possible to quantify and select hESC-derived committed cardiac progenitor cells (hESC-CPCs) and cardiomyocytes (hESC-CMs) after differentiation based on their green fluorescence.

Later Schwach et al. 2017 targeted the *COUP-TFII* gene, a highly expressed gene in hESC derived atrial myocytes (hESC-AMs), for the insertion of mCherry in the NKX2.5 reporter line [27]. It enabled the distinguishing of hESC-AMs from hESC-ventricular myocytes (hESC-VMs) if the GFP positive cardiomyocytes were also positive for mCherry expression. Analysis of the action potentials of the beating cardiomyocytes showed that double positive cells had a quicker repolerisation among other features which are recognisable for atriomyocytes.

Now a new target gene was chosen to enrich this dual reporter cell line for proper identification of sinoatrial node (SAN) cells, *SHOX2*. The mRNA expression in mouse embryo hearts showed that the transcription factor, SHOX2, is only expressed in the SAN in high quantities [28]. Also SHOX2 overexpression stimulates differentiation to pacemaker cells. The transcription factor, which is mostly located in the nucleus, plays an important role in maintaining the pacemaker program and regulates other key SAN proteins like Hcn4, Isl1, and Tbx3 [29]. An antagonistic mechanism of SHOX2 against NKX2.5 is shown to be indispensable and causes SAN cells to be NKX2.5 negative.

Embryonic bodies with SHOX2 overexpression had electrophysiological hallmarks which mimicked those of actual SAN cells and had improved beating frequency and automaticity [28]. Ex vivo, SHOX2+ cells even had the ability to control the beating of neighbouring SHOX2- cells [29]. These pacemaker properties can be seen in SHOX2+ cells, regardless of the expression of other prominent transcription factors for the SAN program, like Tbx3 and Isl1, making it a worthy gene for reporting functional pacemaker cells.

1.6 mTAGBFP fusion with SHOX2

The reporter line will need a different colour fluorescent protein to distinguish SHOX2 expression from the green GFP fluorescent coupled with NKX2.5 expression and the red mCherry fluorescent linked with COUP-TFII expression. The blue coloured mTAGBFP fluorescent protein [30] is the protein chosen for this. For the insertion of the sequence coding for the mTAGBFP protein, the pSp-Cas9-sgRNA vector and the mTAG-HDR were designed and produced. They were designed to fuse the insert into the SHOX2 gene in such a way that the SHOX2 protein can still correctly fold into its functional vorm, while also having the fluorescent mTAGBFP protein covalently bound to it. Without these considerations the full functional expression of the SHOX2 protein will be diminished and the differentiation program could be in less efficient.

This rainbow reporter line will thus be able to distinguish SAN cells via their unique expression

of mTAGBFP and mCherry. Since SHOX2 has an antagonistic mechanism against NKX2.5 [29], the GFP expression will be less compared with for example CPCs.

Having a reporter line which can help determining SAN cells will have short-term benefits in research like: developing a standardised differentiation protocol for SAN cells, being able to select and purify SAN cells, determining the functional characteristics of in vitro developed SANs. These short-term benefits will help in the long term goal of replacing the battery powered pacemaker with a biological one in patients. So in this thesis the *SHOX2* gene of the dual reporter line from Schwach et al. 2017 [27] will be targeted for insertion of mTAGBFP.

1.7 MEF2C in cardiac development, reprogramming and diseases

The myocyte enhacer factor 2c (MEF2C) has been found to enhance transcription of proteins known for myocyte differentiation programs through the binding of DNA sites [31]. Furthermore, using a transcriptionally active MEF2C binding reporter in mouse embryos, binding activity was discovered in the heart [32]. A null mutation of MEF2C in the development of mouse embryo hearts showed multiple irregularities in morphogenesis and gene expression, deeming MEF2C essential for cardiac myogenisis [33]. The importance of MEF2C in cardiac development is clear.

However, during the regeneration of damaged dog hearts MEF2C was also detected [34], causing for new interest in MEF2C for cardiac regeneration. Direct reprogramming of fibroblasts to a cardiomyocyte-like state was possible through transducing three transcription factors (GATA4, Tbx5, and MEF2C) [35]. In infarcted hearts, the transfer of the genes for these transcription factors enhanced cardiac differentiation [36]. Wang et al. 2020 [37] even used deadCas9 transcription activation of the three transcription factors and HAND2 genes to reprogram fibroblasts into CPCs. These cases showed the potential of MEF2C research for the regeneration of scarred hearts. Further stressing the importance of research into mechanisms surrounding MEF2C is the fact that it is a protein often linked with cardiac diseases. For example, congenital heart disease has been linked to small mutations in the *MEF2C* gene in humans [39]. Also the RNA splicing of the *MEF2C* gene has a role in the development of cardiac hypertrophy and heart failure [40]. Cardiac hypertrophy is additionally induced by the downregulation of MEF2C, diminishing the expression of a micro-RNA, miR-133a, a hypertrophy reducing factor [41].

These multiple mechanisms which might affect either MEF2C functionality or expression could be investigated with a functional MEF2C detecting system.

1.8 Stable detection of MEF2C DNA binding

For these research ideas a stable detection of functional MEF2C is needed in living cells. The concept of using the MEF2C transcription enhancing abilities via DNA binding to transcribe a reporter protein, as can be seen in Naya et al. 1999 [32], can be used for this. The binding sequence of MEF2C in humans is approximately from 5 to 3 end 'CTATAAATAC' [42]. Multiple bound MEF2C proteins will form a complex with a p300 protein promoting transcription as can be seen in figure 4 [38].

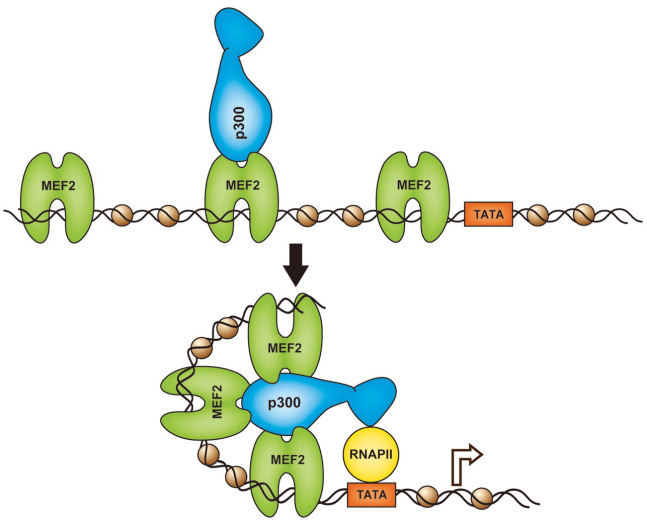


Figure 4: A hypothetical model of an enhanceosome consisting of a MEF2C-TATA box construct with three MEF2C DNA binding sites [38].

The distance of the TATA box to the transcription starting site (TSS) also plays a role in the transcription of the gene. The binding spot of RNA polymerase II with the sequence 5'-TATATAAT-3', also known as the TATA box, should generally have a small window of 30 to 39 bp after its sequence before the start codon 5'-ATG-3' starts [43]. Other TATA-TSS distances are often-

times non-functional or show less expression, since the polymerase structurally can't reach the start codon from its binding spot [44].

For the integration of a MEF2C detecting construct into the genomic DNA a location should be chosen where insertion of a gene won't affect the cell. The locus should also be able to allow the binding of transcription associated proteins for persistent expression. The human target sequence for adeno-associated virus integration (AAVS1) is a suitable location for these requirements [45]. The location even allows for expression after differentiation [46]. Furthermore, the protein that will be transcribed and translated should be detectable and have no further effect on the cell. The protein chosen for this is mRuby, due to its brightness and resistance to denaturation [47].

All taken together, an HDR vector targeting the AAVS1 locus for the insertion of a MEF2C construct promoting transcription of a mRuby sequence is needed for the stable detection of MEF2C in cells. With the ability to detect MEF2C possible mechanisms behind for example heart failure can be researched. In this thesis the production of the HDR vector will be a goal.

1.9 Research aims, goals and hypotheses

The primary objective of this thesis was to achieve subgoals within larger research projects using CRISPR/Cas technology. The discovery of CRISPR/Cas' genome engineering possibilities opened many doors in stem cell research. Based on literature for one of the first research project, it was theorised that a human embryonic stem cell (hESC) rainbow reporter line could be made with NKx2.5-GFP, COUP-TFII-mCherry, and SHOX2-mTAGBFP out of the dual reporter line from Schwach et al. [27]. This new cell line will help identify SAN pacemaker cells after differentiation. For the second research project, a HDR vector for the insertion of a MEF2C inducible mRuby gene into the AAVS1 locus should be made. This vector can later be used to create a stable cell line with the ability to indicate MEF2C expression.
To reach the primary objective these subgoals should be achieved.

Aim 1: Make the NKx2.5-GFP, COUP-TFII-mCherry, and SHOX2-mTAGBFP hESC rainbow reporter line.
Methods: Transfect the dual reporter line of Schwach et al. [27] with Cas9/sgRNA and SHOX2-mTAGblue HDR vectors. Select the puromycin resistant transfected cells. Sort the cells to create single cell colonies. Screen the clonal expansion for successful HDR with PCR of the insert region. Sequence for indels in the genes wild-type.
Expected results: A successful infusion of mTAGBFP on the SHOX2 gene, making the rainbow reporter line.

Aim 2: Develop the MEF2C-mRuby HDR vector.
Methods: Sticky end PCR the MEF2C-Tata construct. Linearize the mRuby-AAVS1 vector. Ligate the sticky end PCR into the mRuby-AAVS vector. Transform vector into E. coli bacteria for vector amplification. Isolate vector and screen with restriction enzyme digestion and sequencing.
Expected results: The MEF2C-mRuby plasmid for HDR insertion into AAVS1 is ready for transfection.

2 Materials and methods

2.1 MEF2C induced expression HDR template

Acquiring the MEF2C-TATA insert

The pGL2-Basic vector with the MEF2C-TATA construct was sequenced with Sanger sequencing services from GATC (Eurofins Genomics). See Supplemental Experimental Details for extra information, sequences and sequencing primers. Multiple sticky end PCR primers were designed and tested for the amplification of the construct. PCR primers are shown in Supplemental Experimental Details. PCR was done with the Phusion High-Fidelity PCR Kit (Thermo Fisher Scientific) according to the manufacturer's instructions in a T100 Thermal Cycler (Bio-Rad). The products were run on an agarose gel and extracted from gel (see Agarose Gel and Extraction). The two PCR products with extensions were mixed (1:1), denatured on 94°C for 4 min, annealed at 65°C for 1 min and annealed at RT for 14 min, to make the sticky end PCR product.

Digest and ligation of mRuby-AAVS1 HDR vector

A vector with the AAVS1 homology arms with an mRuby2-N1 sequence in between the arms was lend from another project at AST Utwente (See Supplemental Experimental Details for sequence). A double digest and two test digests with XmaJI (AvrII), HindIII restriction enzymes, and Fast Digest Green Buffer (all Thermo Fisher Scientific) were done overnight at 37°C. The linearised plasmid was directly put on an agarose gel without loading buffer and extracted (see Agarose Gel and Extraction). The linearised plasmid and the sticky end product mix were added to a ligation mix, Ligase 10X Buffer and T4 DNA Ligase (Promega), in a ratio of 1:3 in accordance with the Promega ligation protocol.

Cloning

The ligation product was transformed into One Shot TOP10 Competent Cells (Invitrogen) according to Invitrogen cloning protocol and 150 µl of the S.O.C. Medium (Invitrogen) with the transformed cells was spread over a LB agar (Miller, Fisher BioReagents) plate with 50 µg/ml Kanamycin (Sigma Aldrich). After 15 hours in an incubator at 37°C multiple colonies were picked for stab cultures in LB broth (Miller, Fisher BioReagents) with the same concentration of Kanamycin and these cultures were incubated with agitation at 37°C. After 8 hours these were added to 50 ml midipreps and were incubated further. The plasmids were isolated into Nuclease-Free Water with the PureYield Plasmid Midiprep System (Promega) and the concentration was measured with NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific). A glycerol back-up stock with the *E. coli* in a 40% Glycerol in ddMilliQ mixture was made and stored in Cryovails (Greiner Bio-One) at -80°C.

Digest screening and sequencing

The plasmids from the different conditions were linearised with MluI, XhoI, and EcorI restriction enzymes, and Fast Digest Green Buffer (all Thermo Fisher Scientific) overnight at 37°C. The products were run on agarose gels to determine product sizes and check for the inserted construct. The plasmids with the construct insert were sequenced with Sanger sequencing services from GATC (Eurofins Genomics).

2.2 Rainbow reporter line

Imaging COUP-line

Both hESC derived AMs and VMs, from the COUP-line of Schwach et al. 2017 [27], were provided by another project and kept in Cardiomyocyte Maintenance Medium (Gibco). Also undifferentiated hESCs from the same line were provided and kept in Essential 8 (E8) medium (Gibco). Fluorescent and brightfield images were taken with an inverted fluorescent microscope (Nikon). The exposure time and intensity were kept the same for all conditions.

Vectors and sequence

The Cas9 sgRNA and SHOX2 HDR vectors were stored in a glycerol stock from a previous student. Clones were picked and the vectors were isolated as previously described. The vectors were sequenced by Eurofins Genomics (for the vector sequences see Supplemental Experimental Details) and the integrity of the sgRNA and the HDR arms with the insert was determined.

HESC culture, transfection and selection

The hESCs were cultured on mouse embryonic fibroblasts (MEFs) in hESC medium to keep them from differentiating. At a confluency of 60-75% (0.35-1.0 x 10^6 hESCs) the co-culture was transfected with 4 µg of sgRNA1 or 2 vector and 1-3 µg of HDR vector utilising Lipofectamine Stem Transfection reagent (Thermo Fischer Scientific) for 24 hours. Selection of sgRNA vector transfected cells was done for 24 hours using hESC medium with 0.5 µg/ml Puromycin (Sigma Aldrich).

Mixed population screening

From the mixed transfected populations the DNA was extracted with QuickExtract DNA Extraction Solution (Lucigen) according to the manufacturer's instructions. Forty PCR cycles were done on the DNA with primer combinations for SHOX2 and mTAGBFP (See supplementary Table 2 for primer sequences) as described earlier. The products were run on gel electrophoresis for size analysis.

hESC passaging and sorting

The selected mixed population was dissociated with EDTA (Invitrogen) and transferred in E8 medium (Gibco) on Vitronectin (Thermo Fischer Scientific). During passaging also backup vials were made with cells stored in vapor phase liquid nitrogen below –135°C. For the sort the cells were taken up in Flow Cytometry Staining Buffer (Thermo Fischer Scientific) and they were sorted with the Sony SH800s Cell Sorter (for sort selection boundaries see Supplemental Experimental Details). Three cells were sorted into every well of a with Vitronectin coated 384 wells plate for clonal expansion. The medium used for the first 4 days of the clonal expansion was E8 (Gibco) with (1:10) CloneR (Stemcell Technologies), (1:100) RevitaCell (Gibco), and (1:500) Primocin (Fisher Scientific). After 4 days the medium was changed back to simple E8 (Gibco).

2.3 Analysis

Agarose Gel and Extraction

The PCR results were mixed with 6X DNA Gel Loading Dye (Thermo Fisher Scientific) and were analysed with agarose gel electrophoresis and 1 Kb plus DNA Ladder (Invitrogen) as comparison. The gels made from 1X TRIS-acetate-EDTA (TAE) Buffer with (1:10000) SYBR Safe DNA Gel Stain (Invitrogen) were run with Horizontal Electrophoresis Systems and Power Supplies from Bio-Rad. Extraction of the wanted PCR products was done with the Nucleospin Gel and PCR Clean-up Kit (Macherey-Nagel) and DNA concentration of the extract was measured with NanoDrop 1000 Spectrophotometer (Thermo Fisher Scientific).

3 Results

3.1 MEF2C

To insert a product with sticky ends into a vector, first the vector should acquire its sticky ends. In the HDR vector of AAVS1 which already contained the sequence for mRuby this could be done via the digestion of two restriction enzyme sites. The unique restriction enzyme sites HindIII and AvrII allowed for the linearisation of the vector, while leaving the homology arms intact and minimising the amount of basepairs before the start codon of the mRuby sequence. The results of this linearisation via these individual restriction digests can be seen in figure 5. Both restriction enzymes were able to linearise the vector. This is clear since the linearised vector moves faster through the gel than the open circular vector, but slower than the supercoiled vector.

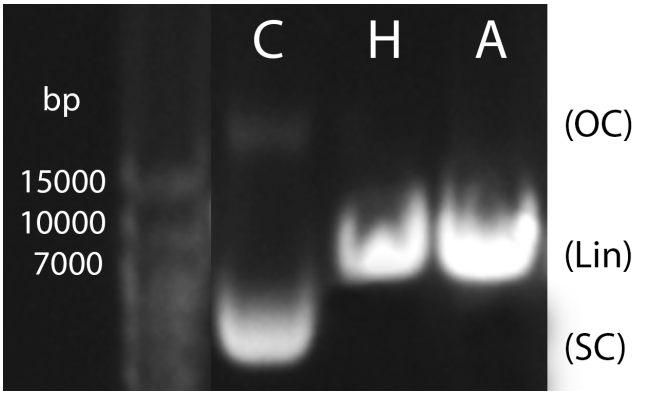


Figure 5: The AAVS1 HDR vector on gel. The letters represent the undigested vector (C), and the vector digested with either the restriction enzyme HindIII (H) or AvrII (A). The uncut vector can be seen in supercoiled (SC) and open circular (OC) state. The fully digested products are linear (Lin).

The linearised vector has overhanging basepairs left after the restriction site digest for HindIII, 5'-AGCT-3', and AvrII, 5'- CTAG-3'. Sticky end PCR primers were designed to amplify the MEF2c-TATA construct and to create extensions complementary to the overhanging basepairs. PCR primers for the sticky end PCR were designed and their expected products are displayed in table 1 (Supplementaries).

The sticky end PCR worked with all primer combinations as can be seen by the obtained product sizes in figure 6b compared to the simulated PCR results in figure 6a. Since the PCR product 2 has a good intensity and since their expected product after ligation with the linearised plasmid will leave 37 bp in between the TATA box and the start codon, this product has been chosen to continue with.

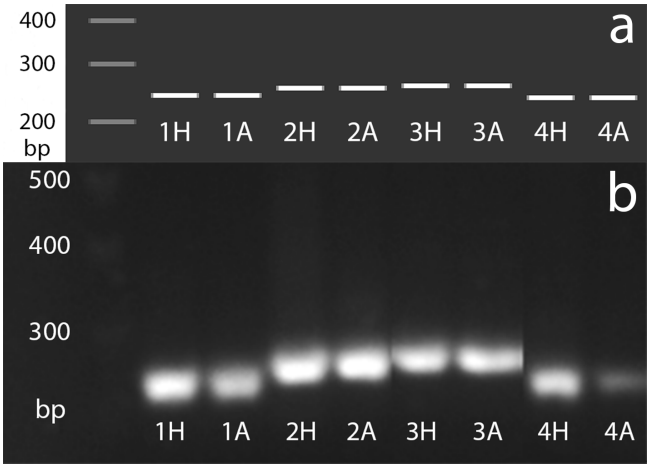


Figure 6: (a) Simulation of the sticky end PCR primer test. (b) The results of the test. The products are noted as either one of the duo with the extension for the HindIII sticky end or with the AvrII sticky end extension (e.g. 1H and 1A respectively). All tested primer combinations made correctly sized products

To increase the chances of transforming ligated vectors into the *E. coli*, first different ligation mixtures with increasing amounts of ligation products were put on gel. In figure 7 it can be seen that ligation mixtures will have the open circular ligated vector which migrates slower than the linearised vector. Increasing the amount of the ligation products did increase the intensity of the open circular band, meaning that the amount of open circular DNA increased. This enforced the reasoning that using the ligation mixture containing more ligated vectors for the transformation might increase the chances of transforming this product. Since the unique overhangs of HindIII and AvrII were not complementary to each other they were not able to be ligated together.

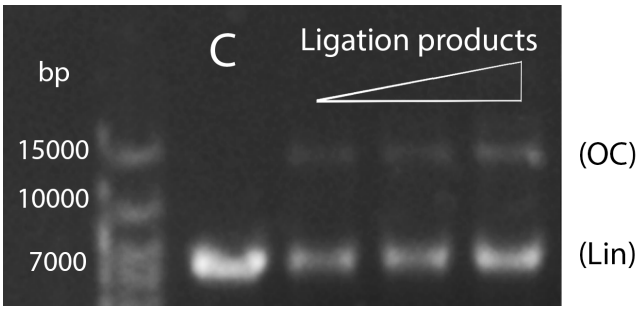


Figure 7: Ligation mixtures with increasing amounts of both HDR vector and MEF2c insert (respectively from left to right). Ligation mixture C did not contain the insert. The ligation mixtures with insert formed slower migrating open circular (OC) vectors from the linearised (Lin) HDR vector and insert.

After the successful transformation and DNA isolation from four clones, the restriction digest screen of these vectors was performed. Comparing the results in figure 9 with the expected results in figure 8, it can clearly be seen that the clones 2 and 4 contained vectors which were cut with the restriction enzymes as expected. Meaning that the vectors isolated from those clones are expected to have the MEF2C construct inserted next to the mRuby sequence. Sequencing these vectors will confirm these expectations and will give further insight in the integrity of the construct.

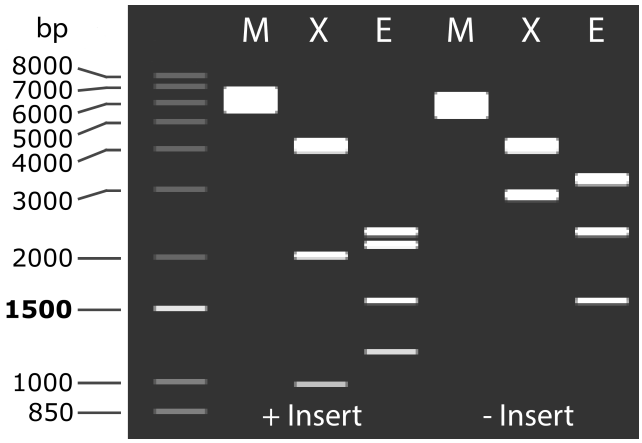


Figure 8: Simulation of restriction digest screen. Restriction enzymes MluI, XhoI, and EcorI digesting the HDR vector with the MEF2C construct in 1, 2, and 3 respectively, and digesting the HDR vector without the construct in 4, 5, and 6 respectively.

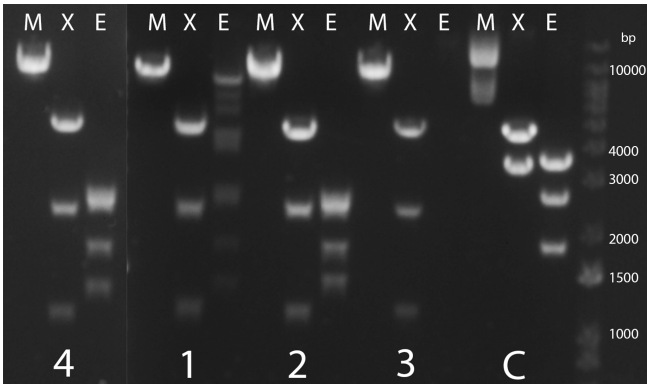


Figure 9: Digest screen of the isolated DNA from the 4 clones. From left to right, the enzymes MluI, XhoI, and EcorI cutting the vector from clones 4, 1, 2, and 3, and the original HDR vector respectively.

The sequencing data of the AAVS1 HDR vector for these clones (See figure 14 and figure 15 for schematics, and see Supplemental Experimental Details for detailed sequences) showed that both clones have a single nucleotide polymorphism (SNP) in the left homology arm. Also in the MEF2c construct an SNP was visible in between a MEF2C binding site and the TATA box (see figure 16). Clone 2 even had an insert which sequence was the same as the three MEF2c binding sites, meaning that clone 2 has four binding sites. The rest of the functional parts (e.g. the mRuby sequence, the right homology arm) were sequenced as expected.

3.2 Rainbow reporter

The hESC COUP-TFII reporter line was imaged for GFP and mCherry expression in three states: undifferentiated, differentiated towards AM, and differentiated towards VM (see figure 10 A, B and C respectively). The undifferentiated cells showed no expression for either fluorescent protein compared to the hESC-AMs. Possible dead cells or debris did show some sign of fluorescence, but not as intense as the hESC-AMs. The hESC-VMs did show GFP expression, but did not show as heavy amounts of mCherry expression as the hESC-AMs. Since the GFP is expressed when NKX2.5 is being expressed

and the same for mCherry with COUP-TFII expression, it can be concluded that the reporter line still functions as designed.

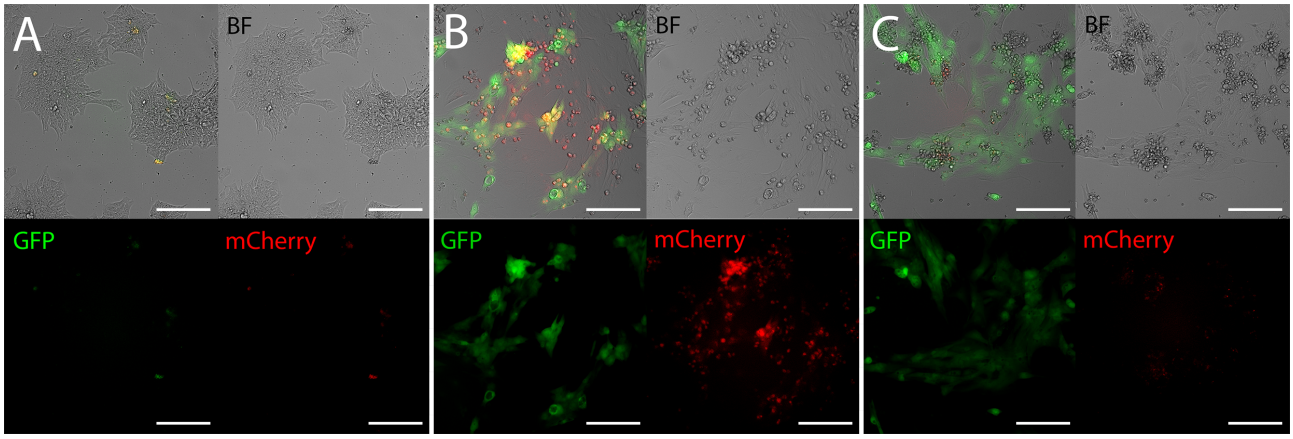


Figure 10: GFP and mCherry expression together with BF images in respectively (A) undifferentiated hESC, (B) hESC-AM, (C) hESC-VM. Scale bar represents 200 µm.

After the DNA isolation of the SHOX2 HDR vector and the two Cas9 sgRNA vectors, the vectors were sequenced. The sequencing results showed that both sgRNA scaffolds were still as designed. For the SHOX2 HDR vector PAM sites did not correlate with the needed PAM sequence for Cas9 binding, meaning that the HDR vector can not be cut by the Cas9 sgRNA complex. The integrity of the mTAGBFP sequence and the left homology arm were intact. However the right homology arm contained multiple mismatches with the design. Thirteen nucleotides of the total 813 did not correlate in the homology arm, meaning a correlation of 98.4%.

The selection after the transfection allowed only cells to survive which were transfected with the Cas9 sgRNA vector and in which the Cas9 came to expression. Enough cells survived the selection to allow for the continuation of the experiment, meaning that the transfection was successful. The new population was mixed with cells which had been targeted for the insertion of mTAGBFP via HDR and cells which were only cut and repaired without the insertion. The screening of this population with a high amount of PCR cycles (forty) would amplify and indicate even small concentrations of DNA with this insert.

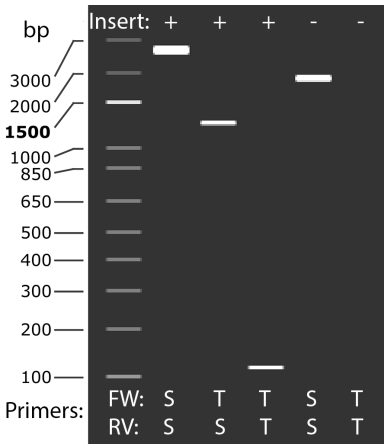


Figure 11: Simulation of SHOX2 PCR products on gel. The products are predicted with (+) or without (-) the mTAGBFP insert in the SHOX2 gene using PCR primers for both the sequence of SHOX2 and mTAGBFP (primers denoted as e.g. SHOX2 forward (FW: S) or mTAGBFP reverse (RV: T)).

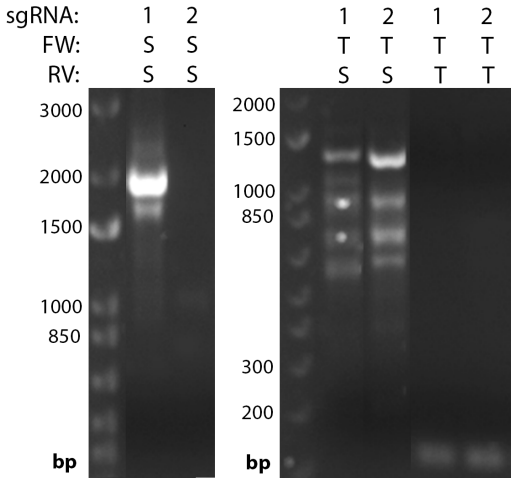
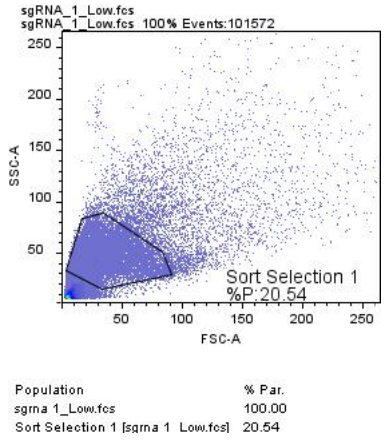


Figure 12: The SHOX2 PCR products of both the sgRNA 1 and 2 targeted conditions after 40 PCR cycles using PCR primers for both the sequence of SHOX2 and mTAGBFP (primers denoted as e.g. SHOX2 forward (FW: S) or mTAGBFP reverse (RV: T)). Products sized as expected for the mTAGBFP insertion were visible for both sgRNAs.

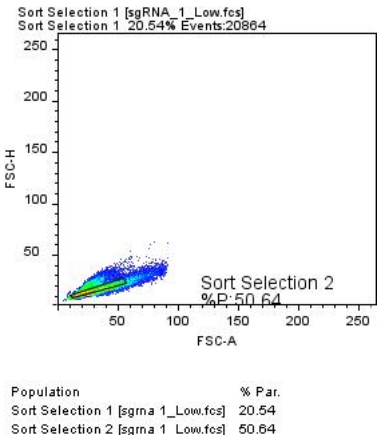
In figure 11 the expected results for products with or without the insert can be seen and in figure 12 the acquired PCR results are shown. With the forward and reverse SHOX2 PCR primers the product shows an intense band just under two thousand bp and no strong band can clearly be seen around the 2800 bp region. This would indicate that either the expected product with the insert wasn't produced or that amplified product had a concentration which was so much lower relative to the product without the insert that the product band is overshadowed.

However correctly sized products are visible for the PCR with the forward and reverse mTAGBFP primers. This means that the mTAGBFP sequence is in the genome of some of the population for both sgRNA conditions. When looking at the products formed with the mTAFBFP forward and SHOX2 reverse PCR primers a band with a relatively higher intensity than the other bands formed in the lane indicated a prominent product was approximately 1300 bp. This product was expected to be formed if the mTAGBFP sequence was inserted into the *SHOX2* gene. These results together show a strong conviction of the insertion of the mTAGBFP sequence in the *SHOX2* gene.

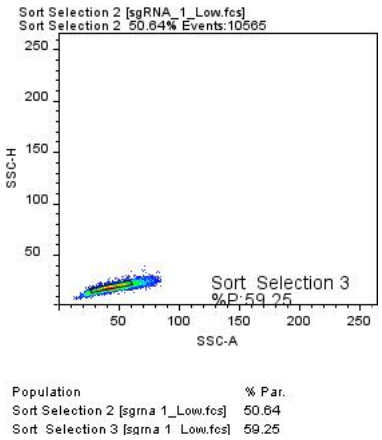
The mixed populations were dissociated and sorted as described in the Methods. The data from the sort and the boundaries for selecting single cells are visible in figure 13, 17, and in Supplemental Experimental Details. The boundaries for selection 1 were based on complexity (side scatter area) and size (forward scatter area) of the cells. Since the data is relative only to the events that were measured, a dense area of events will form representing debris, which are small and simple, and a dense area will form representing the cells from the population. These average cells were selected in selection 1. Selection 2 and 3 sorted for single cells on the basis that doublets show a different height compared to their averaged forward or side scatter.



(a) Selection 1. Forward scatter (FSC) area vs. Side scatter (SSC) area.



(b) Selection 2. FSC area vs. FSC height.



(c) Selection 3. SSC area vs. SSC height.

Figure 13: The selection for single cells from one condition targeted with sgRNA 1.

Ten days after the sort zero out of the approximately seven hundred wells contained a colony, so the passaging and screening of the clonal expansion could not be done.

4 Discussion

The linearisation of the AAVS1 HDR is pointed out by the difference in the migrated distance of the linearised vector on gel compared to the circular and supercoiled vector. Circular vectors tend to wind up and supercoil since it is a more energetically favourable structure [48]. And because it is wound up more compact it will experience less resistance moving through the gel allowing for a larger migrated distance. The linearised vector is not able to coil up as compact and will thus experience more resistance. Relaxed circular vector or nicked vectors either haven't undergone supercoiling yet or don't have the same favourable energetic landscape to change their structure [48]. So they are in a conformation which is even less favourable for migration.

The SNP in the MEF2C-TATA box construct is located in between a binding site of MEF2C (5'-CTATAAATAC-3') and the TATA box (5'-TATATAAT-3') as can be seen in figure 16. It is not uncommon for the nucleotides next to the MEF2C binding site to differ from the regulatory regions of different genes [42]. Furthermore, this specific region in the construct is not included in the three times repeating sequences which were specifically designed as the MEF2C binding sites. So it seems unlikely that the functional binding will be hindered. Also an enhanceosome as in figure 4 does not seem to give any function for this region. However, only via testing of the vector for its ability to express mRuby through MEF2C promotion can the functional integrity be confirmed for sure.

The final isolated vector for AAVS1 HDR currently requires the HDR to insert about 3.3 kb (see MEF_mRuby in Supplemental Experimental Details). For this particular AAVS1 HDR vector there was already an isolated and sequence confirmed vector containing the mRuby sequence. This vector even contained unique restriction enzyme sites, of which the restriction enzymes were available, that would allow for the sticky end PCR ligation with the needed requirements like the distance from the TATA box to the TSS. However this vector did have a large unnecessary sequence right after the mRuby sequence. As discussed earlier these larger inserts have lower HDR efficiencies [8]. To reduce the size of the insert region a restriction enzyme can be used that has two restriction sites in this unnecessary region, MfeI. Through a digest with this enzyme, followed by a ligation the insert sequence can be reduced with 1.7 kb to 1.6 kb, and thus making the efficiency of HDR higher.

For future production of HDR vectors a different technique could be used which would allow for even leaner vectors, in-fusion cloning [49]. This technique uses the same principle of amplifying wanted products via PCR with extended sequences on the ends of the products. The extended ends should be designed to be compatible in whatever way the different products should be ordered. These extended ends will be opened up to the compatible overhangs by an in-fusion enzyme. And finally these products with compatible overhangs will be transformed into bacteria which will ligate the vector together. This technique allows for perfect control of what the vector should consist of [49].

In the imaging of the COUP-TFII reporter line, small amounts of mCherry fluorescence where still visible during the imaging of the hESC-VMs (see figure 10). When also looking at the brightfield images where this fluorescence might come from it seemed to not only come from debris, but also some possible living cells. This could be because the differentiation process of the cells is not a hundred percent efficient [27]. Some smaller amount of cells could differentiate towards other cell types, causing the expression of proteins which would not be expected in a sample perfectly selected on cell type.

The amount of homology in the homology arms with the targeted gene improves the efficiency of the HDR [5], [9]. So the SNPs in the homology arms most likely did have an effect on the amount of effective insertions in the *SHOX2* gene, but as shown in figure 12 it did not become impossible.

The reasons behind the failing of the clonal expansion are not known for sure, but are very likely to be technical. The state in which the cells are right before the sort, the harshness of the conditions during the sort and after could play a significant role in the potency of the cells surviving the sort. Supplements like CloneR and RevitaCell are supposed to improve these conditions. However, also

technical deviations in the cyto flowmetry or wrongly chosen boundaries for the selection could result in sorting less potent cells.

Since the expression and functional binding of MEF2C is a prominent factor in the pathology of heart diseases such as cardiac hypertrophy and congenital heart disease [39], [40], research towards underlying mechanisms or downstream mechanisms would be essential to understand these diseases. The AAVS1 HDR vector allowing the stable insertion of a MEF2C DNA binding construct, which promotes the transcription of a mRuby sequence, will help with the detection of functional MEF2C proteins.

The fact that SHOX2 expression is found in cells with pacemaker properties, like controlling the beating of neighbouring SHOX2- cells and having the electrophysiological hallmarks of actual SAN cells, regardless of the expression of other SAN prominent transcription factors, show that it is a protein which can report functional pacemaker cells. The reporter line would help in the processes needed for better research in the understanding of pacemaker cells. The standardisation of protocols for pacemaker differentiation and the possible selection process of SAN cells will make great leaps in development in the characterisation of hESC derived SAN cells and maybe even leaps in the development of a biological pacemaker worthy of replacing the artificial one.

In summary, HDR vector capable of inserting a MEF2C DNA binding construct with the ability to promote transcription of a mRuby sequence into the *AAVS1* gene for stable detection of functional MEF2C was produced. Via sticky end PCR was it able to insert this construct into an already existing vector. The stable detection of functional MEF2C could in further research be used to uncover underlying mechanisms of MEF2C related diseases like cardiac hypertrophy and congenital heart disease. Also a population able to report SHOX2 expression via the genetical fusion of a sequence coding for mTAGBFP was developed through targeting with CRISPR/Cas9 HDR on a NKX2.5/COUP-TFII dual reporter line. This population will be used to create the reporter line able to detect functional pacemaker cells. With such a rainbow reporter the development of a biological pacemaker could potentially be accomplished in the future.

References

[1] F. Jiang and J. A. Doudna, "CRISPR-Cas9 Structures and Mechanisms," vol. 46, pp. 505–529, May 2017. DOI: 10.1146/annurev-biophys-062215-010822.

[2] J. D. Sander and J. K. Joung, "CRISPR-Cas systems for editing, regulating and targeting genomes," *Nature Biotechnology*, vol. 32, no. 4, pp. 347–350, Mar. 2014. DOI: 10.1038/nbt.2842.

[3] A. Pickar-Oliver and C. A. Gersbach, "The next generation of CRISPR–Cas technologies and applications," *Nature Reviews Molecular Cell Biology*, vol. 20, no. 8, pp. 490–507, Aug. 2019. DOI: 10.1038/s41580-019-0131-5.

[4] P. D. Hsu and E. S. Lander, "Development and applications of CRISPR-Cas9 for genome engineering," *Cell*, vol. 157, no. 6, pp. 1262–1278, Jun. 2014. DOI: 10.1016/j.cell.2014.05.010.

[5] F. A. Ran and P. D. Hsu, "Genome engineering using the CRISPR-Cas9 system," *Nature Protocols*, vol. 8, no. 11, pp. 2281–2308, Oct. 2013. DOI: 10.1038/nprot.2013.143.

[6] C. D. Richardson and G. J. Ray, "Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA," *Nature Biotechnology*, vol. 34, no. 3, pp. 339–344, Mar. 2016. DOI: 10.1038/nbt.3481.

[7] B. Wang and K. Li, "Highly efficient CRISPR/HDR-mediated knock-in for mouse embryonic stem cells and zygotes," *BioTechniques*, vol. 59, no. 4, pp. 201–208, Oct. 2015. DOI: 10.2144/000114339.

[8] K. Li and G. Wang, "Optimization of genome engineering approaches with the CRISPR/Cas9 system," *PLoS ONE*, vol. 9, no. 8, e105779, Aug. 2014. DOI: 10.1371/journal.pone.0105779.

[9] J. P. Zhang and X. L. Li, "Efficient precise knockin with a double cut HDR donor after CRISPR/Cas9-mediated double-stranded DNA cleavage," *Genome Biology*, vol. 18, no. 1, pp. 1–18, Feb. 2017. DOI: 10.1186/s13059-017-1164-8.

[10] M. Bibikova and K. Beumer, "Enhancing gene targeting with designed zinc finger nucleases," *Science*, vol. 300, no. 5620, p. 764, May 2003. DOI: 10.1126/science.1079512.

[11] M. Christian and T. Cermak, "Targeting DNA Double-Strand Breaks with TAL Effector Nucleases," *Genetics*, vol. 186, no. 2, pp. 757–761, Oct. 2010. DOI: 10.1534/genetics.110.120717.

[12] J. A. Doudna and E. Charpentier, "The new frontier of genome engineering with CRISPR-Cas9," *Science*, vol. 346, no. 6213, p. 1 258 096, Nov. 2014. DOI: 10.1126/science.1258096.

[13] F. V. Karginov and G. J. Hannon, "The CRISPR System: Small RNA-Guided Defense in Bacteria and Archaea," *Molecular Cell*, vol. 37, no. 1, pp. 7–19, Jan. 2010. DOI: 10.1016/j.molcel.2009.12.033.

[14] V. Globyte and S. H. Lee, "CRISPR/Cas9 searches for a protospacer adjacent motif by lateral diffusion," *The EMBO Journal*, vol. 38, no. 4, e99466, Feb. 2019. DOI: 10.15252/embj.201899466.

[15] V. Mekler, L. Minakhin, and K. Severinov, "Mechanism of duplex DNA destabilization by RNA-guided Cas9 nuclease during target interrogation," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 114, no. 21, pp. 5443–5448, May 2017. DOI: 10.1073/pnas.1619926114.

[16] I. E. Ivanov and A. V. Wright, "Cas9 interrogates DNA in discrete steps modulated by mismatches and supercoiling," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 117, no. 11, pp. 5853–5860, Mar. 2020. DOI: 10.1073/pnas.1913445117.

[17] S. H. Sternberg and B. Lafrance, "Conformational control of DNA target cleavage by CRISPR-Cas9," *Nature*, vol. 527, no. 7576, pp. 110–113, Nov. 2015. DOI: 10.1038/nature15544.

[18] Y. S. Dagdas and J. S. Chen, "A conformational checkpoint between DNA binding and cleavage by CRISPR-Cas9," *Science Advances*, vol. 3, no. 8, Aug. 2017. DOI: 10.1126/sciadv.aao0027.

[19] Z. Liu and O. Chen, "Systematic comparison of 2A peptides for cloning multi-genes in a polycistronic vector," *Scientific Reports*, vol. 7, no. 1, Dec. 2017, ISSN: 20452322. DOI: 10.1038/s41598-017-02460-2. [Online]. Available: /pmc/articles/PMC5438344/%20/pmc/articles/PMC5438344/?report=abstract%20https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5438344/.

[20] M. Kimura and K. Thakar, "Novel approaches for the identification of nuclear transport receptor substrates," in *Methods in Cell Biology*, vol. 122, Academic Press Inc., Jan. 2014, ch. 16, pp. 353–378. DOI: 10.1016/B978-0-12-417160-2.00016-3.

[21] E. J. Aird and K. N. Lovendahl, "Increasing Cas9-mediated homology-directed repair efficiency through covalent tethering of DNA repair template," *Communications Biology*, vol. 1, no. 1, pp. 1–6, Dec. 2018. DOI: 10.1038/s42003-018-0054-2.

[22] H. Bai and G. M. Lester, *Cytoplasmic transport and nuclear import of plasmid DNA*, Dec. 2017. DOI: 10.1042/BSR20160616.

[23] T. A. Brown, *Gene cloning and DNA analysis: an introduction*, 6th. Wiley-Blackwell, Apr. 2010, pp. 88–104, ISBN: 9781444318616.

[24] D. Hanahan, "Studies on transformation of Escherichia coli with plasmids," *Journal of Molecular Biology*, no. 4, pp. 557–580, Jun. 1983. DOI: 10.1016/S0022-2836(83)80284-8.

[25] G. Zeng, "Sticky-end PCR: New method for subcloning," *BioTechniques*, vol. 25, no. 2, pp. 206–208, 1998. DOI: 10.2144/98252bm05.

[26] D. A. Elliott and S. R. Braam, "NKX2-5 eGFP/w hESCs for isolation of human cardiac progenitors and cardiomyocytes," *Nature Methods*, vol. 8, no. 12, pp. 1037–1043, Dec. 2011. DOI: 10.1038/nmeth.1740.

[27] V. Schwach and A. O. Verkerk, "A COUP-TFII Human Embryonic Stem Cell Reporter Line to Identify and Select Atrial Cardiomyocytes," *Stem Cell Reports*, vol. 9, no. 6, pp. 1765–1779, Dec. 2017. DOI: 10.1016/j.stemcr.2017.10.024.

[28] V. Ionta and W. Liang, "SHOX2 overexpression favors differentiation of embryonic stem cells into cardiac pacemaker cells, improving biological pacing ability," *Stem Cell Reports*, vol. 4, no. 1, pp. 129–142, Jan. 2015. DOI: 10.1016/j.stemcr.2014.11.004.

[29] W. Ye and J. Wang, "A common Shox2-Nkx2-5 antagonistic mechanism primes the pacemaker cell fate in the pulmonary vein myocardium and sinoatrial node," *Development*, vol. 142, pp. 2521–2532, 14 Jul. 2015. DOI: 10.1242/dev.120220.

[30] O. M. Subach and I. S. Gundorov, "Conversion of Red Fluorescent Protein into a Bright Blue Probe," *Chemistry and Biology*, vol. 15, no. 10, pp. 1116–1124, Oct. 2008. DOI: 10.1016/j.chembiol.2008.08.006.

[31] M. J. Potthoff and E. N. Olson, "MEF2: A central regulator of diverse developmental programs," *Development*, vol. 134, no. 23, pp. 4131–4140, Dec. 2007. DOI: 10.1242/dev.008367. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/17959722/.

[32] F. J. Naya and C. Wu, "Transcriptional activity of MEF2 during mouse embryogenesis monitored with a MEF2-dependent transgene," *Development*, vol. 126, no. 10, pp. 2045–2052, May 1999. DOI: 10.1242/dev.126.10.2045.

[33] Q. Lin, J. Schwarz, C. Bucana, and E. N. Olson, "Control of mouse cardiac morphogenesis and myogenesis by transcription factor MEF2C," *Science*, vol. 276, no. 5317, pp. 1404–1407, May 1997. DOI: 10.1126/science.276.5317.1404. [Online]. Available: http://science.sciencemag.org/.

[34] A. Linke and P. Muller, "Stem cells in the dog heart are self-renewing, clonogenic, and multipotent and regenerate infarcted myocardium, improving cardiac function," *Proceedings of the National Academy of Sciences of the United States of America*, vol. 102, no. 25, pp. 8966–8971, Jun. 2005. DOI: 10.1073/pnas.0502678102.

[35] "Direct reprogramming of human fibroblasts toward a cardiomyocyte-like state," *Stem Cell Reports*, vol. 1, no. 3, pp. 235–247, Sep. 2013. DOI: 10.1016/j.stemcr.2013.07.005.

[36] K. Inagawa and K. Miyamoto, "Induction of cardiomyocyte-like cells in infarct hearts by gene transfer of Gata4, Mef2c, and Tbx5," *Circulation Research*, vol. 111, no. 9, pp. 1147–1156, Oct. 2012. DOI: 10.1161/CIRCRESAHA.112.271148. [Online]. Available: http://circres.ahajournals.org/lookup/suppl/doi:10.1161/CIRCRESAHA.112..

[37] J. Wang and X. Jiang, "Lineage reprogramming of fibroblasts into induced cardiac progenitor cells by CRISPR/Cas9-based transcriptional activators," *Acta Pharmaceutica Sinica B*, vol. 10, no. 2, pp. 313–326, Feb. 2020. DOI: 10.1016/j.apsb.2019.09.003.

[38] J. He and J. Ye, "Structure of p300 bound to MEF2 on DNA reveals a mechanism of enhanceosome assembly," *Nucleic Acids Research*, vol. 39, no. 10, pp. 4464–4474, May 2011. DOI: 10.1093/nar/gkr030.

[39] X. H. Qiao and F. Wang, "MEF2c loss-of-function mutation contributes to congenital heart defects," *International Journal of Medical Sciences*, vol. 14, no. 11, pp. 1143–1153, Sep. 2017. DOI: 10.7150/ijms.21353. [Online]. Available: https://pubmed.ncbi.nlm.nih.gov/29104469/.

[40] C. Gao and S. Ren, "RBFox1-mediated RNA splicing regulates cardiac hypertrophy and heart failure," *Journal of Clinical Investigation*, vol. 126, no. 1, pp. 195–206, Jan. 2016. DOI: 10.1172/JCI84015.

[41] X. Chen and B. Gao, "MEF2 signaling and human diseases," *Oncotarget*, vol. 8, no. 67, pp. 112 152–112 165, Dec. 2017. DOI: 10.18632/oncotarget.22899.

[42] V. Andres, M. Cervera, and V. Mahdavi, "Determination of the consensus binding site for MEF2 expressed in muscle and brain reveals tissue-specific sequence constraints," *Journal of Biological Chemistry*, vol. 270, no. 40, pp. 23 246–23 249, 1995. DOI: 10.1074/jbc.270.40.23246.

[43] J. Ponjavic and B. Lenhard, "Transcriptional and structural impact of TATA-initiation site spacing in mammalian core promoters," *Genome Biology*, vol. 7, no. 8, R78, Aug. 2006. DOI: 10.1186/gb-2006-7-8-r78.

[44] X. Liu and D. A. Bushnell, "RNA polymerase II transcription: Structure and mechanism," *Biochimica et Biophysica Acta*, vol. 1829, no. 1, pp. 2–8, Jan. 2013. DOI: 10.1016/j.bbagrm.2012.09.003.

[45] J. R. Smith and S. Maguire, "Robust, Persistent Transgene Expression in Human Embryonic Stem Cells Is Achieved with AAVS1-Targeted Integration," *Stem Cells*, vol. 26, no. 2, pp. 496–504, Feb. 2008. DOI: 10.1634/stemcells.2007-0039.

[46] F. Oceguera-Yanez and S. I. Kim, "Engineering the AAVS1 locus for consistent and scalable transgene expression in human iPSCs and their differentiated derivatives," *Methods*, vol. 101, pp. 43–55, May 2016. DOI: 10.1016/j.ymeth.2015.12.012.

[47] S. Kredel and F. Oswald, "mRuby, a bright monomeric red fluorescent protein for labeling of subcellular structures," *PLoS ONE*, vol. 4, no. 2, Feb. 2009. DOI: 10.1371/journal.pone.0004391.

[48] V. Patrick Higgins and A. V. Vologodskii, "Topological Behavior of Plasmid DNA," *Plasmids: Biology and Impact in Biotechnology and Discovery*, vol. 3, no. 2, pp. 105–131, 2015. DOI: 10.1128/9781555818982.ch7.

[49] M. Raman and K. Martin, "One solution for cloning and mutagenesis: In-Fusion® HD Cloning Plus," *Nature Methods*, vol. 11, no. 9, pp. iii–v, Sep. 2014. DOI: 10.1038/nmeth.f.373.

5 Supplementals

Table 1: The different possible PCR products which can be made with the designed sticky end PCR primers. Desired couple products are numbered 1 to 4 and the distance of their TATA box to the start codon of mRuby after ligation is shown.

PCR primers	MEF_STICKY_1	MEF_STICKY_2	MEF_STICKY_5	MEF_STICKY_6
MEF_STICKY_3	2 HindIII extension TATA to TSS: 37 bp	Double extension	3 HindIII extension TATA to TSS: 41 bp	Double extension
MEF_STICKY_4	No extension	2 AvrII extension TATA to TSS: 37 bp	No extension	3 AvrII extension TATA to TSS: 41 bp
MEF_STICKY_7	No extension	4 AvrII extension TATA to TSS: 37 bp	No extension	1 AvrII extension TATA to TSS: 41 bp
MEF_STICKY_8	4 HindIII extension TATA to TSS: 37 bp	Double extension	1 HindIII extension TATA to TSS: 41 bp	Double extension

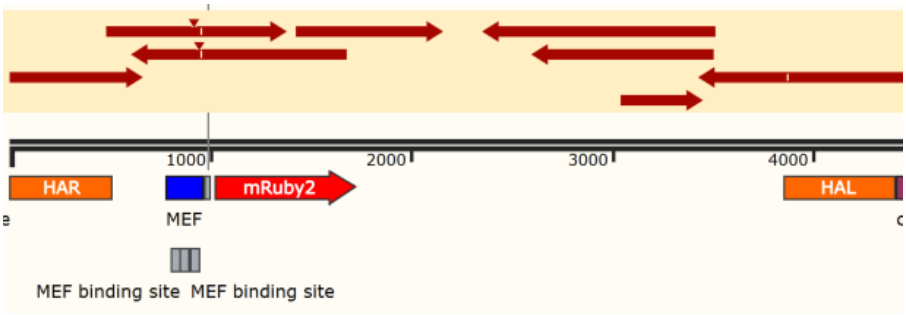


Figure 14: Schematics of the sequencing of clone 2. With one SNP in the HAL and in the MEF2c construct. Also an insert is in the MEF2c construct.

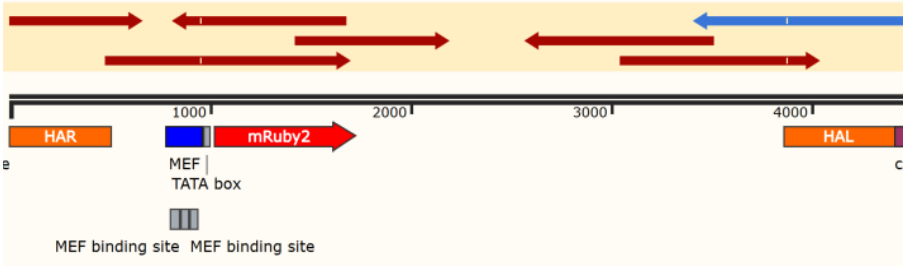


Figure 15: Schematics of the sequencing of clone 4. With one SNP in the HAL and in the MEF2c construct.

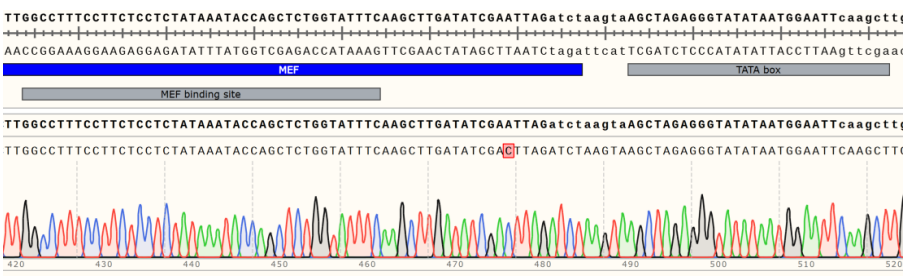
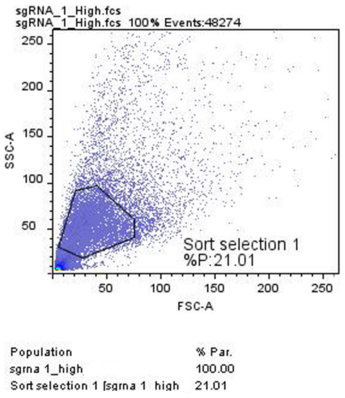
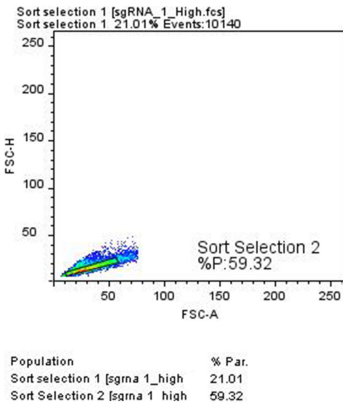


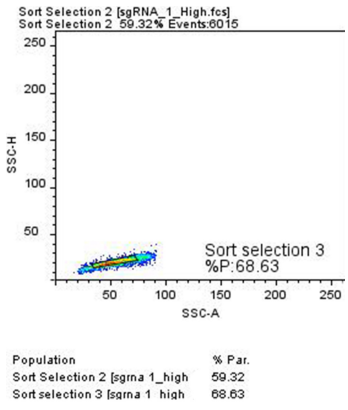
Figure 16: Sequencing result of clone 4 showed clear signs of an SNP in between the MEF2c binding site and the TATA box.



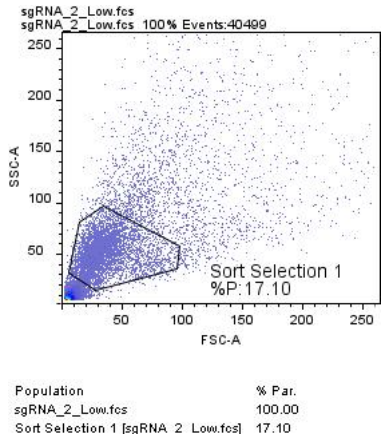
(a) Selection 1. Forward scatter (FSC) area vs. Side scatter (SSC) area.



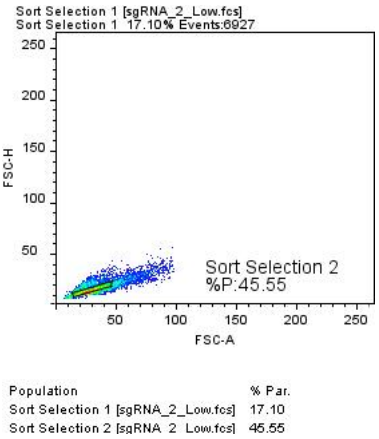
(b) Selection 2. FSC area vs. FSC height.



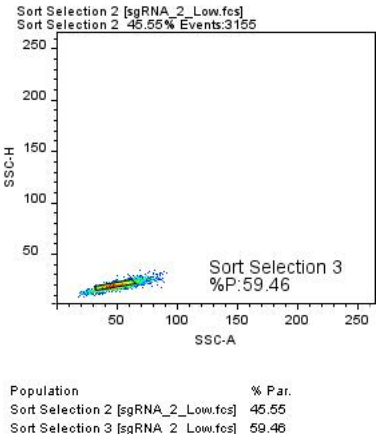
(c) Selection 3. SSC area vs. SSC height.



(d) Selection 1. Forward scatter (FSC) area vs. Side scatter (SSC) area.



(e) Selection 2. FSC area vs. FSC height.



(f) Selection 3. SSC area vs. SSC height.

Figure 17: The selection for single cells from one condition targeted with sgRNA 1 and one condition targeted with sgRNA 2 (a to c, and d to f respectively).