UNIL | Université de Lausanne
Faculté de biologie et de médecine

Centre hospitalier universitaire vaudois

SCHOOL OF MEDICINE – MASTER THESIS 5778

GENOMIC TAGGING OF ENDOGENOUS SYT-SSX FUSION PROTEIN IN SYNOVIAL SARCOMA CELL LINE HS-SYII

Student
Johan Hungerbühler

Tutor
Prof. *Ivan Stamenkovic*
Institut Universitaire de Pathologie – CHUV

Expert
Prof. *Tatiana Petrova*
Institut Universitaire de Pathologie – CHUV

Lausanne, March 2019

TABLE OF CONTENTS

List of Figures

List of tables

List of abbreviations

ChIP-Seq	Chromatin immunoprecipitation and sequencing analysis
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
crRNA	CRISPR-associated RNA
DSB	Double strand break
EmGFP	Emerald Green Fluorescent Protein
FBS	Fetal bovine serum
HDR	homology-directed repair
KRAB	Kruppel Associated Box
LHA	Left homology arm
MOI	Multiplicity of infection
Neo	Neomycin
NeAA	Non-essential amino-acids
NHEJ	Non-homologous end joining
PS	Penicillin-streptomycin
PCR	Polymerase chain reaction
PAM	Protospacer adjacent motif
RT-qPCR	Real-time quantitative polymerase chain reaction
rAAV	Recombinant adeno-associated virus
RHA	Right homology arm
sgRNA	Single-guide RNA
SSX-RD	SSX repression domain
SS	Synovial sarcoma
tracrRNA	Trans-activating RNA
TALE	Transcription activator-like effector
TALEN	Transcription activator-like effector nuclease

Abstract

Synovial Sarcoma is a soft tissue malignancy harboring a pathognomonic chromosomal translocation t(X;18)(p11.2;q11.2) producing a chimeric translocation protein called SYT-SSX. It is an aggressive tumor type mainly occurring in children and young adults, with a poor prognosis and no specific therapy. A better understanding of the molecular pathogenesis is needed to develop new therapeutic tools.

In recent publications, it has been proven that this translocation protein deeply affects the epigenetic pathways of thecell. To better understand these modifications, a ChIP-Seq analysis can be performed, creating the need for a strong antibody target within the sequence of this chimeric protein.

In this work, we induced a double strand break using a CRISPR-Cas9 system and designed a Donor DNA sequence with the aim to use the homology-directed repair mechanism to insert a V5 genomic tag in the HS-SYII cell genome. We then started preliminary work to enable precise co-infection with both the specific CRISPR-Cas9 system and our Donor DNA to induce the V5 tag expression in SYT-SSX.

Keywords: Synovial Sarcoma, genome editing, SYT-SSX, cell culture, HS-SYII

1 Introduction

1.1 Synovial sarcoma

Sarcomas are a rare group of malignancies that derive from mesenchymal tissue and comprise a wide range of tumor types. The primary tumors can appear in a variety of different locations including bone and soft tissues such as fat and muscle. They can be subdivided into two groups based on the genetic events that are believed to drive their pathogenesis: tumors associated with unique, “signature” chromosomal translocations and multiple events, including point mutations, and chromosomal translocations, inversions and deletions without any specific pattern (1).

Synovial sarcoma (SS), which represents 5 to 10% of all soft tissue sarcomas (2), is characterized by a specific translocation. It can occur at any age but is most commonly seen within the 15-35-year-old range. Contrary to what its name suggests, this type of tumor is unrelated to synovium and less than 5% arise in a joint or a bursa. Whereas any site of the body can be affected (3), 80% of SS arise into the deep soft tissue, particularly around the knee next to the joint and tendons. Typical primary SS appear as an infiltrative mass of 3-10cm in diameter or circumcised if slowly growing (3). There is no known predisposing factor leading to SS formation.

It is believed that the basis of SS development resides in its pathognomonic translocation between chromosomes 18 and X t(X;18)(p11.2;q11.2) that results in the fusion gene called *SYT-SSX*. Depending on the SSX gene to which SS18 is fused, the fusion gene can be translated into three fusion proteins, namely SYT-SSX1, SYT-SSX2 or rarely SYT-SSX4.

Histological analysis of SS tumors shows features that are clearly different from synovial tissues. The tumors typically consist of spindle cells which are small uniform ovoid cells with sparse cytoplasm and pale nuclei and nucleoli and/or epithelioid cells with abundant cytoplasm often arranged in glandular structures. Each tumor can be either monophasic, biphasic or poorly differentiated, with the majority – more than two thirds – displaying a monophasic phenotype (4). SS may be difficult to diagnose with standard immunohistochemical techniques. Therefore, molecular diagnostic tests are crucial. The biphasic tumors mainly harbor a SYT-SSX1 translocation and the vast majority of SYT-SSX2 tumors are monophasic (4).

Treatment of SS barely changed in the last decades and relies on surgery for localized disease combined in some cases with (neo)adjuvant chemotherapy, radiotherapy or both (5). However, given the rarity of SS, it is difficult to conduct large prospective trials and to determine which is the best strategy for SS patients. In advanced disease cases, cytotoxic chemotherapies (anthracyclines) remain the font-line treatment, while a tyrosine kinase inhibitor (pazopanib) has also been approved as a SS treatment drug (6). While these treatments are showing promising results, none of them are SS-specific.

1.2 SS molecular pathogenesis.

The SS fusion gene *SYT-SSX* encodes an oncogenic protein that drives the tumorigenesis. SS18 (SYT) is a ubiquitously expressed protein harboring SNH-domain and is an integral part of the chromatin

remodeling complex SWI/SNF, within which it interacts with the BRG1 subunit. BRG1 was also found to bind the IGF2 promoter, thereby modulating IGF2 expression, which appears to be required for tumor formation (7)(8). SS18 also harbors a C-terminal transactivating sequence rich in glutamine, proline, glycine and tyrosine (QPGY domain) containing transcriptional activator sequences (9).

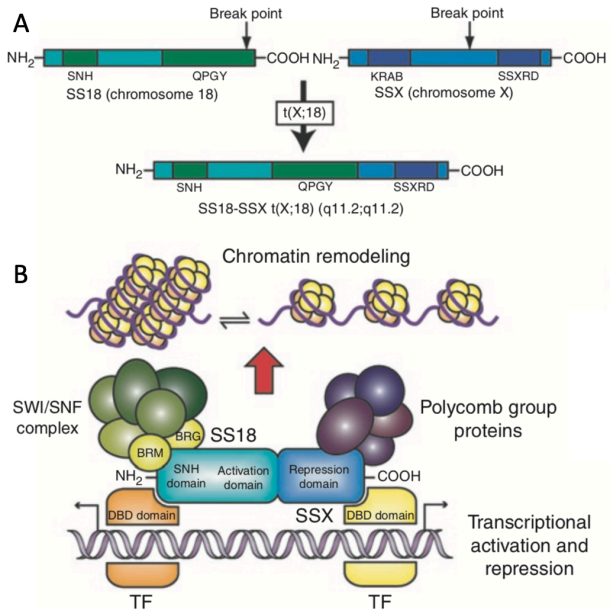


Fig. 1: Molecular genetics of SS. A: Fusion gene formation, showing breakpoints on SYT and SSX as well as recombination. B: fusion protein, its interaction with DNA, proteins and its effect on chromatin remodeling. Image adapted from (1)

SSX contains a repression domain (SSXRD) which contributes to transcriptional repression and a Kruppel-Associated box (KRAB), which is not retained in the translocation. When fused, the *SYT-SSX* gene encodes a fusion protein whose component parts have both transcriptional activation and repression activity.

In recent years, the molecular pathogenesis of sarcomas has been an active field of study, and several associations with known tumorigenic mediators have been identified, opening possibilities for potential new therapies. For example, it has been shown that the majority of SS tumors are high Bcl-2 expression neoplasms (10), without showing this protein to be a direct target of SYT-SSX. However, lower Bcl-2 levels in cell lines without the SYT-SSX fusion protein point to a possible association (11). SYT-SSX also has been shown to have a role in maintaining stem-cell properties and poor differentiation through partial Wnt-pathway activation in the absence of Wnt ligands (12). On an epigenetic level SYT-SSX is suggested to be a cellular context-dependent modifier, requiring a permissive cell state for expression and function. It participates in altering chromatin conformation to render its recognition sites accessible to resident transcription factors and induce gene expression. Many of the induced genes are developmentally regulated and participate in driving malignancy (13).

Because SYT-SSX does not have a DNA-binding domain, its interaction with DNA requires association with other proteins. Several of these have been established, particularly within the SWI/SNF complex, but additional possible interactions remain to be explored as does the mechanism of chromatin remodeling.

1.3 Genome engineering techniques

Genome editing techniques include an array of tools that can be used to manipulate virtually any gene in the genome. These techniques are based on endonucleases whose activation induces a double strand break (DSB) in the DNA, thereby activating the endogenous DNA repair mechanisms. These proteins are coupled to sequence-specific guides targeting specific DNA sites, thereby ensuring a precise DSB at a desired location of the genome. There are three mainly used genome editing techniques: zinc-fingers, TALENs and the CRISPR-Cas9 system.

An individual Zinc-finger is composed of around 30 amino-acids in a beta-beta-alpha configuration with several amino acids on the outer alpha-helix recognizing a DNA triplet in the genome. Specific DNA-targeting guides can then be developed by assembling a chain of zinc-finger proteins. The guide is coupled to a FokI endonuclease. By composing complement zinc-fingers targeting the other DNA strand also coupled to a FokI endonuclease, a DSB can be then induced (Fig. 2A)(14).

Transcription activator-like effector nucleases (TALENs) also induce a DSB by FokI endonuclease action. In this system, the DNA-sequence targeting guide is composed of proteins whose amino-acid repeats each recognize a single base pair. These transcription activator-like effectors (TALEs) are, as zinc-fingers, arranged in series and can recognize a specific DNA sequence (Fig. 2B)(14).

The clustered regularly interspaced short palindromic repeats (CRISPR) constitutes together with the Cas enzyme a microbial adaptive system whose function is to cleave foreign genetic elements and has been derived to enable precise cuts in the genomic DNA of mammalian cells. In this system, the DSB is achieved by Cas9 nuclease, guided by a CRISPR-associated RNA (crRNA). An additional auxiliary trans-activating RNA (tracrRNA) is necessary to recruit the nuclease to the target site. In experimental uses of this system, these two RNAs can be fused to form a single-guide RNA (sgRNA) consisting of a 20-nucleotide sequence complementary to the target DNA sequence. One requirement for the CRISPR-Cas9 system to work is the presence of a protospacer adjacent motif (PAM) sequence in the DNA directly after the target sequence, allowing for Cas9 recruitment and a DSB event. This PAM sequence consists of a 5'-NGG sequence (N for any nucleotide) and is present every 8-12 base pairs in the human genome(Fig. 2C) (15).

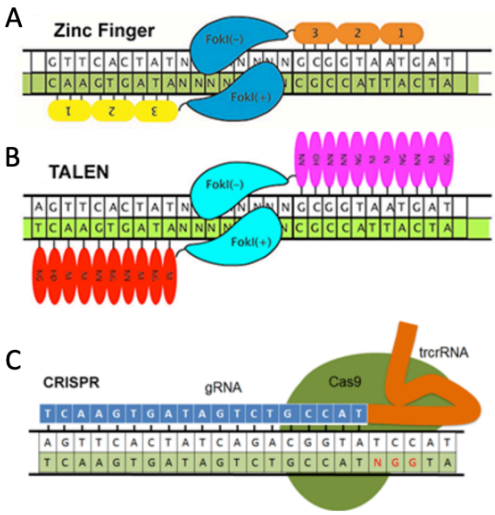


Fig. 2 : Genome editing techniques. A: zinc finger B: TALEN C: CRISPR-Cas9. Image adapted from https://www.addgene.org/genome-engineering/

Compared to the other editing techniques, CRISPRs offers a few advantages. First, the sequence coding for the sgRNA can be easily cloned and manipulated in a short time compared to TALENs (16). Then, Cas9 has a specific cut site between the 17th and 18th base pairs in the target sequence (17). In comparison, TALENs have a non-specific cut site between their monomer-pair (18). By its ease of targeting, co-delivery of different sgRNAs can also be performed simultaneously, allowing manipulation of multiple targeted sites in the cell genome(16).

1.4 DNA repair mechanisms

When a DSB event occurs, two repair pathways are induced: non-homologous end joining (NHEJ) and homology-directed repair (HDR)(16).

The NHEJ is the main repair mechanism in a mammalian cell after a DSB. It consists of the juxtaposition of the blunt ends and a DNA ligation. This process often results in the loss of one or more base pairs (19), which can be used for gene knockouts by transcription frame shifts resulting in an aberrant protein.

HDR is based on the presence of a normal copy in the other allele. It uses this undamaged sequence as a template to repair the broken region, recognizing the intact allele using homology arms on each side of the broken area. This mechanism can be used to introduce a sequence of choice in the DNA if flanked by homology arms.

To increase the insertion efficiency of a new element (payload) into the genomic DNA, a few strategies can be used:

- The Donor DNA containing the payload can be delivered as a single stranded DNA sequence. A recombinant adeno-associated virus (rAAV) can be employed for this purpose (20). This method is shown to increase HDR frequency by three magnitudes (21).
- Creating a DSB by using a CRISPR-Cas9 system, which is more precise and easier to manipulate compared to Zinc Fingers and TALENs.
- NHEJ can be inhibited by a drug, Scr7, increasing the HDR frequency up to 19 times (22).
- Providing Donor DNA homology arms of different sizes would create an asymmetric Donor DNA that increases the insertion efficiency (23).

1.5 Aim of the project

To enable the discovery of specific treatments for SS, the underlying molecular mechanisms of SS need to be better defined and interactions between SYT-SSX fusion protein and DNA should be explored. Chromatin immunoprecipitation and sequencing analysis (ChIP-Seq) is one of the most frequently used techniques to investigate protein-DNA interactions.

To conduct a successful ChIP-Seq assay, there is a need for a highly specific antibody that is able to immunoprecipitate the protein of interest efficiently while keeping a low count of off-target proteins. To date, there are no available antibodies uniquely recognizing th SYT-SSX fusion protein without cross-reacting with the wild type SYT and/or SSX protein. To overcome this problem, in this study we aim to

modify genome of an established SS cell line (HS-SYII) to introduce an artificial V5 epitope tag in the fusion protein, which can be subsequently recognized by a V5 specific antibody.

Tagging of genomic SYT-SSX is a gene knock-in that relies on HDR to drive targeted gene replacement. Insertion of V5 tag sequence to the intended locus will be triggered by the CRISPR-Cas9 system that creates a double strand break near the desired genome editing point. The HDR donor sequence will be introduced into cells through a rAAV transduction. It will contain two homologous regions that flank the V5-tag, the 2A sequence and a positive selection marker, neomycin (Neo). After HDR occurs, the bulk population of recombinant cells harboring V5- tagged SYT-SSX can be selected by neomycin treatment. The expression and localization of SYT-SSX-V5 can be confirmed by polymerase chain reaction (PCR), western blot and immunofluorescence. Upon confirmation of SYT-SSX- V5 harboring cells, ChIP-Seq studies will be started.

2 Results

2.1 Finding the exact breakpoint region of SYT-SSX translocation in the HSSYII genome

While genomic sequences coding for SYT and SSX are well known and available on NCBI gene database (24), the exact break point location of the chromosomal translocation in between SYT on chromosome 18 and SSX on chromosome X, t(X;18)(p11;q11) has not been established for each cell line yet.

To investigate the breakpoint region of the SYT-SSX translocation in the HSSYII cell line, we performed genome sequencing covering the possible breakpoint regions on the HSSYII genome. The translocation process results in a loss of an unknown length of genomic sequence, which needed to be exactly defined. The initial steps to get the exact sequence was the extraction of genomic DNA and subsequent amplification of the possible breakpoint region by PCR covering the SYT-SSX transition, using Forward cloning primer (FCP) and Reverse cloning primer (RCP); namely primers 1 & 2 (Table 1).

	Primer Name	Primer sequence
1	Forward cloning primer (FCP)	5’-TGGATATGACCAGGTAAGTAGTTTG-3’
2	Reverse cloning primer (RCP)	3’-GTGTAATATTTCTAGTACGGG-5’
3	Forward sequencing primer (FSP)	5’-CCTCAACTAGAAACTCCAATAC-3’
4	Reverse sequencing primer (RSP)	3’-AACCCAACTCCCAGATCC-5’
5	Homology arms forward primer (HAF)	5’-CCCTTTGAAATCCCTCCACTTG-3’
6	Homology arms reverse primer (HAR)	3’-CCCGTCTTGTTCCCTGATGTC-5’
7	Right homology arm forward primer (RHAF)	5’-ATCTTGTCATTTGTTTCCCAAATC-3’
8	Right homology arm reverse primer (RHAR)	3’-CACTCAGAACTGCCCTCAGTAG-5’
9	Left homology arm forward primer (LHAF)	5’-ACTTGCTGCAGCTTCTCCATCA-3’
10	Left homology arm reverse primer (LHAR)	3’-CATGGGCCTTCTTTATCCAG-5’

Table 1 : Primer list

Since the breakpoint mapping of SYT-SSX has not been performed for the HSSYII cell line, the exact length of the PCR product using FCP and RCP primers could not be determined ahead. Thus, we tried different extension times for the amplification of the PCR product, namely 1,2 and 3 minutes. The PCR products resulting from different elongation times were run on an agarose gel (0.7%). We could only get a PCR product with the elongation time of 3 minutes, which was around 5 to 7 kb long (Fig.3). Most probably, the elongation times of 1 and 2 minutes were insufficient to create a PCR product about 5-7 kb in length for the DNA polymerase which is able to synthetize 2kb of DNA in a minute originally. Therefore, it needs more than two minutes to polymerize a product longer than 4kb.

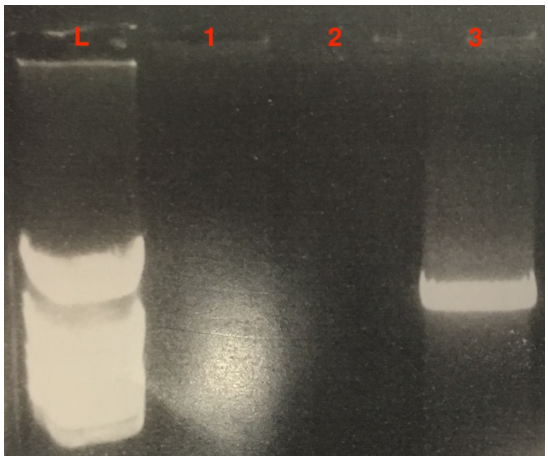


Fig. 3 : Gel run of PCR products (1), (2), (3) produced by PCR elongation times of 1, 2 and 3 minutes respectively. lambda DNA-BstEII digest ladder (New England Biolabs) (L) was run as a reference.

The corresponding product with 3 minutes of elongation time contained the breakpoint region. Hence, it was gel extracted and inserted into a TOPO vector by blunt end ligation. This plasmid named TOPO-H1 was subsequently sequenced. Finding the exact breakpoint region of the translocation was achieved by sequencing the PCR product inserted into Topo vector. The breakpoint was found using Forward sequencing primer (FSP) and Reverse sequencing primer (RSP); namely: primers 3 & 4 (Table 1).

Comparing to the full-length genomic sequences of SYT and SSX genes, we found that a 8625 bp-long region was lost during the SYT-SSX translocation in the HS-SYII genome.

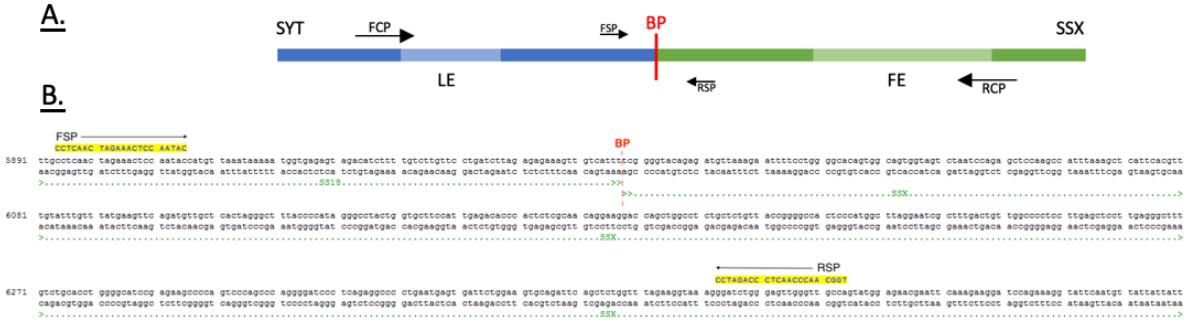


Fig. 4 : Translocation breakpoint. A: Forward cloning primer (FCP) initiates 5' to 3' transcription from SYT last exon (LE) while reverse cloning primers (RCP) initiates 3' to 5' transcription from SSX first exon. PCR amplification covers the breakpoint (BP). B: Exact sequence of breakpoint (BP) region, sequenced using forward sequencing primer (FSP) and reverse sequencing primer (RSP), which showed an overlap in the sequences they allowed to acquire.

2.2 Design and cloning of the CRISPRs targeting the region of interest

In this study we wanted to modify genome of HSSYII to introduce an artificial V5 epitope tag in the fusion protein, which can be subsequently recognized by a V5 specific antibody.

This gene knock-in uses the HDR pathway to drive targeted gene replacement. Insertion of V5 tag sequence to the intended locus was triggered by a DSB induced by a CRISPR-Cas9 system.

A CRISPR-Cas9 complex is composed of two components. A 20 nucleotide RNA (guide RNA) which is the reverse complement to the targeting region on DNA, and Cas9 protein that induces a double strand break near the desired genome editing point, which in turn activates the DNA repair mechanisms.

The target region on the genome, reverse complement to the guide RNA, needs to be directly adjacent to a PAM sequence. We searched a target region in the vicinity of the last SYT exon and its stop codon using online tools: namely crispr.mit.edu and chopchop.cbu.uib.no. Two guide RNA sequences with the highest scores were chosen to continue with cloning experiments.

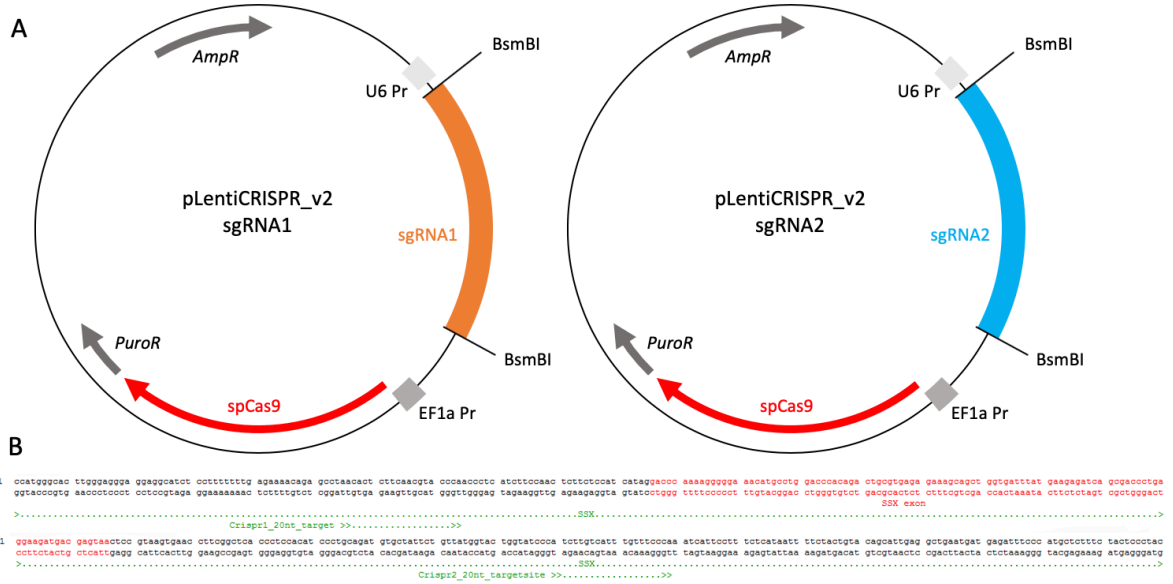


Fig. 5: Crispr system. A : sgRNA 1&2 in a pLentiCRISPRv2 vector containing Ampicillin resistance gene (AmpR), puromycin resistance gene (PuroR) and Streptococcus-derived Cas9 nuclease. U6 promoter & EF1-alpha promoter regulate gene expression in this plasmid. BsmBI restriction sites allow for sgRNA insertion. B: CRISPR 1 & 2 target sites around the last SYT-SSX exon.

The oligonucleotides coding for the top two guide RNAs were inserted into a plasmid producing Cas9 protein (#52961, Addgene). In this vector, the expression of Cas9 is under the control of EF1-alpha promoter and the transcription of the sgRNA is regulated by U6 promoter (Fig 5A). Furthermore, the backbone carries antibiotic resistance genes, for instance; ampicillin is to select transfected bacteria and puromycin is to select infected HSSYII cells.

2.3 Detailed design of the donor DNA

We design a HDR donor sequence (donor DNA) which can be introduced to cells through a rAAV. Donor DNA contains 3 elements: a left homology arm (LHA), a right homology arm (RHA) and the payload. The payload consists of the V5-tag sequence, 2A sequence, two LoxP sites and the Neomycine resistance gene. Besides, LHA and RHA of different sizes flank the payload asymmetrically (Fig 6 & 7).

We used a traditional cloning based on restriction sites and ligation to generate our rAAV plasmid carrying homology arms and payload sequence, all of which creates donor DNA (Fig 6).

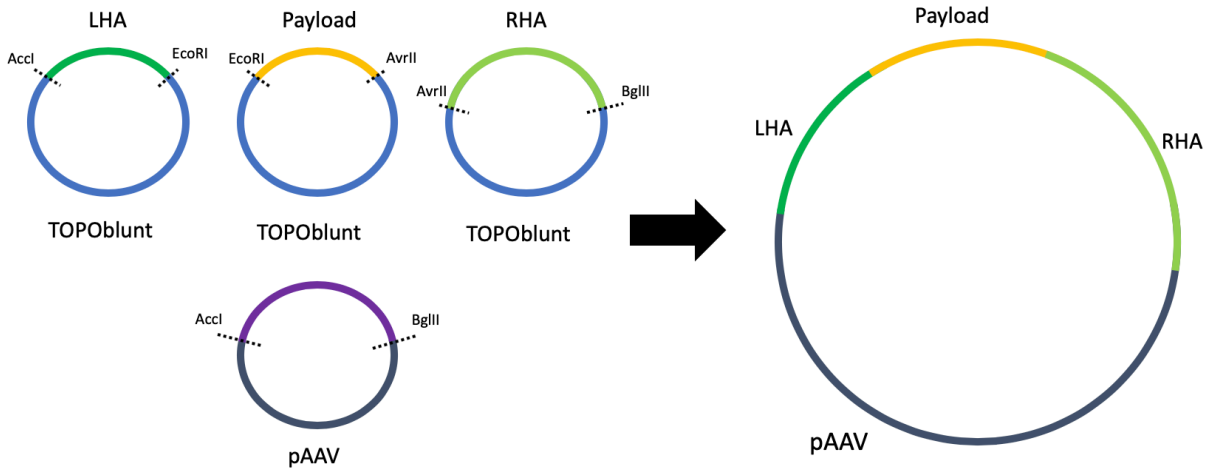


Fig. 6: Assembly of donor plasmid. Left homology arm (LHA) is digested by AccI and EcoRI restriction enzymes. Donor DNA (payload) is digested with EcoRI and AvrII restriction enzyme. Right homology arm (RHA) is digested with AvrII and BglII restriction enzymes. These three parts were extracted from a TOPOblunt backbone. The Adeno-Associated Virus backbone (pAAV) is digested with AccI and BglII restriction enzymes. The four parts are then assembled by four-way ligation to construct the definitive plasmid carrying the donor DNA. Schemes are not in scale.

2.3.1 Homology arms

The genomic DNA sequence containing both of the homology arms were amplified from the previously established TOPO-H1 plasmid using primers 5&6 (Table 1). The subsequent PCR product was gel extracted and cloned into a TOPO vector. Then, we created RHA and LHA using pairs of primers 7-8 and 9-10, respectively. The PCR products were gel extracted and inserted into a TOPOblunt vector individually to create LHA-TOPOblunt and RHA-TOPOblunt plasmids (Fig 6).

2.3.2 Payload

The payload DNA is the sequence carrying the V5 tag, which is in frame with the stop codon. It is also designed to provide tools to detect random insertions and to allow for later knockout experiments.

By designing a promoterless system, we insured that the antibiotic selection process would select only the cells in which the payload is inserted in the target location. This way, the Neo resistance is acquired only if the donor sequence is inserted into the target location, which is under the control of SSX exon promoter. In this design the *Neo* region is preceded by a 2A sequence allowing for the independent transcription of Neomycin resistance while the V5 tag is being expressed.

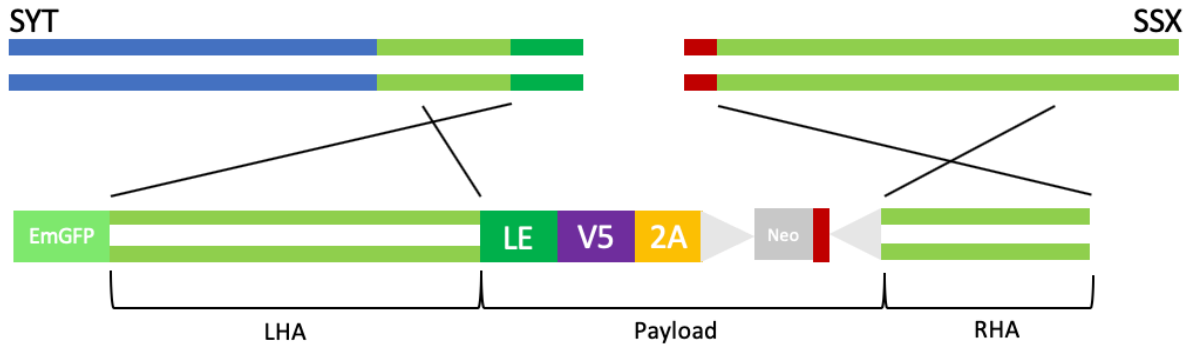


Fig. 7: HDR scheme. Upper panel: double strand break in the SYT-SSX fusion chromosome, around the stop codon (in red). Lower panel: Donor DNA with Emerald Green Fluorescent Protein (EmGFP), left (LHA) and right (RH) homology arms flanking the payload (GBlock). The payload is constructed with the sequences of the last exon (LE), the V5 tag, the 2A sequence and the LoxPsites (triangles) around the Neo region and the stop codon (in red). Schemes are not in scale.

The Emerald Green Fluorescent Protein (EmGFP) is a tool to detect random insertions which can happen in off-target parts of the genome. Normally, it is expected that the cells having a random insertion would not express EmGFP or any other coding sequences in payload since donor DNA does not contain any promoter. However, the cells containing a random insertion could still express EmGFP, if only the random insertion is in frame with an active gene being expressed which is very unlikely. Even it happens, cells with EmGFP expression can be detected and eliminated by cell sorting. On the other hand, being placed upstream of the left homology arm in 5' direction, the EmGFP sequence would not be introduced in the genome in case of HDR event in the targeted DNA region.

```
361  acccaaccct catcttccaa ctcttctcca tcataggacc caaaaggggg aaacatgcct ggacccacag actgcgtgag agaaagcagc
     tgggttggga gtagaaggtt gagaagaggt agtatcctgg gttttccccc tttgtacgga cctgggtgtc tgacgcactc tctttcgtcg
                                                                  SSX exon2
     >............'HR SSX 5'............>>
     >.....................................gblock.....................................>
     >..>> Crispr1_20nt_target

451  tggtgattta tgaagagatc agcgaccctg aggaagatga cgagggtaag cctatcccta accctctcct cggtctcgat tctacgggaa
     accactaaat acttctctag tcgctgggac tccttctact gctcccattc ggatagggat tgggagagga gccagagcta agatgccctt
     >.....................................gblock.....................................>

                                          >>.................V5tag...................>>
                                                                                  2A >>.>

541  gcggagctac taacttcagc ctgctgaagc aggctggaga cgtggaggag aaccctggac ctatgataac ttcgtatagc atacattata
     cgcctcgatg attgaagtcg gacgacttcg tccgacctct gcacctcctc ttgggacctg gatactattg aagcatatcg tatgtaatat
                                                                            Lox66
     >.....................................gblock.....................................>
     >............................2A............................>>

631  cgaacggtac catgggatcg gccattgaac aagatggatt gcacgcaggt tctccggccg cttgggtgga gaggctattc ggctatgact
     gcttgccatg gtaccctagc cggtaacttg ttctacctaa cgtgcgtcca agaggccggc gaacccacct ctccgataag ccgatactga
     >.....................................gblock.....................................>
               >>.................................Neomycine.................................>
     :
     :
1441 tctaaataac ttcgtataat gtatgctata cgaacggtac tccgtaagtg aaccttcggc tcaccctcca catccctgca gatgtgctat
     agatttattg aagcatatta catacgatat gcttgccatg aggcattcac ttggaagccg agtgggaggt gtagggacgt ctacacgata
                       iLox71
     >.....................................gblock.....................................>
     >..>> Neomycine
                                        >>.................HR SSX 3''.................>
```

Fig. 8 : Payload sequence. The Payload sequence consists in the last SYT-SSX exon followed by the V5 tag sequence, the 2A sequence and Neo region, which is flanked by two LoxP sites facing each other.

Moreover, there are *LoxP* sites on both sides of the *Neo* sequence, allowing for knockout experiments once the payload is inserted. This knockout would be initiated by *Cre* enzyme inducing the sequence to flip over, shifting the stop codon in the non-sense mediated RNA decay window and thereby inducing the mRNA degradation and knocking the SYT-SSX fusion protein out.

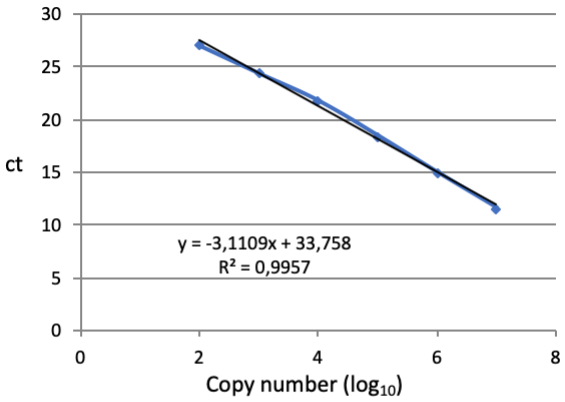
The synthetic DNA fragment containing the payload sequence was inserted into a TOPOBlunt backbone. Eventually, it was cloned into pAAV backbone together with LHA and RHA (Fig. 6).

2.4 Viral production and infection of HSSY2 cells

2.4.1 CRISPR lentivirus

To produce the lentiviruses for infection of HSSYII cells with CRISPR vectors, we performed liposome-based transfection of HEK293T cells using FuGENE HD transfection Reagent. We co-transfected the CRISPR1 & CRISPR2 plasmids together with envelope and packaging plasmids to produce lentiviruses. After 48h, the cells were harvested and the virus were collected (see materials and method, part 4.6), thereby producing a virus stock.

A



B

Sample	Copy nb/ul	Infected cell nb	MOI (copy/cell)
Crispr1_v2	213'384,8	0,4 M	53,3
Crispr2_v2	125'248,5	0,4 M	31,3
HS-SYII_WT	17,8	0,4 M	0,004

Fig. 9: Lentiviral titration results. A : Graph showing the ct values (y-axis) according to the copy numbers (x-axis, in log_{10} scale). B: table with effective virus copy number, number of cells infected and multiplicity of infection for each CRISPR lentivirus.

After transfection and virus harvest, we separately infected 0.4 M HS-SYII cells with CRISPR1 & 2 and titrated the virus using RT-qPCR. An uninfected sample (HS-SYII_WT) of the same cell number was also analyzed as control. From these results we then calculated the multiplicity of infection (MOI), corresponding to the viral copy number integrated in the genome of the infected cells. We were also able to calculate the viral copy concentration (copy number per ul) of our lentiviral stock harvested from HEK293T cells. These results can be used for later dual infection with a wanted MOI, insuring for the vast majority of cells to be infected with our virus.

2.4.2 rAAV production

The adeno-associated virus was produced in a similar manner to lentiviruses. It was co-transfected on HEK293T cells with envelope and packaging sequences and thereby producing a stock from which we used different volumes to infect HS-SYII cells.

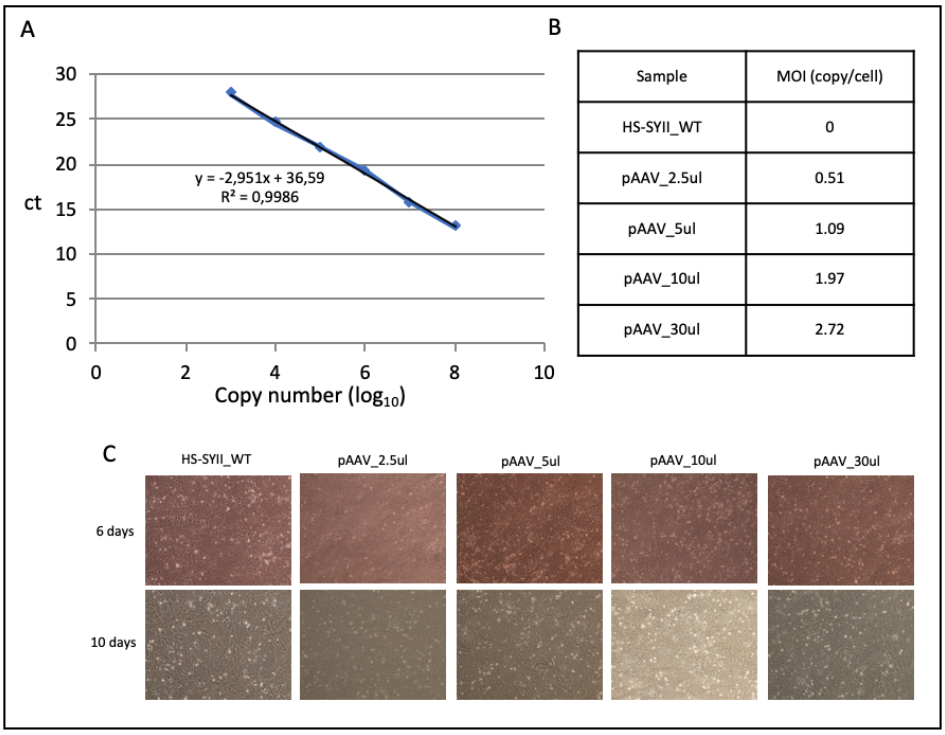


Fig. 10: Adeno-associated virus titration results and cell viability test. A: Graph showing the ct values (y-axis) according to the copy number of donor sequence (x-axis, in log_{10} scale). B: Table showing MOI results with different volumes of virus stock and the MOI corresponding to each volume. C: cell viability test after infection, pictures at 10x after 6 and 10 days of infection.

We titrated the viral infection by real-time quantitative polymerase chain reaction (RT-qPCR) (fig 10A) and calculated the MOI from the results (fig. 10B). The cells were then kept in culture to determine their viability (fig. 10C) after infection. Cells kept growing after infection showing that the infection with pAAV plasmid had no toxicity up to an infection with 30 µl of our virus stock.

2.5 Dual infection

A dual transduction with rAAV-Donor and both CRISPR vectors was performed as a preliminary infection with use of Scr7. The co-infected HS-SYII cells were then selected for 10 days using Neomycin. At the end of the selection process, only a few bulk populations of cells were found on the plate which are expected to have V5 tag integration in their genome (see fig. 11).

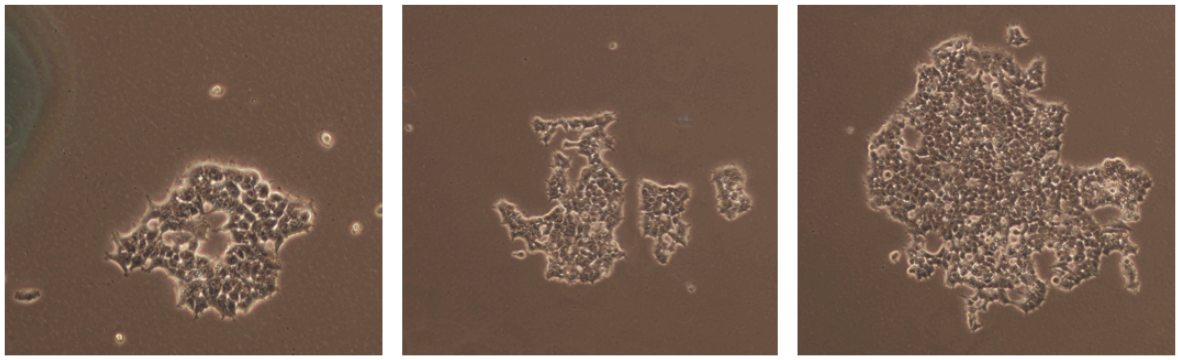


Fig. 11 : Bulk HS-SYII populations after a 10-days selection under Neomycin.

3 Discussion

In this project we have demonstrated the work needed to introduce a high-affinity and low cross-reacting antibody target sequence specific to the SYT-SSX fusion protein, namely the V5 tag sequence. This genomic engineering requires using the HDR pathway.

Because HDR events are rare, we enhanced their efficiency by delivering our payload in a rAAV vector and flanked by two asymmetrical homology arms and used a CRIPR-Cas9 system to induce a precise DSB in the required location. We inhibited NHEJ by administering Scr7 at the time of infection.

The payload was designed to harbor the V5-tag sequence in frame with the last *SYT-SSX exon.* It also contains a *Neomycin resistance* gene to detect random insertions, which is preceded by a 2A sequence allowing for the transcription of an independent mRNA simultaneously as the SYT-SSX transcription. This *Neo* region is flanked by two LoxP sites for later knockout of SYT-SSX by Cre recombination.

We demonstrated that the HSSYII cells are infectible with rAAV-Donor as well as CRISPR lentiviruses. A dual infection was performed as a preliminary experiment and the cells that survived the selection process remain to be tested for the harboring of the V5-tag. The V5-tag can be detected by Western-blotting using a specific anti-V5 antibody and looking for a protein with a molecular weight compatible in size with SYT-SSX-V5.

A precise double infection with one of the CRISPRs and the rAAV-Donor virus still has to be performed, with controlled volumes from the produced virus stocks aiming for a MOI of 4-5. This would insure the infection of approximately 99% of the cells with at least one viral particle. Again, the presence of the SYT-SSX fusion protein would be checked by Western-blotting. Following the positive selection with Neomycin, a RT-qPCR will be performed with primers specific to SYT-SSX and to V5-tag. Finally, a genomic sequencing of the targeted DNA region will insure for the correct sequence of the inserted payload.

This insertion allows for a ChiP-seq analysis precipitating the SYT-SSX fusion protein. This will allow the analysis of DNA binding sites and chromatin state in edited cells. Recent works using a similar technique have shown the fusion protein to deregulate developmental programs by hijacking the

repressive complex KDM2B-PRC1.1 (25). This in turn is proposed to drive synovial sarcoma oncogenesis.

4 Material and Methods

4.1 Extraction of genomic DNA and amplification of possible breakpoint locus

The genomic DNA was extracted from HSSYII cells using a kit (QIAgen) according to manufacturer's instructions. Possible breakpoint locus on genome was amplified by PCR using Phusion polymerase (New England Biolabs) and an elongation time of 1, 2 and 3 minutes. The products were run on a gel and the PCR product of the 3 min elongation time was extracted from the gel (Qiagen). It was then inserted into a TOPO vector by ligation with an incubation time of 10 min at 25°C.

4.2 Insertion vector

The donor DNA was inserted into pAAV backbone in a two-step digestion with restriction enzymes Acc1 and Bgl2 with an incubation time of 1hr at 37°C. The two products were separated from each other by electrophoresis in an agarose gel. The product corresponding to the backbone had a total length of 4150bp and was then extracted.

The 5'HA homology arm was amplified from the previously extracted genomic DNA (TOPO-H1) and inserted into TOPOBlunt vector PCR blunt ligation with an incubation time of 10min at 25°C. The result was then precipitated using butanol and resuspended in H_2O. It was electroporated into Stbl3 and incubated at 37°C overnight. After PCR screening two colonies were chosen, miniprepped and sent to sequencing. It was then prepared for ligation with a one-step double digestion by Acc1 and EcoR1-HF at 37°C for 1h and products separated using electrophoresis on an agarose gel. The corresponding product was identified with its length of 1010bp and extracted.

The 3'HA homology arm was also amplified from the extracted genomic DNA and ligated by a PCR blunt ligation for 10min at 25°C, resuspended in H_2O and electroporated into Stbl3 and incubated overnight. Three colonies from the PCR screening were miniprepped and sent to sequencing. It was then prepared for ligation with a two-step double digestion by Acc1 and EcoR1-HF at 37°C for 1h and products separated using electrophoresis on an agarose gel. The corresponding product was identified with its length of 581bp and extracted.

The payload insert was designed and ordered from Integrated DNA technologies. It was inserted in a TOPOblunt vector by blunt-end ligation with an incubation time of 10 min at 25°C. The product was precipitated with butanol, resuspended in H_20, electroporated in Stbl3 strain of *E. coli* and incubated overnight. After PCR screening 3 colonies were chosen and sent to sequencing. The preparation for ligation with homology arms consisted in a one-step double digestion using first AvrII and then Bgl2, both with an incubation time of 1h at 37°C. The products were separated with an electrophoresis on an agarose gel. The expected product of 1721bp was then extracted.

The ligation of the four products was performed in a 4-way ligation using T4 ligase with an incubation time of 10min at 25°C. The product was then precipitated with butanol, resuspended in H_2O and electroporated into GLSHOT strain of *E. coli* and incubated for 72h at 24°C. After a PCR screening, 3 colonies were chosen for the subsequent miniprep plasmid extraction and sent for sequencing. One of them with the correct sequence was then chosen to continue.

All the gel extractions were performed using QIAquick extraction kit from Qiagen and the products were then resuspended in elution buffer from the kit.

All sequencings were performed using Sanger sequencing method. This process was performed by Microsynth AG.

4.3 CRISPR cloning

The two guide RNA sequences cloned into CRISPR plasmid were designed using crispr.mit.edu and chopchop.cbu.uib.no online tools. They were designed with overhangs 5'-3' CACC and 3'-5' AAAC, making them compatible for subsequent sticky-end ligation with the lentiviral vector

The guide RNA oligonucleotides were suspended in annealing buffer and annealed on a thermocycler with 70 cycles consisting of 5minutes at 95°C followed by 30seconds at -1°C, producing two 10 uM annealed CRISPR stocks. Before insertion into the lentiviral vector, the annealed stocks were diluted 1/100 to a concentration of 100nM.

The lentiviral vector lentiCRISPRv2 was ordered on Addgene website (26). It was suspended according to manufacturer specifications. It was prepared for ligation by digestion with restriction enzyme BsmbI with an incubation time of 1h at 37°C. The products were separated with agarose gel electrophoresis and the product of 12988bp, corresponding to the vector, was then extracted.

The annealed guide RNA oligonucleotides and vector backbone were ligated together by sticky-end ligation with T4 ligase with an incubation time of 10min at 65°C. The product was precipitated with butanol, electroporated into GLSHOT strain of *E.coli* and incubated overnight. After miniprep, five colonies were chosen and sent to sequencing. One colony of each CRISPR construct was chosen to continue.

4.4 Cell culture

HS-SYII cells were cultured in DMEM with 10% fetal bovine serum (FBS)(Gibco) and 1% penicillin-streptomycin (PS)(Gibco). They were cultured in 15cm petri dishes and kept at 37°C, 5% CO_2.

HEK 293T cells were cultured in DMEM with 10% FBS, 1% PS and 1% non-essential amino-acids (NeAA)(Gibco). They were cultured in 15cm petri dishes and keep at 37°C, 5% CO_2.

4.5 Virus production

The viruses were produced using the Fugene HD transfection kit (Promega).

Regarding both lentivirus production of CRISPR constructs, we used pMD2G as envelope plasmid and PCMVDelta R8.74 as packaging plasmid. They were then transferred on 4 million HEK 293T and the cell medium was changed after 24h.

pHelper and pRC2 were used respectively, as envelope and packaging plasmids for the production of rAAV containing the Donor DNA. The reaction was transferred onto 12 million HEK 293T cells and the cell medium was changed after 12h.

4.6 Virus harvesting

The rAAV virus carrying the donor DNA was harvested using AAVpro Helper Free System (TAKARA BIO Inc.) reagents 48h after transfection.

The CRISPR viruses were harvested using Lenti-X Concentrator (Clontech Laboratories Inc.) reagent 48h after transfection.

4.7 Infection and selection of HSSYII cells

We first infected cells with only CRISPR 1 or 2 lentivirus and rAAV independently.

HSSYII cells were infected with the lentiviruses carrying CRISPR backbones or rAAV containing the donor DNA. The mediums of the cells were changed 10h after infection. We selected the infected cells using Neomycin (1mg/ml) for 10 days.

A dual infection was performed as a preliminary experiment with rAAV extracted particles (2.5ul of stock) together with the filtrated medium of CRISPR1 infected HEK293T cells. We used 2ul of Scr7, thereby inhibiting NHEJ.

4.8 Virus titration

rAAV-Donor viruses were titrated and mulptiplicity of infection was calculated by RT-qPCR using AAVpro Titration kit (TAKARA BIO Inc.) reagents.

CRISPR 1&2 lentiviruses were titrated using a protocol developed by Barde et al. (27), and adapted for our experiment.

Acknowledgements

I would like to thank Prof. Ivan Stamenkovic for his kindness and availability.
A special thanks goes to Dr. Tugba Keskin for her never-ending patience, her availability, her teaching and support.
I would also like to thank Dr. Carlo Fusco and Dr. Luisa Cironi for their answers to my questions.
Thanks to Prof. Tatiana Petrova for her time and contribution as an expert.
Finally, I would like to thank all the people working in Prof. Stamenkovic's Lab for making me feel part of the lab.

Bibliography

1. Anderson JL, Denny CT, Tap WD, Federman N. Pediatric sarcomas: translating molecular pathogenesis of disease to novel therapeutic possibilities. Pediatr Res [Internet]. 2012 Aug [cited 2017 Aug 29];72(2):112–21. Available from: http://www.nature.com/doifinder/10.1038/pr.2012.54
2. Kransdorf MJ. Malignant soft-tissue tumors in a large referral population: distribution of diagnoses by age, sex, and location. Am J Roentgenol [Internet]. 1995 Jan [cited 2018 Dec 10];164(1):129–34. Available from: http://www.ajronline.org/doi/10.2214/ajr.164.1.7998525
3. Fletcher CDM, Unni KK, Mertens F, Weltgesundheitsorganisation, International Agency for Research on Cancer, editors. Pathology and genetics of tumours of soft tissue and bone ;. Lyon: IARC Press; 2002. 427 p. (World Health Organization Classification of tumours).
4. Lange S, Kremer J, Schultze M, Van Tine B. Current Management and Molecular Targets of Synovial Sarcoma. Austin J Cancer Clin Res. 2014;1(2):1–9.
5. Vlenterie M, Jones RL, van der Graaf WTA. Synovial sarcoma diagnosis and management in the era of targeted therapies: Curr Opin Oncol [Internet]. 2015 Jul [cited 2018 Dec 14];27(4):316–22. Available from:
6. Stacchiotti S, Van Tine BA. Synovial Sarcoma: Current Concepts and Future Perspectives. J Clin Oncol [Internet]. 2018 Jan 10 [cited 2018 Dec 14];36(2):180–7.
7. de Bruijn DRH, Allander SV, van Dijk AHA, Willemse MP, Thijssen J, van Groningen JJM, et al. The Synovial Sarcoma–Associated SS18-SSX2 Fusion Protein Induces Epigenetic Gene (De)Regulation. Cancer Res [Internet]. 2006 Oct 1 [cited 2018 Dec 14];66(19):9474–82.
8. Sun Y, Gao D, Liu Y, Huang J, Lessnick S, Tanaka S. IGF2 is critical for tumorigenesis by synovial sarcoma oncoprotein SYT-SSX1. Oncogene [Internet]. 2006 Feb [cited 2018 Dec 14];25(7):1042–52.
9. Thaete C, Brett D, Monaghan P, Whitehouse S, Rennie G, Rayner E, et al. Functional domains of the SYT and SYT-SSX synovial sarcoma translocation proteins and co-localization with the SNF protein BRM in the nucleus. Hum Mol Genet. 1999 Apr;8(4):585–91.
10. Krsková L, Kalinová M, Břízová H, Mrhalová M, Sumerauer D, Kodet R. Molecular and immunohistochemical analyses of BCL2, KI-67, and cyclin D1 expression in synovial sarcoma. Cancer Genet Cytogenet [Internet]. 2009 Aug [cited 2018 Dec 16];193(1):1–8.
11. Albritton KH, Randall RL. Prospects for targeted therapy of synovial sarcoma. J Pediatr Hematol Oncol. 2005 Apr;27(4):219–22.
12. Cironi L, Petricevic T, Fernandes Vieira V, Provero P, Fusco C, Cornaz S, et al. The fusion protein SS18-SSX1 employs core Wnt pathway transcription factors to induce a partial Wnt signature in synovial sarcoma. Sci Rep [Internet]. 2016 Apr [cited 2018 Dec 14];6(1). A
13. Tamaki S, Fukuta M, Sekiguchi K, Jin Y, Nagata S, Hayakawa K, et al. SS18-SSX, the Oncogenic Fusion Protein in Synovial Sarcoma, Is a Cellular Context-Dependent Epigenetic Modifier. PLoS ONE [Internet]. 2015 Nov 16 [cited 2017 Jun 6];10(11).
14. Gaj T, Gersbach CA, Barbas CF. ZFN, TALEN, and CRISPR/Cas-based methods for genome engineering. Trends Biotechnol [Internet]. 2013 Jul [cited 2018 Dec 17];31(7):397–405.
15. Cong L, Ran FA, Cox D, Lin S, Barretto R, Habib N, et al. Multiplex Genome Engineering Using CRISPR/Cas Systems. Science [Internet]. 2013 Feb 15 [cited 2018 Dec 17];339(6121):819–23.
16. Ran FA, Hsu PD, Wright J, Agarwala V, Scott DA, Zhang F. Genome engineering using the CRISPR-Cas9 system. Nat Protoc [Internet]. 2013 Nov [cited 2018 Dec 17];8(11):2281–308.
17. Jinek M, Chylinski K, Fonfara I, Hauer M, Doudna JA, Charpentier E. A Programmable Dual-RNA-Guided DNA Endonuclease in Adaptive Bacterial Immunity. Science [Internet]. 2012 Aug 17 [cited 2018 Dec 17];337(6096):816–21.
18. Miller JC, Tan S, Qiao G, Barlow KA, Wang J, Xia DF, et al. A TALE nuclease architecture for efficient genome editing. Nat Biotechnol [Internet]. 2011 Feb [cited 2018 Dec 17];29(2):143–8.
19. Alberts B, editor. Molecular biology of the cell. 4th ed. New York: Garland Science; 2002. 1548 p.
20. Chen X, Gonçalves MAFV. Engineered Viruses as Genome Editing Devices. Mol Ther

[Internet]. 2016 Mar [cited 2018 Dec 17];24(3):447–57.
21. Russell DW, Hirata RK. Human gene targeting by viral vectors. Nat Genet [Internet]. 1998 Apr [cited 2018 Dec 17];18(4):325–30.
22. Maruyama T, Dougan SK, Truttmann MC, Bilate AM, Ingram JR, Ploegh HL. Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining. Nat Biotechnol [Internet]. 2015 May [cited 2018 Dec 17];33(5):538–42.
23. Richardson CD, Ray GJ, DeWitt MA, Curie GL, Corn JE. Enhancing homology-directed genome editing by catalytically active and inactive CRISPR-Cas9 using asymmetric donor DNA. Nat Biotechnol [Internet]. 2016 Mar [cited 2018 Dec 27];34(3):339–44.
24. Home - Gene - NCBI [Internet]. [cited 2019 Jan 5]. https://www.ncbi.nlm.nih.gov/gene
25. Banito A, Li X, Laporte AN, Roe J-S, Sanchez-Vega F, Huang C-H, et al. The SS18-SSX Oncoprotein Hijacks KDM2B-PRC1.1 to Drive Synovial Sarcoma. Cancer Cell [Internet]. 2018 Mar [cited 2018 Dec 27];33(3):527-541.e8.
26. LentiCRISPRv2 [Internet]. https://www.addgene.org/52961/
27. Barde I, Salmon P, Trono D. Production and Titration of Lentiviral Vectors. In: Crawley JN, Gerfen CR, Rogawski MA, Sibley DR, Skolnick P, Wray S, editors. Current Protocols in Neuroscience [Internet]. Hoboken, NJ, USA: John Wiley & Sons, Inc.; 2010 [cited 2019 Jan 5].