NTNU – Trondheim
Norwegian University of
Science and Technology

Role of FOXA genes in maintaining transformed phenotype in human bronchial epithelial cells

Construction of a CRISPR-Cas9 vector system for site-specific integration of *FOXA1*

Oda Astrid Haarr Foss

Biotechnology
Submission date: June 2015
Supervisor: Aage Haugen, IBT
Co-supervisor: Steen Mollerup, STAMI
Audun Trygge Haugen Bersaas, STAMI

Norwegian University of Science and Technology
Department of Biotechnology

Acknowledgemens

This project was carried out at the National Institute of Occupational Health (STAMI) in Oslo as a part of the Master of Science in Biotechnology at the Norwegian University of Science and Technology (NTNU). I wish to thank my supervisor Professor Aage Haugen, head of the toxicology department for giving me this opportunity, and a special thanks to my main supervisor in this project, Steen Mollerup (PhD), for excellent guidance, answering all my questions and trusting me with challenging work. I wish also to thank my co-supervisor Audun T. Bersaas for great guidance and cooperation with *in silico* design, for your inspiring enthusiasm and our interesting academic conversations.

For tutoring in laboratory and teaching me invaluable methods and techniques, I would like to thank research fellow Iselin Rynning and staff engineer Rita Bæra. Especially I would thank Iselin for being such a great example of immaculate lab work.

I thank the Toxicology department for making me feel included by always being helpful with questions regarding lab work, and by showing me a great work environment and inviting me to social events.

Working on this project would not be the same without my fellow students Heidi Notø and Umer Anayyat in the fifth floor, thanks for good times.

And lastly my supporting family, friends and my beloved Jon Helge S. Jakobsen; I thank you for always wishing me the best.

Oslo, June 2015

Oda Astrid Haarr Foss

Abstract

Lung cancer is the number one cause of worldwide cancer deaths for men, and second cause of cancer deaths for women. Tobacco smoking is strongly associated with lung cancer, and carcinogenic compounds inhaled when smoking cause DNA damage of lung tissue when they are metabolized in the cells. Epithelial-mesenchymal transition (EMT) is a process in tumorigenesis where the cells gain an invasive and migratory phenotype, and it mediates metastasis of the primary tumor. EMT is characterized by specific markers expressed from the mesenchymal cells. A possible connection between steroid receptor pathways and carcinogens has been proposed, and transcription factors (FOXA family) regulating these pathways may therefore play an important part in regulating carcinogen metabolism. *FOXA1* and *FOXA2* are downregulated in human bronchial epithelial cells (HBECs) that are exposed to cigarette smoke condensate. As these cells also have gained a mesenchymal phenotype, this downregulation supports the notion that FOXA factors may be involved in EMT.

Ectopic expression of *FOXA1* in the transformed HBECs is an interesting way of studying whether the factor alone plays a significant part in EMT. In this thesis, we aimed to

1) Develop a method for controlled integration of *FOXA1* into transformed HBEC cells

2) Restoring close to normal expression levels in order to study the role of FOXA1 in regulating mesenchymal markers and phenotype

To achieve these goals, we used CRISPR-Cas9 system for RNA guided site-specific genome engineering, and designed an expression construct to be integrated into the target genome at a genomic safe harbor target locus. The CRISPR-Cas9 system creates a double strand break at a desired locus, and this break may be repaired by homologous recombination using a repair construct containing the gene of interest (*FOXA1)*. The *FOXA1* construct was chemically synthesized as three gene fragments (gBlocks) of ~2000 bp to be assembled using the Gibson Assembly method. After assembly, the construct would be inserted into the AAVS1 safe harbor locus, a site widely used for gene knock-in with no adverse effects reported. In the course of this thesis, assembly of the gBlocks into a *FOXA1* construct could not be achieved, and this was probably caused by the overlapping regions needed for Gibson Assembly being too long, and by unspecific products in the PCR-amplified gBlocks inhibiting the reaction. The two custom guide RNA-sequences required for the CRISPR-mediated double strand break were designed, and one of them (gRNA2) was successfully cloned into a CRISPR delivery plasmid vector.

The remaining steps required to investigate the role of FOXA1 in EMT is to redesign and successfully assemble the *FOXA1* construct, clone gRNA1 into a CRISPR delivery vector, and co-transfect HBEC cells with the *FOXA1* construct and guide RNA-containing CRISPR vectors so *FOXA1* can be expressed in the cells.

Once established, the *in vitro* site-specific knock-in system will provide a versatile tool for future knock-in studies in HBEC cells and other human *in vitro* models.

Sammendrag

Lungekreft er hovedårsaken til kreftdødsfall hos menn over hele verden, og andre årsak til kreftdødsfall hos kvinner. Røyking av tobbakk er sterkt assosiert med lungekreft, da kreftfremkallende stoffer inhalert fra tobbaksrøyk forårsaker genetisk skade når de metaboliseres lungecellene. Epitelial-mesenchymal transisjon (EMT) er en prosess i kreftutvikling hvor cellene får en invasiv og migratorisk fenotype, og det medierer metastase av primærtumoren. EMT karakteriseres av spesifikke markører som uttrykkes fra de mesenchymale cellene. En mulig sammenheng mellom steroidreseptorsignalveier og karsinogenmetabolisme har blitt foreslått, og transkripsjonsfaktorer (FOXA-familien) som regulerer disse signalveiene kan derfor spille en viktig rolle i reguleringen av karsinogenmetabolisme. *FOXA1* og *FOXA2* er nedregulert i humane bronkiale epitelceller (HBEC) som er eksponerte for sigarettrøykkondensat. Da disse cellene også har fått en mesenchymal fenotype, støtter dette under oppfatningen om at FOXA-faktorer kan være involvert i EMT.

Ektopisk ekspresjon av *FOXA1* i de transformerte HBEC-cellene er en interessant måte å studere rollen til denne faktoren i EMT. I denne oppgaven var målene å

1) Utvikle en metode for stabil integrering av *FOXA1* i de transformerte HBEC-cellene

2) Oppnå nær normale nivåer av genuttrykk slik at effekten av FOXA1 i regulering av mesenchymale markører og -fenotype kunne studeres

For å få til dette brukte vi CRISPR-Cas9-systemet for stedsspesifikk genetisk endring, og designet et ekspresjonskonstrukt for integrasjon i målcellens genom i et genomisk såkalt ”trygg havn”-locus. CRISPR-Cas9-systemet lager et dobbelttrådet brudd på et ønsket locus, og bruddet kan repareres ved hjelp av homolog rekombinering med, i dette tilfellet, *FOXA1*-konstruktet som reparasjonstemplat. *FOXA1*-konstruktet ble kjemisk syntetisert som tre genfragmenter (gBlokker) ca 2000 bp i størrelse som skulle sammenkobles ved bruk av Gibson Assembly-kloning. Etter sammenkobling skulle konstruktet settes inn i den genomiske ”trygge havnen” AAVS1, som er et locus mye brukt for innsetting av DNA uten uønskede bieffekter. I løpet av denne oppgaven ble sammensetning av gBlokkene til et *FOXA1*-konstrukt ikke oppnådd, og dette skyldtes trolig at de overlappende regionene som behøves for Gibson Assembly var for lange, og at uspesifikke produkter dannet under PCR-amplifisering av gBlokkene hindret reaksjonen. De to spesialdesignede guide-RNA-sekvensene som var nødvendige for CRISPR-mediert dobbelttrådbrudd ble syntetisert, og en av dem (gRNA2) ble klonet inn i en CRISPR-plasmidvektor.

Trinnene som gjenstår for å kunne undersøke rollen til FOXA1 i EMT er å designe på nytt og sette sammen *FOXA1*-konstruktet, klone inn gRNA1 i et CRISPR-plasmid, og ko-transfektere HBEC-celler med *FOXA1*-konstruktet og CRISPR-plasmidene med de to guide-RNA sekvensene slik at *FOXA1* kan uttrykkes i cellene.

Straks metoden for stabil stedsspesifikk integrering av et gen er etablert, vil det være et allsidig verktøy som muliggjør mange fremtidige knock-in studier i HBEC-celler så vel som andre humane *in vitro* modeller.

Abbreviations

A	adenine
a.a.	amino acids
AAVS1	adeno-associated virus integration site 1
Amp	ampicillin
ATP	adenosine tri phosphate
B[*a*]P	benzo[*a*]pyrene
BLAST	basic local alignment search tool
bp	base pairs
C	cytosine
Cas9	CRISPR-associated protein 9
Cas9n	CRISPR-associated protein 9, nickase mutant
CDH	cadherins
CRISPR	clustered regularly interspaced short palindromic repeats
crRNA	CRISPR RNA
CSC	cigarette smoke condensate
CYP1A1	cytochrome P450, family 1, subfamily A, polypeptide 1
CYP1B1	cytochrome P450, family 1, subfamily B, polypeptide 1
DMSO	dimethyl sulfoxide
DNA	deoxyribonucleic acid
dNTP	deoxynucleoside triphosphate
DSB	double strand break
DTT	dithiothreitol
ECM	extracellular matrix
E. coli	*Eschericia coli*
EMT	epithelial-mesenchymal transition
FOXA1	forkhead box protein A1
G	guanine
gBlock	synthesized DNA fragment
GFP	green fluorescent protein
gRNA	guide RNA
HBEC	human bronchial epithelial cells
HDR	homology directed repair
HNH	domain in Cas9
HygroB	hygromycin B
IRES	internal ribosome entry site
Kan	kanamycin
kb	kilobases
LB	luria-bertani
MET	mesenchymal-epithelial transition
mRNA	messenger RNA
NEB	New England Biolabs
NHEJ	non-homologous end joining
NSCLC	non-small cell lung cancer
nt	nucleotides
PAH's	polycyclic aromatic hydrocarbons
PAM	protospacer adjacent motif
pA	polyadenylated tail
pAAVS1	pAAVS1 donor plasmid (Addgene plasmid # 22075)
PBS	phosphate buffered saline
PNK	polynucleotide kinase
poly-A tail	polyadenylated tail
pSpCas9n	pSpCas9n(BB)/PX460 (Addgene plasmid # 48873)
PCR	Polymerase chain reaction
PPP1R12C	protein phosphatase 1, regulatory unit 12C
RNA	ribonucleic acids
RT	room temperature
RuvC	domain in Cas9
S. pyogenes	*Streptococcus pyogenes*
sgRNA	single guide RNA
SpCas9n	S. pyogenes Cas9 nickase (D10A mutant)
SOC	super optimal broth
STAMI	National Institute of Occupational Health
T	thymine
T2A	*Thoseaasigna* virus 2A cleaving peptide
TAE	tris-acetate-EDTA buffer
tracrRNA	trans-activating CRISPR RNA
U	uracil

Table of Contents

1 Introduction

1.1 Lung cancer

Lung cancer is worldwide the leading cause of cancer deaths for men, and second leading cause for women. Tobacco smoke is the major cause of lung cancer development. Other risk factors include exposure to various occupational and environmental carcinogens such as asbestos, arsenic, radon, particulate air pollution and polycyclic aromatic hydrocarbons. [1] Although controversial, several lines of evidence point towards a difference in carcinogen susceptibility between men and women, and female smokers may be more prone to lung cancer than male smokers [2]. This may be at least partly due to differences in carcinogen metabolism [3].

Histologically, lung cancer can be divided in two main groups, small cell lung cancer (SCLC) and non-small cell lung cancer (NSCLC), where SCLC is the most aggressive type, and represent 15 % of all lung cancers. NSCLC can be divided further into large cell carcinoma, adenocarcinoma and large cell carcinoma, and they represent 85 % of lung cancer cases. All lung cancer types are strongly associated with smoking [1].

Eight physiological traits have been described as the hallmarks of cancer by Hanahan and Weinberg, and these hallmarks are considered to be common to virtually all cancers. In this description of carcinogenesis, cancer is seen as a cell-based genetic disease where DNA mutations lead to uncontrolled cell proliferation. Six traits were initially characterized as hallmarks of cancer in 2000, but the same authors have since added two additional traits, published in 2008. The eight hallmarks of cancer are: Self-sufficiency in growth signals, insensitivity to anti-growth signals, resistance to apoptosis, limitless replicative potential, sustained angiogenesis, activation of invasion and metastasis, genome instability and mutation, and tumor-promoting inflammation. Together these traits enable cells to avoid anti-cancer mechanisms, and allow them to proliferate and develop into a malignant tumor. [4]

1.1.1 Lung carcinogens

The lung is a target organ for all carcinogens inhaled from the environment. Cigarette smoke contains more than 5000 identified chemical compounds, and over 70 of these have been classified as carcinogens by the International Agency for Research on Cancer (IARC). The compounds are classified in three groups according to their carcinogenicity: Group 1 is carcinogenic to humans, group 2A is probably carcinogenic, and group 2B is possibly carcinogenic to humans. Polycyclic aromatic hydrocarbons (PAHs) and the tobacco smoke specific nitrosamines are considered of major importance. [5, 6]

Polycyclic aromatic hydrocarbons (PAHs) are a group of genotoxic compounds formed by incomplete combustion of organic matter. These carcinogens are found in e.g. tobacco smoke, broiled food and exhaust engines. PAHs are defined as chemicals combined by two or more benzene rings without heteroatoms. Because of its known carcinogenicity, benzo[*a*]pyrene (B[*a*]P) is a PAH that is used extensively in experimental studies, and is often utilized as a reference compound for carcinogenicity [7].

Nitrosamines are another type of carcinogens found in tobacco smoke. While PAHs are formed during combustion, the nitrosamines are formed during tobacco processing, and are

thus also present in unburned tobacco. Examples of nitrosamines are 4-(methylnitrosamino)-1-(3-pyridyl)-1butanone (NNK) and N-methyl-N-nitrosurea (MNU). [6]

1.1.2 Carcinogen metabolism

Many carcinogens only exert their carcinogenic effects after being metabolized or bioactivated by Phase I (functionalization) and phase II (conjugation) metabolism [7]. These reactions have the purpose of converting compounds to water soluble forms that can be excreted from the body, but intermediates produced in the process can cause DNA damage [5].

The compounds induce expression of Phase I enzymes such as cytochrome P450 1A1 (CYP1A1) and CYP1B1, and phase II enzymes such as epoxide hydrolases and glutathione *S*-transferases. The inducer (PAH) binds and activates the aryl hydrocarbon receptor (AHR), a transcription factor which in turn regulates the CYP enzymes. [7] Upregulated CYP enzymes then metabolize the carcinogens. Metabolites of the carcinogens may form DNA adducts, covalently bonding with DNA, and the adducts give rise to mutations [5]. In mice lacking the AHR receptors, carcinogenicity of exposure to B[*a*]P is reduced [8, 9].

B[*a*]P is capable of inducing a cancer-like phenotype in human lung cells though epithelial-mesenchymal transition (EMT) [10]. EMT is explained further in the following chapter.

1.2 Epithelial-mesenchymal transition (EMT)

1.2.1 Epithelial tissue

Epithelial cells are highly differentiated cells that cover skin and mucosa, and line the intestines and ducts of the body. There are many variants of epithelia, but forming a selective barrier between the environment and body is what they have in common. Different types of epithelia have secretory function, absorbtive properties and protect sterile environments of the body. The epithelial cells are connected to a basement membrane (basal lamina), from which they receive their nourishment, and they have a basal-apical orientation. The epithelial cells are held together with anchoring junctions that convey strength by connecting the cells to each other and to the extracellular matrix, and occluding junctions that form a tight seal at the apical part of the cells. Channel forming junctions are transmembrane channels that allow passage of small molecules and ions between the epithelial cells. [11]

The anchoring junctions consist of transmembrane anchoring proteins such as cadherins and integrins, which connect to the cytoskeleton at its cytoplasmic domain, and to other anchoring proteins or ECM proteins at its extracellular domain. The connection with the cytoskeleton is through a protein complex including proteins such as β-catenin and other catenins. Typically cadherins are part of cell-cell junctions, and integrins are part of cell-ECM anchoring junctions. The occluding junctions consist of transmembrane proteins called claudins and occludins, and claudins are the main functional proteins. In animal cells, the channel forming junctions (gap junctions) are made up of transmembrane proteins called connexins, which are assembled to form connexons (channels). [11] An overview of the various junctions that connect the epithelium is seen in **Figure 1.1.**

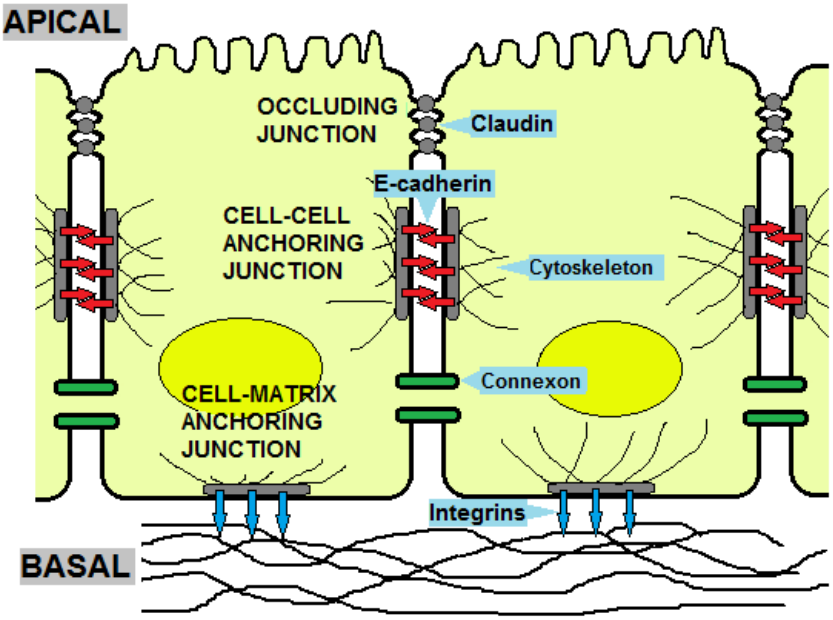


Figure 1.1 Overview of the various junctions connecting the epithelial cells. Occluding junctions are situated at the apical part of the epithelium, sealing the gap between the cells. Anchoring junctions connects the cells to each other and to the ECM by being connected to the cytoskeleton. Channel forming junctions links the cytoplasm of the cells, allowing molecules and ions to pass freely across the epithelium. Recreated from Alberts *et al*. [11]

1.2.2 Mechanisms of EMT

EMT is a process where epithelial cells alter their phenotype, and acquire mesenchymal characteristics. Changes in expression levels of cytoskeletal and cell adhesion proteins cause the cells to lose their basal-apical polarity, and increase their migratory and invasive abilities. EMT is an essential process during embryogenesis and organ development, as well as tissue regeneration following trauma/inflammation, but it is also associated with cancer progression and metastasis. [12]

Important pathways and factors mediating and regulating EMT have been identified (**Figure 1.2**), including TGF-β, EGF, FGF and the Wnt pathways, as well as transcriptional factors such as Snail, Twist, Slug and Zeb. Protein markers expressed by cells are frequently used to test for EMT. While markers such as E-cadherin, cytokeratins, claudin and occludin are downregulated in mesenchymal cells, N-cadherin, vimentin, desmin, brachyury, Slug, Snail and Twist are upregulated. [13] Having identified important initiating factors, “EMT master genes” such as *Snail* and *Twist*, it’s possible to induce EMT *in vitro* by ectopic expression of these genes [14].

1.2.3 Tumor metastasis mediated by EMT

Mesenchymal cells are typically seen at the invasive front of primary tumors. The cells invade through the basement membrane, and migrate through the circulatory systems in lymph or blood vessels. When the migratory cells colonize at their new location, they may regain the phenotype of the primary tumor, indicating that they undergo a mesenchymal-epithelial transition (MET) when settling at the new location (**Figure 1.3**). It is hypothesized that this reversal may be due to the absence of the signals and factors that caused the cells to undergo EMT at the site of the primary tumor. [12]

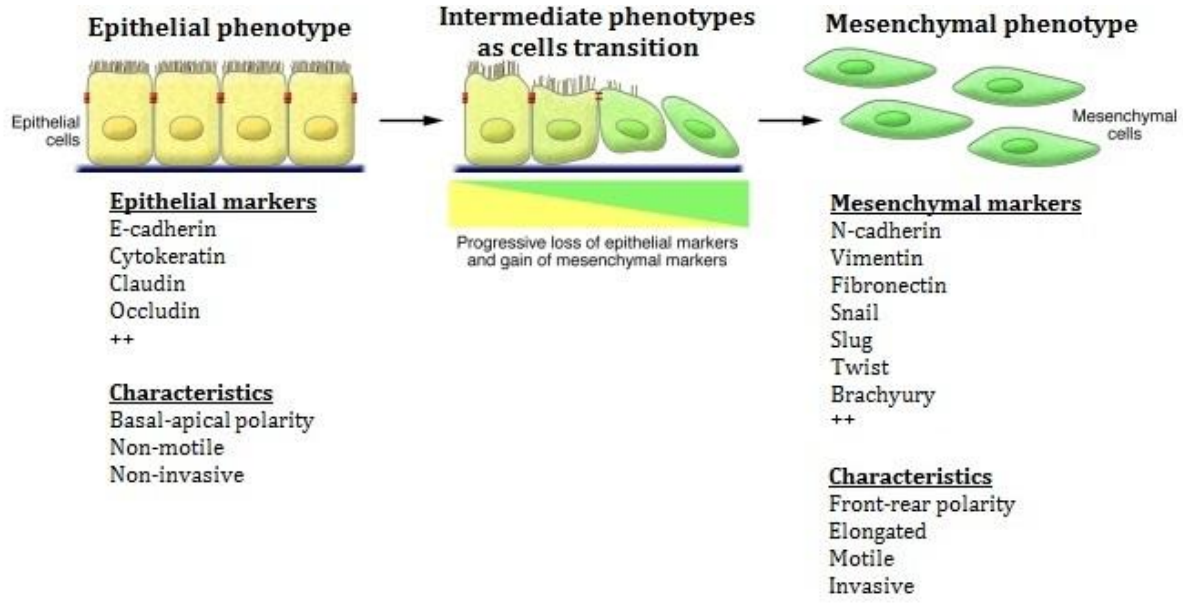


Figure 1.2 During EMT, the cells lose their basal-apical polarity, and gain motile and invasive properties. Some of the common epithelial and mesenchymal markers used for identification of EMT is listed in the figure. Modified from Kalluri *et al*. [12].

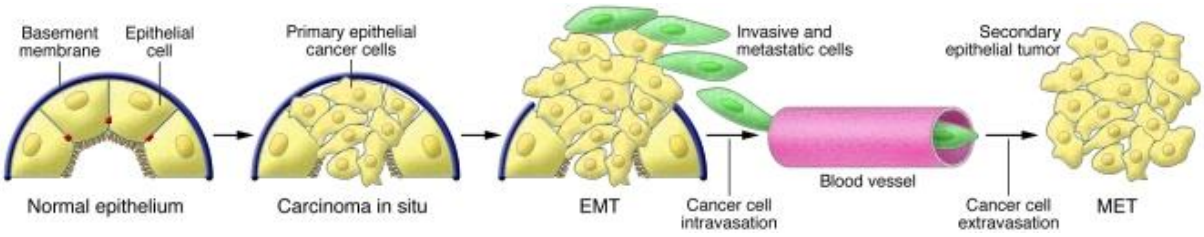


Figure 1.3 Metastasis of a malignant primary epithelial tumor. After having gone through EMT, the tumor cells may invade through the basement membrane and migrate via lymph or blood vessels. Once the mesenchymal cells reach a suitable destination, they may return to an epithelial phenotype through MET [12]

There is evidence that malignant cells may also metastasize by cohesive epithelial migration, and not MET exclusively [15]. Even though this indicates that EMT and subsequent MET is not the only mechanism of metastasis, several studies have shown a possible correlation between mesenchymal marker expression and unfavorable patient prognosis in multiple cancer types [13]. This may be due to the fact that cells that have undergone EMT acquire stem cell-like properties, which makes them more resistant to chemotherapeutic treatment [16]. While the majority of the tumor cells are treated efficiently, a small subpopulation of stem cell-like cells may remain and can form new tumors. Immunotherapeutic intervention targeting mesenchymal markers/drivers is a possible and promising strategy of removing these cells [17].

1.2.4 The cadherin switch

The cadherin superfamily, proteins of cell-cell anchoring junctions, consists of many tissue specific cadherins that bind to other cadherins of the same kind. This homophilic binding ensures that differentiated cells can form highly specialized tissues. The epithelium is held together by E-cadherin. The downregulation of the anchoring protein E-cadherin and balanced upregulation of N-cadherin observed in mesenchymal cells is named the “cadherin switch”, and is a hallmark of EMT. While E-cadherin expression acts as invasion suppressors in epithelium, N-cadherin expression increases motility and invasiveness of the cells. [18]

1.2.5 Role of forkhead box protein A1 (FOXA1) in EMT

The forkhead box A (*FOXA*) gene family are transcription factors that are involved with regulation of steroid receptor pathways. FOXA1 and FOXA2 binding sites are widespread in the mammalian genome, particularly in enhancer regions, and the factors are important during cell differentiation, embryogenesis and longevity. Binding of the FOXA factors to DNA gives an open chromatin structure. [19] Several studies have found that FOXA factors may be important regulators in cancer development, and that aberrant expression of *FOXA1* and *FOXA2* is associated with malignancy [20, 21]. FOXA2 knockdown has shown to induce expression of the EMT marker Slug in human lung cancer cells, which in turn downregulates E-cadherin [22]. Knockdown of either *FOXA1* or *FOXA2* induce EMT in pancreatic cancer cells, and a combined knockdown of both is an even stronger inducer of EMT. This suggests that these genes might cooperatively inhibit EMT. [20] Overexpression of *FOXA1* is also associated with malignancy, as *FOXA1* is amplified in lung-, esophageal- prostate-, thyroid- and estrogen receptor positive breast cancer [23].

Little is known about FOXA transcription factors role in lung cancer, but there are studies indicating that the estrogen receptor pathways regulated by FOXA factors may interact with carcinogen metabolism [24, 25]. This cross-talk may be partly the reason for sex differences in carcinogenicity of PAHs [26, 27]

1.3 *In vitro* transformation model of human bronchial epithelial cells (HBECs)

At STAMI, an *in vitro* transformation model has been established based on the work by Damiani *et al*. [28]. The cells used in this model (**Table 1.1**) are human bronchial epithelial cells (HBECs) that have been immortalized by retroviral transfection with cyclin dependant kinase 4 (Cdk-4) and human telomerase reverse transcriptase (hTERT). These genetic alterations make the cells continuously replicate while maintaining an epithelial phenotype, thus making them suitable for studies of lung cancer pathogenesis [29].

In the *in vitro* transformation model, the HBECs were exposed to carcinogenic compounds, including a cigarette smoke condensate (CSC), B[*a*]P, N-methyl-N-nitrosurea (MNU) and diesel exhaust particles for up to 26 weeks. This exposure has lead to transformed cells with cancer cell-like phenotype. Morphological changes as well as changes in transcriptional program (increase / decrease in expression of specific marker genes) indicate that these cells have gone through an EMT. [30, 31]

Table 1.1 Characteristics of the donor patients for HBECs used in the *in vitro* transformation model at STAMI. NSCLC - non small cell lung cancer

Cell Line	Tumor	Tumor Subtype	Age	Gender	Notes
HBEC2-KT	Lung	Immortalized Normal	68	M	NSCLC; informed consent
HBEC3-KT	Lung	Immortalized Normal	65	F	No cancer; informed consent
HBEC12-KT	Lung	Immortalized Normal	55	F	NSCLC

1.3.1 Downregulation of *FOXA1* in transformed HBECs showing a mesenchymal phenotype

Studies at STAMI have shown that long time exposure of HBECs to (CSC) causes morphological changes towards a mesenchymal phenotype, and the transformed cells show expression patterns characteristic to EMT. Two cell lines (HBEC2-KT and HBEC12-KT) were exposed to CSC for up to 15 weeks, cultured in soft agar for colony formation, and isolated. They were also subjected to an invasion assay, observing significantly increased capability of invasion by the transformed cells. Significant aberrations in the expression pattern of the *FOXA* genes were observed. After 12 weeks of exposure, the transformed cells showed significant downregulation of *FOXA1* and *FOXA2*. The experiments were performed by Audun Bersaas (research fellow at STAMI), and Mari Sjøberg (former master's student at STAMI). Results of *FOXA1* expression is displayed in **Figure 1.4**. [30, 31]

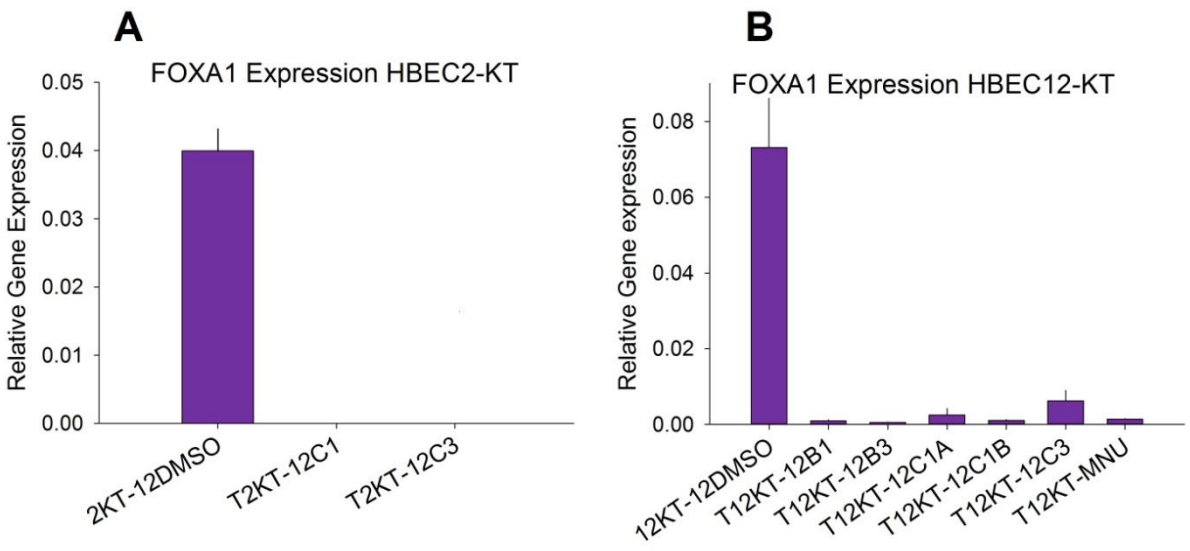


Figure 1.4 Expression levels of *FOXA1* in transformed cell lines after 12 weeks of exposure to CSC (1 µg/mL and 3 µg/mL), and B[*a*]P (0.3 µM and 1 µM). The cells were exposed to exposed DMSO as negative control, and MNU (0.5-1 µM) as positive control. A: CSC induced a transformed phenotype and significant downregulation of *FOXA1* in HBEC2-KT cells, but B[*a*]P and the positive control MNU failed to induce transformation in this cell line. B: CSC and B[*a*]P induced transformed phenotype and significant downregulation of *FOXA1* in HBEC12-KT cells. [31]

1.4 Site-specific transfection of *FOXA1*

The downregulation of *FOXA* genes in the exposure experiments performed at STAMI, and their unclear, but certain role in cancer development and EMT makes them interesting genes for further study. As *FOXA1* is downregulated in HBEC cells after long time exposure to CSC, and the cells have gone through an EMT, ectopic expression of *FOXA1* could elucidate the role of *FOXA1* in EMT. The goal of this project, explained in detail in chapter **1.8 –** Aim of study, was to restore an approximately normal expression of *FOXA1* in transformed HBEC cell lines. To achieve this, it was decided that the *FOXA1* gene must be inserted into a genomic safe harbor (GSH). A human GSH is defined as regions in the genome that can accommodate predictable expression of an integrated expression cassette without adverse effects in the host cell/organism; adverse effects being malignant transformation or change in any of the important functions of the cell [32]. The possible GSHs that are currently being used for knock-in experiments are AAVS1, CCR5 and human ROSA26. The AAVS1 site is a preferred target location for the adeno-associated virus (AAV), which is the only virus known to be capable of site-specific integration in the human genome [33]. The AAV virus uses specific Rep proteins to mediate its integration into the intron of the gene *PPP1R12C*. *CCR5* is a gene coding for the major co-receptor for HIV-1. This gene had been investigated as a possible GSH, but it's proximity to cancer related genes makes it a questionable choice for gene knock-in. A study also showed that GFP knock-in at this site gave lower expression levels than insertion at the AAVS1 site [34]. The human *ROSA26* gene is the human orthologue of mouse ROSA26, which has become a standard site for genome engineering in mouse embryonic stem cells. In humans, this site has yet to be investigated, but it unfortunately also has close proximity to cancer related genes. [35]

Raitano *et al*. (2015) did an insertion of progranulin (*GRN*) by homologous recombination into the AAVS1 site, restoring expression levels in cells with a null mutation of the *GRN* gene. The mutation causes frontotemporal dementia resulting from defects in neurogenesis. The results in the study showed that GRN transcript levels in the transfected cells did not differ significantly from the positive control cells, and that there were no significant acquired genetic abnormalities. [36] The study undoubtedly has similarities to this project, as the goal here is also to restore expression levels of a gene that has been downregulated in the target cells. Merling *et al*. obtained a 70-80 % constitutive expression of subunits from the phagocyte oxidase enzyme by inserting the subunit "minigenes" into the AAVS1 locus [37]. In a 2014 study, Tiyaboonchai *et al*. demonstrated not only exogenous gene expression from the AAVS1 locus, but also that short hairpin RNAs with a microRNA backbone expressed there could suppress expression of a secondary gene [38]. A different study found that expression form AAVS1 is stable, robust, and can be maintained for more than 18 weeks without selection [39].

1.5 Genome editing with CRISPR-Cas9

1.5.1 CRISPR in prokaryotes

Prokaryotes (bacteria and archaea) have evolved an adaptive immune system to protect themselves from repeated invasion of pathogens such as phages and plasmids. This immune system is mediated by clustered regularly spaced palindromic repeats (CRISPR) and CRISPR associated (Cas) enzymes. CRISPR-loci are arrays of approx. 20-50 bp palindromic repeats in the host genome, that are separated by spacers of about the same

length. The spacers are snippets of foreign DNA acquired from pathogens, which are expressed as RNA, and these RNA-sequences guide Cas nucleases to cleave foreign DNA upon future encounters. The sequence in the pathogen that is recognized by the spacer is called a protospacer, and must be flanked by a short motif, the protospacer adjacent motif (PAM) sequence. The PAM sequence is necessary for cleavage by Cas nucleases, but the exact sequence of the PAM differ in various bacterial and archaeal species. CRISPR loci are transcribed in a long transcript, and processed into shorter CRISPR-RNAs (crRNAs) containing a single palindromic repeat and the unique spacer. crRNA then associates with Cas enzymes and thus form a “surveillance complex” ready to recognize and cleave pathogens. The specific configuration of the surveillance complex varies, and therefore, the CRISPR systems can be divided in three types. In type II CRISPR systems (**Figure 1.5**), which are of relevance in this thesis, the crRNA forms an RNA duplex with trans-acting antisens RNA (tracrRNA). The tracrRNA contain a complementary sequence to the palindromic repeats, which enables the binding. [40]

1.5.2 Type II CRISPR-Cas9 system

The mechanisms of processing the CRISPR transcript into crRNAs vary for the different CRISPR systems. In many system types, specific Cas proteins function as RNases. In other systems, such as type II CRISPR-Cas9, tracrRNA directs crRNA processing by binding with an almost exact complementarity to the repeat sequences in the crRNA. The RNA duplex is recognized and diced by RNase III [42]. In 2011, Sapranauskas *et al*. discovered that Cas9 is the sole Cas protein required for DNA cleavage of the invading pathogen [43]. They also demonstrated that the type II CRISPR systems can be heterologously expressed, by successfully transferring a functional CRISPR-Cas system from *Streptococcus thermophilus* to the phylogenetically distant *Escherichia coli.* Cong *et al.* later found that RNase III did not have to be heterologously expressed for the type II CRISPR-Cas9 system to function in mammalian cells, suggesting that RNases expressed by the mammalian host assist in crRNA maturation [44]. Requiring only guide RNA and a single Cas protein for CRISPR interference, the type II system is thus well suited for genome engineering.

The PAM sequence is essential to the function of CRISPR interference as Cas9 ignores matching sequences if they don’t have the PAM sequence [45], and the invading pathogen can circumvent CRISPR interference if its PAM is mutated [46, 47]. The Cas9 from *Streptococcus pyogenes* recognizes a simple trinucleotide PAM sequence, NGG (**Figure 1.6).** This trinucleotide is very common in the human genome, occurring on average every 8-12 bp. The type II CRISPR-Cas9 system derived from *S. pyogenes* is therefore able to create a DSB virtually anywhere in the genome [48].

For increased target efficiency, crRNA:tracrRNA can be combined into a single molecule RNA chimera that forms the same base pairing interactions as the duplex RNAs. The chimera, named single guide RNA (sgRNA, **Figure 1.6**), has conserved the site-specific Cas9-guiding properties of the crRNA:tracrRNA duplex, but is more practical for heterologous expression and subsequent genome editing in a host cell. [46]

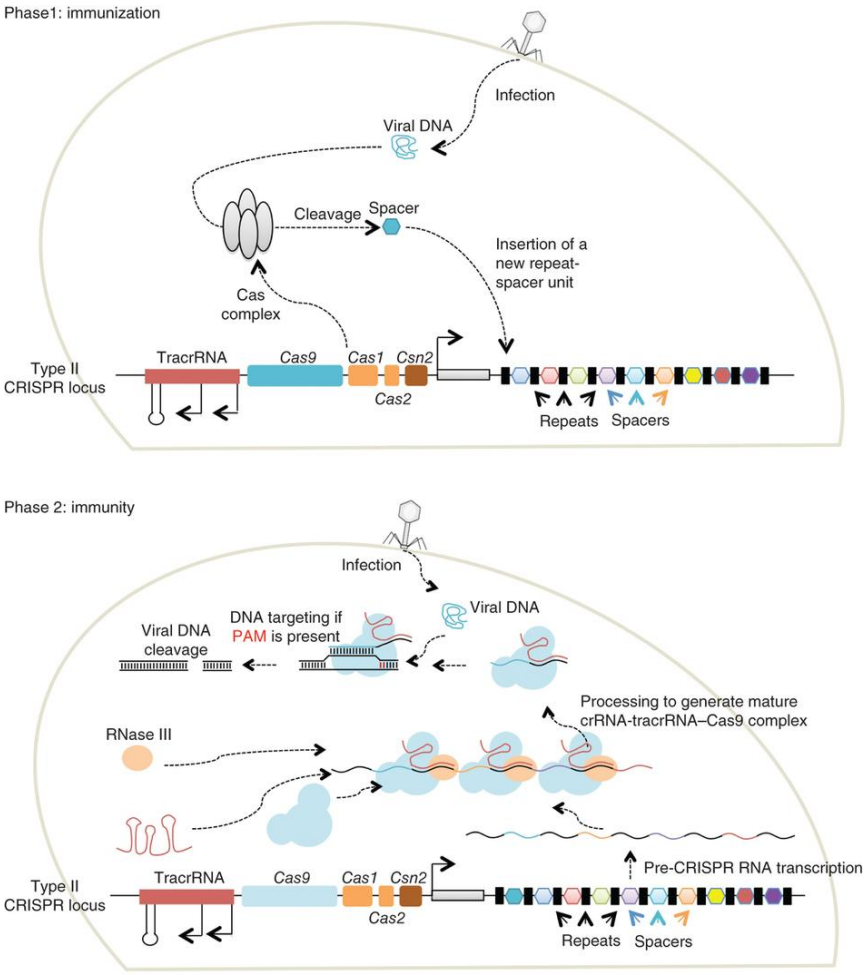


Figure 1.5 Type II CRISPR-Cas system. During first encounter with a pathogen, a short sequence (spacer) is inserted to the host genome between repeats in a CRISPR array. The host then transcribes the array, tracrRNA binds to the repeats, and RNase III processes the transcript into discrete crRNAs. Cas9 and the crRNA:tracrRNA duplex then form a complex that recognizes the complementary protospacer sequence in the invading DNA, and cleaves it upon next encounter [41].

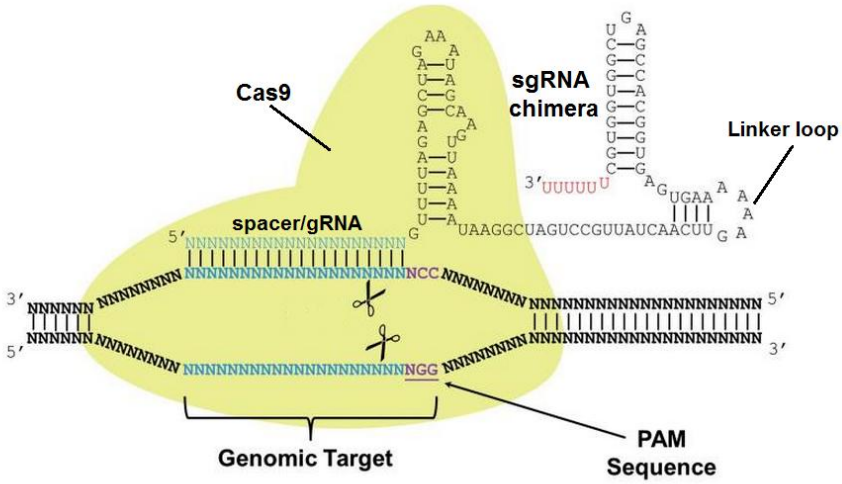


Figure 1.6 Type II CRISPR-Cas9 system derived from S. pyogenes with a chimeric sgRNA. The crRNA with a 20 nt spacer/gRNA sequence is fused with the tracrRNA, and the chimera binds to Cas9, mediating a site-

specific DSB at a genomic target. The non-complementary strand of the target has to be adjacent to a PAM. Protospacer shown in blue, spacer/gRNA in green and PAM in purple. Figure modified from DiCarlo *et al*. [49].

1.5.3 How to utilize CRISPR for genome editing

The DSB mediated by the CRISPR-Cas9 system is typically repaired by one of two major repair pathways, non-homologous end joining (NHEJ) and homology directed repair (HDR) (**Figure 1.7**). NHEJ is useful for knocking out a gene, as combining the ends often results in an insertion or deletion mutation, shifting the reading frame and creating premature stop codons. Larger sections of the DNA can also be removed by creating multiple DSBs. Introduction of an exogenous repair template can however activate the HDR pathway. For insertion of larger DNA fragments such as genes, the required repair template is a double stranded construct with homology arms of at least 500 bp flanking the genomic target, introduced in the cell in a plasmid vector. Another type of repair template is a single stranded, short DNA sequence with at least 40 bp homology arms flanking the genomic target, and this kind of repair template is useful for creating small edits to genomes, for example single nucleotide mutations. [48] The HDR pathway is limited to dividing cells, as it occurs mainly in the S-phase, when sister chromatids are normally available in the cell [50].

1.5.4 Cas9 nickase mutation for a double nicking strategy

One of the biggest drawbacks of the CRISPR-Cas9 system is unspecific binding of the spacers (gRNAs). Multiple mismatches between the gRNA and target protospacer can in some cases be tolerated by the CRISPR-Cas9 complex, creating off-target effects [51]. Studies have shown that increasing the length of the spacer sequence does not increase the specificity [52], but rather the opposite; shortening the typical 20 nt spacer by up to 3 nucleotides can decrease undesired mutagenesis at some off-target sites by 5,000-fold without compromising the on-target efficiency [53].

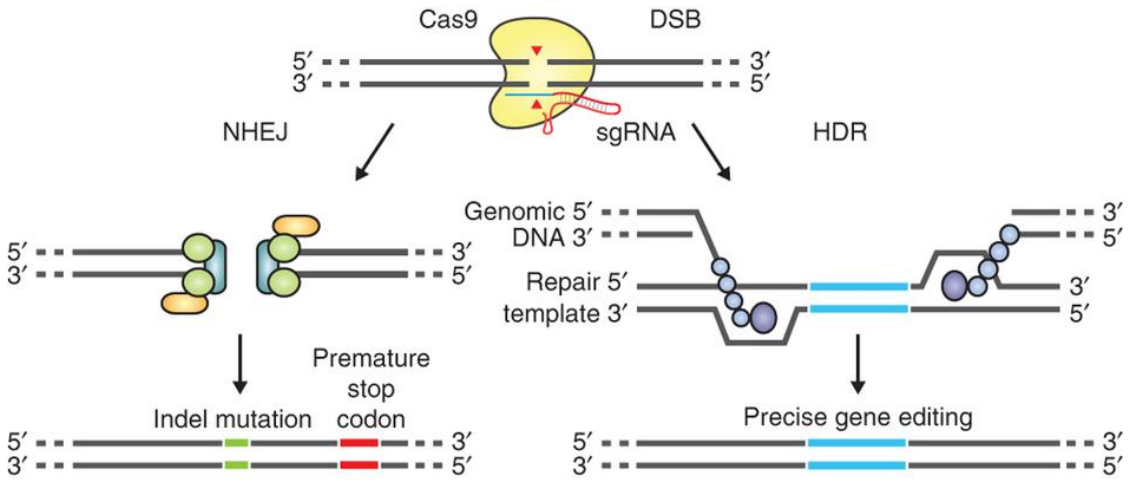


Figure 1.7 DSBs created by CRISPR-Cas9 system can be repaired by non-homologous end joining (NHEJ) or homology directed repair (HDR). NHEJ creates in/del mutations, and premature stop codons, thus knocking out the target gene. HDR enables repair using an exogenous repair template, and makes it possible to insert DNA or create specific scarless mutations. [48]

Attempts to elongate the complementary gRNA sequence have proven futile, as human cells process the longer gRNAs into 20 nt. In search for another way of increasing the number of complementary bases, and thus specificity, an alternative approach has been developed; double nicking with paired sgRNAs. In this strategy the Cas9 enzyme is mutated to create a single nick on one of the target DNA strands. [52] The Cas9 nuclease has two domains which each cleave one strand of the target DNA. The HNH nuclease domain cleaves the complementary strand of the target DNA, and the RuvC-like domain cleaves the non-complementary strand. A mutation in one of these domains can inactivate the catalytic seat and turn Cas9 to a nickase (Cas9n). [43, 46] When two sgRNAs recognize target sequences on each strand in close proximity, two nicks are caused, and a DSB with a 5' overhang is created (**Figure 1.8**). Using the Cas9 with a D10A mutation to the RuvC-like domain, design guidelines for position of the two sgRNAs have been developed. An offset between the two gRNAs should be as small as possible to increase the chance of a DSB occurring, and a -4 to 20 bp offset gives efficient DSBs. Increasing the offset beyond 20 bp also may create DSB, but with a reduced efficiency. The probability for a DSB to occur off-target is significantly reduced with double nicking, as two sgRNAs binding in sufficient proximity is highly unlikely. Single stranded nicks created by off-target binding by a single sgRNA are preferably repaired by the high-fidelity BER pathway, and should thus not result in mutagenesis. [52]

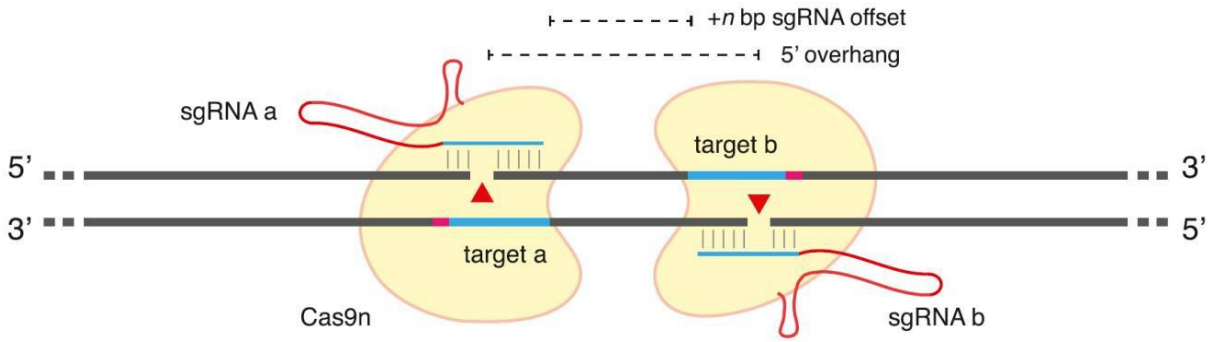


Figure 1.8 Two sgRNAs, one on each strand, bind with a -4 to 20 bp offset, causing Cas9n to nick at the target. If the two nicks are created at sufficient proximity, a DSB is created. [52]

1.5.5 Applications of CRISPR-Cas9

The CRISPR-Cas9 system is very useful in creating *in vitro* and *in vivo* cancer model systems, as it eliminates the need for time consuming and laborious cross breeding or germline/embryo manipulation. With CRISPR-Cas9, it's possible to induce loss-of-function or gain-of-function mutations simultaneously or sequentially, and recreate the specific combination of genetic aberrations for particular cancer phenotypes. [54, 55] Drost *et al*. were able to create useful colorectal cancer models by introducing oncogenic mutations in the four most commonly mutated genes in colorectal cancer (*KRAS*, *APC*, *P53* and *SMAD4*) in human cells, demonstrating that the cells developed cancerous characteristics [56]. Blasco *et al*. demonstrated that specific oncogenic chromosomal rearrangements could be modeled by generating double strand breaks on both the Alk and Eml4 genomic loci, creating the Elm4-Alk gene fusion often found in non-small-cell lung cancers, and thus inducing cancer formation in adult mice [57].

Genes essential for growth and viability can also be knocked down through the CRISPR-Cas9 system without cell death. Park *et al*. obtained complete knock-in of a DD tag on both alleles of a gene by simultaneously transfecting 293T cells with two repair templates containing the DD tag plus puromycin and blasticidine resistance genes respectively. The DD tag confers instability of the tagged protein which can be regulated by an external compound. In this way, the effects of knock-down can be rapidly induced, and also restored by removal of the external compound. [58]

Epigenetic studies is another area where CRISPR-Cas9 can be useful. A mutation in both catalytic domains of Cas9 creates a catalytically dead Cas9 (dCas9), which upon targeted binding can silence genes by blocking the transcription complex [59]. dCas9 can additionally be fused to effector proteins such as transcription activators or epigenetic modifiers. The epigenetic fusion proteins makes it possible to install or remove specific epigenetic marks (e.g. DNA methylation, histone acetylation), and perform targeted gene regulatory events without altering the DNA sequence [60].

In addition to creating disease models, CRISPR-Cas9 is a promising therapeutic tool for genetic disorders. Monogenic recessive disorders caused by loss-of-function mutations could potentially be treated by correcting the mutation with a small repair template, or inserting the whole functional gene into a safe harbor. Dominant disorders caused by a gain-of-function mutation could be treated inducing an allele-specific DSB (gRNA targeting the mutated sequence) and knocking the gene out with NHEJ-repair. Trinucleotide repeat disorders could be treated by inducing two DSBs, excising the excessive repeats, and either subjecting the DSBs to NHEJ-repair or introducing a repair template with the correct number of trinucleotide repeats for HDR repair [60].

Infectious diseases also have potential therapeutic approaches by CRISPR-Cas9. The human immunodeficiency virus (HIV) can be circumvented by targeting the CCR5 co-receptors on $CD4^+$ T-cells with NHEJ-mediated knock out. Tebas *et al*. showed that injecting patients with $CD4^+$ T-cells that were modified at the CCR5 locus decreased HIV DNA levels [61]. Sequences of an essential gene of a specific virus or bacteria may also be targeted for cleavage by Cas9, and even antibiotic resistance in bacteria may be battled by targeting resistance genes in bacteria [62].

1.6 Construct design for site-specific integration

For a gene to be heterologously expressed in mammalian cells, it must contain regulatory sequences that respond to the target cell's transcriptional/translational machinery. Promoters and enhancers are sequences where gene regulatory proteins can bind, and they are therefore capable of initiating transcription. Transcriptional terminators (poly-A tail) marks the end of the gene, and when transcribed they are recognized by proteins that release the mRNA from the transcription complex. It's important for translation that the construct contains a ribosomal binding site, and a start codon (ATG), that initiates translation and sets the reading frame for the following codons. The gene must also have a stop codon (TAG, TAA or TGA) that signals translation termination. [63]

1.7 Gibson Assembly® cloning

Traditionally, assembly of genetic elements have relied on restriction digestion and ligation reactions, and thus been limited by the presences of specific restriction sites in the DNA sequences. In 2009, Gibson *et al*. at the J. Craig Venter institute described a restriction enzyme-free method of combining DNA molecules up to several hundred kilobases. In this method, DNA molecules as large as ~500 kb were assembled *in vitro* in a one-step isothermal reaction [64]. It was a simplified version of a two-step enzymatic assembly method used to assemble the complete genome of *Mycoplasma genitalium* (582,970 bp) from 101 chemically synthesized oligonucleotides (~5000-7000 bp). The oligonucleotides contained overlapping sequences of 80-360 bp, and cassette 101 overlapped cassette 1, thus forming a complete circle. [65] The one-step enzymatic assembly method, named Gibson Assembly, builds on the same principle. The DNA fragments to be assembled need to have overlapping regions. 5' exonuclease removes nucleotides exposing a single strand, the overlapping regions anneal to each other, and DNA polymerase extends the 3' ends to fill in the single stranded gaps. When the DNA is polymerized, a ligase seals the nicks (**Figure 1.9**). It was demonstrated in the study that overlaps of 40 bp was sufficient for Gibson Assembly. [64] New England Biolabs (NEB) offers a ready-made Gibson Assembly kit, where all three enzymes are mixed in one buffer solution, which confers simple enzymatic assembly of DNA molecules of choice.

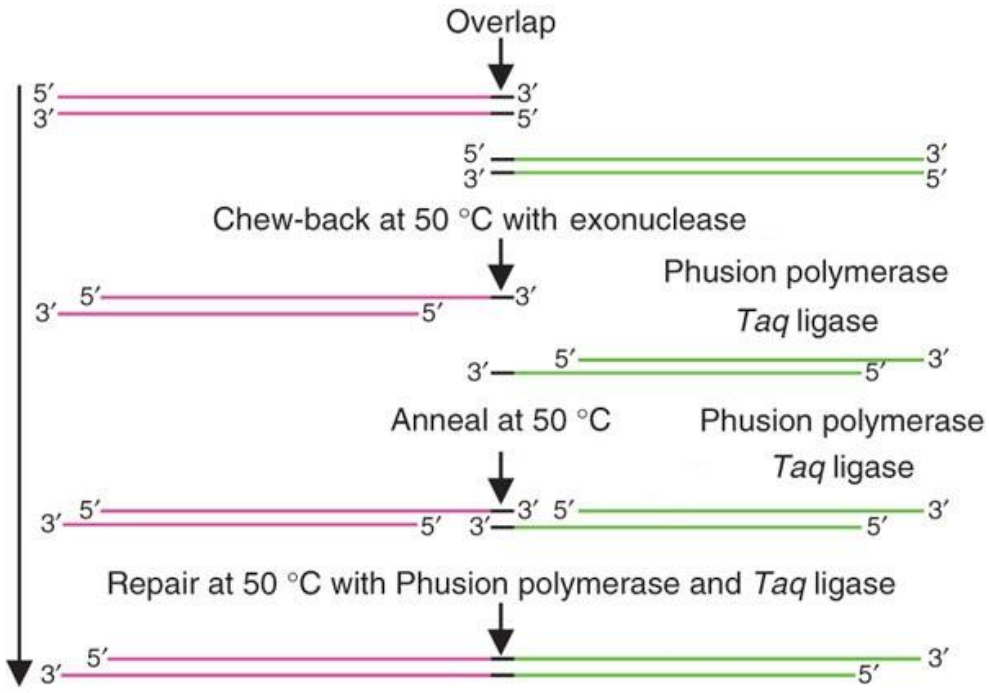


Figure 1.9 Gibson Assembly cloning: Two adjacent DNA fragments (magenta and green) shares terminal sequence overlaps (black). T5 exonuclease removes nucleotides from the 5' ends of double-stranded DNA molecules, complementary single-stranded DNA overhangs anneal, Phusion DNA polymerase extends the 3' ends and *Taq* DNA ligase seals the nicks. [64]

1.8 Aim of study

The aims of this project were divided in two parts

1) Development of a vector system for site-specific insertion of a gene of interest in HBEC cells

   This part of the aims for the project had the following specific goals:

   - Design an expression cassette containing *FOXA1* with selection markers/reporters for differentiation of transfected cells
   - Choose a suitable delivery system, and successfully clone the *FOXA1* construct into the delivery vector
   - Insert the construct containing *FOXA1* in a specific locus without adverse effects using the CRISPR-Cas9 system
   - Achieve close to normal expression of *FOXA1* in the transfected cells
   - Make the system versatile and open for editing, so the method can be applied for other genes of interest

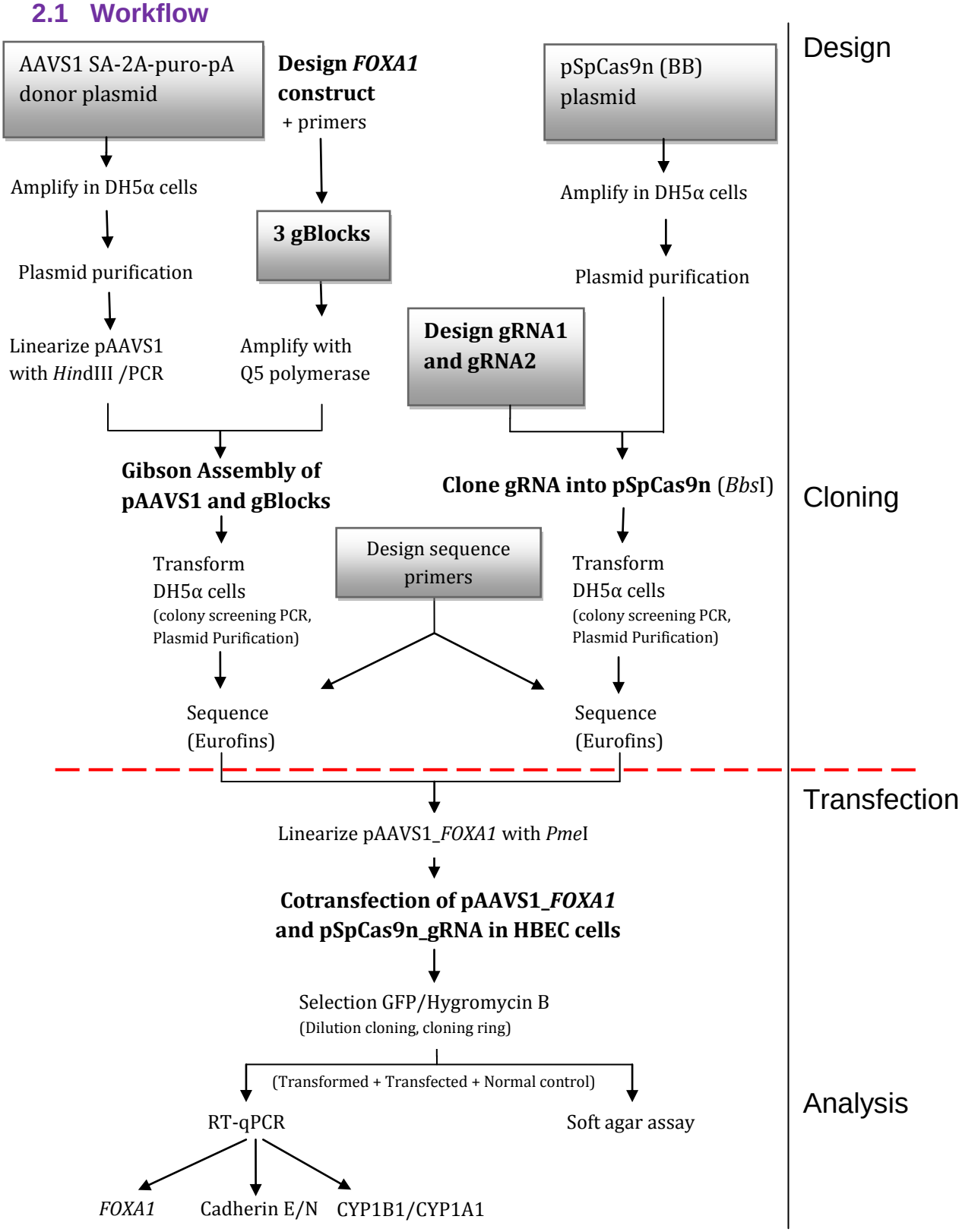
2) Determine role of the gene (*FOXA1)* in maintaining transformed phenotype of HBEC cells exposed to CSC

   The second part of the aims was about elucidating the role of *FOXA1* in EMT by observing potential changes in phenotype or EMT marker expression. The following hypotheses were proposed:

   - Ectopic expression of *FOXA1* from a safe harbor locus will result in close to normal expression levels of the *FOXA1* transcription factor
   - Normal expression levels of *FOXA1* will decrease expression of mesenchymal markers and increase expression of epithelial markers
   - Normal expression levels of *FOXA1* will decrease the number of colonies formed in soft agar
   - Normal expression levels of *FOXA1* will reduce the invasive properties of the cells

2 Materials and methods

2.1 Workflow



2.1: Workflow chart showing the steps of work in the whole project. The figure is divided in four main parts, in silico design of DNA sequences, cloning, transfection of HBEC cells, and analysis of the transfected cells. Steps above the red dotted line indicate the scope of this thesis.

2.2 Donor construct design

gBlocks™ are chemically synthesized, double stranded gene fragments that are compatible with use in methods such as gene assembly, genome editing and qPCR standards. gBlocks synthesized by IDT had a max size limit of 2000 bp, and accordingly was the ~4000 bp *FOXA1*-construct divided into three fragments with overlapping regions to be assembled with Gibson Assembly cloning. The synthesis method had certain restrictions to which sequences could be synthesized, and these criteria had to be considered when designing the construct. High GC contents and repetitive elements had to be limited, as they would be likely to give rise to 3D structures and difficulties in both synthesis and subsequent Gibson® assembly. The gBlock entry tool contained built-in quality control scans that predicted GC content and repetitive elements in both whole and parts of the gBlocks, and a sequence with predicted difficulties would not be accepted.

Codon optimization in genes with high GC-content and/or repetitive elements was conducted in IDT's Codon Optimization Tool. This software suggests optimal codons customized for either gBlocks or gene expression from a FASTA nucleotide sequence input. The desired organism of expression can also be specified. In this project the genes in the construct were optimized for gBlock synthesis.

GeneDesigner 2.0 bioinformatics software from DNA 2.0 is an intuitive software for design of de novo sequences [66]. The complete sequence of the *FOXA1*-construct was built in this software, adding the components (genes, promoters, terminators etc) sequentially and connecting them as discrete elements. The sequence of the pAAVS1_*FOXA1* assembled plasmid was also obtained through this software, by cloning the pAAVS1_*puro* vector and the construct together with the built in restriction cloning tool. This software was additionally used to generate the plasmid map of pAAVS1_*FOXA1* that can be seen in chapter **3.1.7**, **Figure 1.1**.

SerialCloner 2.6.1 is a software tool for analysis, manipulation and visualization of nucleotide sequences. The software was mainly used for marking features of the construct and plasmid sequences, including gBlocks, overlapping regions, genes and homology arms. Many features, such as genes and common components of plasmid vectors are already specified in the software, but these may also be added manually and highlighted in the desired colors. SerialCloner 2.6.1 was also used for identifying restriction sites. The sequence could be viewed in addition to the FASTA-format, as linear or circular graphic maps showing the relative positions of all specified features and restriction sites, or sequence maps displaying both the sense and antisense strand with corresponding amino acid sequences for all three possible reading frames. Small changes to the sequences, such as mutation of PAM sequence, were done by directly changing the nucleotides in the SerialCloner interface.

Sequences were aligned in the Jalview workbench software using muscle nucleotide alignment with preset parameters. The Muscle algorithm allows for alignment of large number of sequences simultaneously. [67, 68] The alignment was used to compare gBlocks to the final construct and ensure that the sequences matched, to quality check correct positions of small edits conducted on the sequences, and also to visualize the overlapping regions between the gBlocks.

Nucleotide sequences were generally imported as FASTA format, and viewed in jEdit programmer's text editor, as this text editor assigns no formatting to the text.

2.3 gRNA design

gRNA design was performed using the Zhang Lab CRISPR design tool from MIT [69]. The tool is developed to simplify the selection of CRISPR gRNA in an input sequence. It uses an algorithm to identify suitable gRNAs adjacent to a PAM, searches for off-target effects in a selected genome and scores the sequences by their target specificity.

Artemis (Sanger Institute), a genome browser and annotation tool, was used to find a suitable region for gRNA oligo search within the intron of the PPP1R12C gene.

Verification of the results was performed in the CRISPR/Cas9 gRNA design tool from DNA 2.0, which uses their own algorithm to search for suitable gRNAs in an input sequence, a gene or a genomic region, and also ranks the hits by specificity and predicts off-target effects in a selected genome [70].

The predicted sequences of pSpCas9n_gRNAs were cloned *in silico* using the Serial Cloner software.

2.4 Primer design

Primers used in this project were designed in IDT's software, PrimerQuest [71]. The software enables entry of a specific sequence or a GenBank/Accession ID, and scans the sequence for PCR/sequencing primers matching the selected parameters. The software provides detailed custom design parameters including amplicon location and size, desired Tm, GC content and primer size, as well as reaction conditions.

NCBIs tool Primer-BLAST (basic local alignment search tool) was used to scan relevant genomes for unspecific binding by the primers [72]. In addition to having many of the same design features as PrimerQuest, this tool can scan entire genomes from various databases, or a custom sequence. Primers can be suggested by the software according to design parameters, or manual query for unspecific binding by existing primers.

2.5 Plasmid vectors

2.5.1 Donor Plasmid

AAVS1 SA-2A-*puro*-pA donor (pAAVS1_*puro*) was a gift from Rudolf Jaenisch (Addgene plasmid # 22075) [73]. The plasmid contains AAVS1 3' and 5' homology arms (800 bp each) flanking a donor cassette consisting of a splice acceptor, a 2A cleaving peptide and a puromycin resistance (*PuroR*) gene. Two *Hin*dIII restriction sites are situated between each homology arm and the donor cassette. This enables removal of the cassette by *Hin*dIII restriction digestion, and subsequent insert of a different donor cassette. The plasmid also contains *AmpR* and *KanR* genes, which enables selective culturing with ampicillin and kanamycin. Plasmid map is shown in **Figure 2.2**.

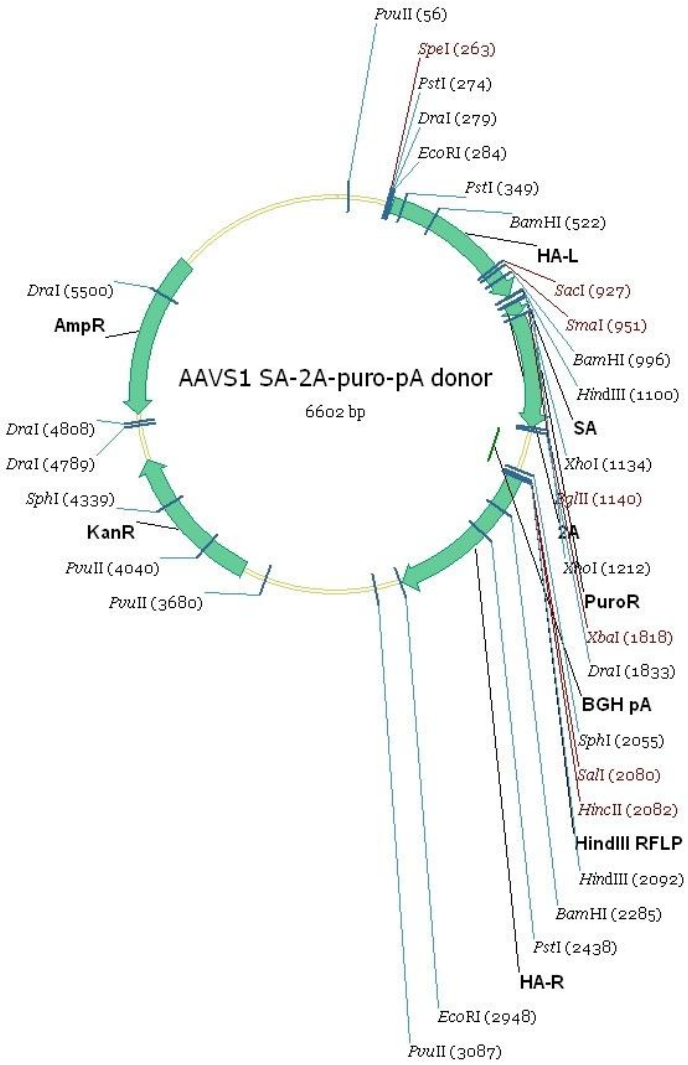


Figure 2.2 Plasmid map showing components of the pAAVS1_*puro* plasmid. The map was uploaded to AddGene by the depositor.

2.5.2 CRISPR plasmid

pSpCas9n(BB)/PX460 (pSpCas9n) was a gift from Feng Zhang (Addgene plasmid # 48873) [48]. The plasmid contains the *Cas9n* (D10A nickase mutant) gene from *S. pyogenes* in an expression cassette with a CBh promoter, a bGH pA terminator and an SV40 NLS. The plasmid also contains an sgRNA scaffold which upon insertion of a 20 bp spacer (gRNA) can be expressed as a functional sgRNA by a U6 promoter. The gRNA can be inserted with restriction ligation cloning using the type IIs restriction enzyme *Bbs*I. Ampicillin can be used to enrich for the plasmid, as it contains an *AmpR* gene. Plasmid map is shown in **Figure 2.3**.

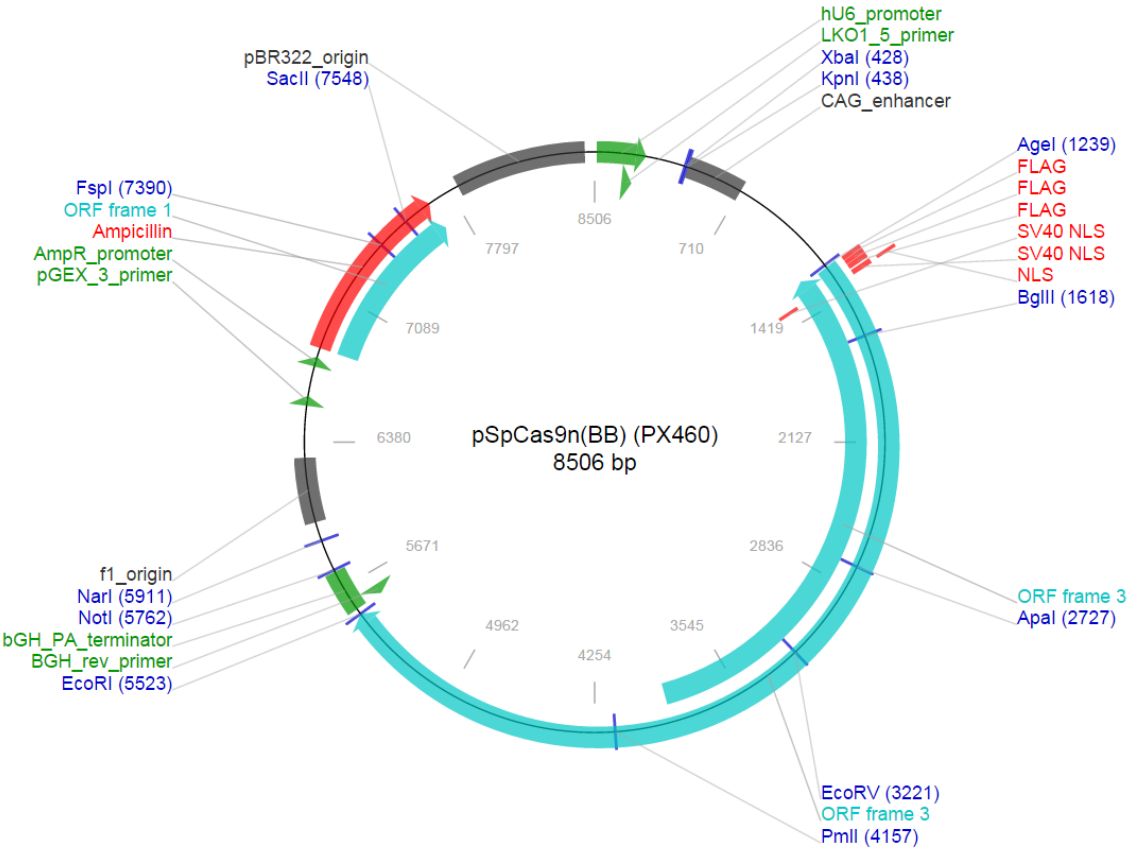


Figure 2.3 Plasmid map showing components of the pSpCas9n plasmid. The map is generated by AddGene from full sequence supplied by depositor

2.5.3 Amplification of plasmids

All bacterial work was performed in an OAS LAF bench; media and equipment were sterilized in a Systec DX90 autoclave. Powdered LB medium and agar were weighted out on a PG503 DeltaRange® scale, dissolved, and autoclaved at 121 °C. The bacteria containing pAAVS1_*puro* were cultured in LB medium containing ampicillin (50µg/mL) and kanamycin (50µg/mL), and bacteria containing pSpCas9n were cultured in LB medium containing ampicillin (50µg/mL).

2.5.3.1 Streaking bacteria to obtain single colonies

Bacterial stab cultures containing the plasmids were received from AddGene, and were streaked out on LB agar plates using sterile plastic inoculation loops (Nunc). Plates were incubated at 37 °C for 18 hours in a Termax B8000 bacteriologic chamber.

2.5.3.2 Inoculating single colonies in liquid cultures

Liquid cultures were incubated in an Infors HT Ecotron shaker incubator at 37 °C. Single colonies were picked from the agar plates with autoclaved pipette tips (Biosphere®) and inoculated for eight hours in starter cultures with liquid LB medium (25 mL). A serial dilution of the suspension into secondary cultures was performed, ranging from 1:1000 to 1:10,000. The secondary cultures were inoculated in 250 mL Erlenmeyer flasks with liquid LB media (100 mL), and incubated for 18 hours. An OD_{600} measurement was conducted on the

cultures to ensure that the cell density was optimal (OD_{600} = 2-3), and determine the amount of bacterial suspension to harvest for plasmid purification.

2.5.4 Glycerol stocks

Glycerol freeze stocks were made for long term storage of the plasmids in bacteria. This was to be able to make more plasmid DNA without transforming new competent cells. Frozen bacterial glycerol stocks are stable for several years at -80 °C, but repeated thawing may reduce viability.

1. An overnight culture of the bacteria containing the desired plasmid was inoculated in liquid LB medium
2. 500 µL of the overnight culture was added to 500 µL glycerol (50 %) in a 2 mL screw top, and gently mixed
3. The vials were stored at -80 °C
4. To recover bacteria from stock, the top layer of frozen bacteria was scraped off with a sterile loop while keeping the vials on ice. The bacteria were then streaked onto a fresh LB agar plate

2.6 Amplification of gBlocks™ with Q5® High Fidelity polymerase

After *in silico* design of gBlock1, gBlock2 and gBlock3, the fragments were synthesized by IDT, and sequence verified before being delivered as dried pellets. The gBlocks are delivered linear, and not cloned, which means that a small percentage of the DNA molecules may contain sequence errors. IDT estimates that the gBlock fragments have 85-90 % purity.

Upon arrival, the gBlocks were resuspended in 20 µL TE buffer for a DNA concentration of 10 ng/µL. Q5 High Fidelity polymerase from New England Biolabs (NEB) was used to amplify the gBlocks before Gibson assembly, and reaction mix is seen in **Table 2.1**.

This polymerase has an error rate >100-fold lower than *Taq* polymerase (1/3,300 bp for *Taq* vs 1/1,000,000 bp for Q5), and can amplify approximately 1 kb in 10 seconds. Both its proof reading capability and speed makes it ideal for amplification of large, complex sequences. Reaction conditions were optimized for all three gBlocks, and details are presented in **Appendix E**.

Table 2.1 PCR reaction mix for amplification of gBlocks (*optimal primer concentration varied between PCR assays)

Component	Volume	Final conc/mass
Q5 Reaction Buffer (5X)	10 µL	1X
dNTPs (10 mM)	1 µL	200 µM
Primer Fw (10 µM)	2.5 µL	0,17-0.5 µM*
Primer Rev (10 µM)	2.5 µL	0,17-0.5 µM*
gBlocks (0.1 ng/µL)	1.0-1.5 µL	100-150 pg
Q5 High Fidelity DNA Polymerase	0.5 µL	20 U/mL
Nuclease free H_2O	32.5 µL	
Total volume	50 µL	

The reaction was run on an MJ Research PTC-220 DYAD Thermal Cycler with the program shown in **Table 2.2**.

Table 2.2 PCR program for amplification of gBlocks (*annealing temperature was optimized for each set of primers, and extension time depended on the length of the amplicons)

Step	Temp	Time
1. Initial denaturation:	98 °C	30 s
2. Denaturation	98 °C	8 s
3. Annealing	58-66 °C*	20 s
4. Extension	72 °C	10-20 s*
5. Repeat step 2-4 for 25 cycles		
6. Final extension	72 °C	2 min

2.7 Enzymatic assembly of gBlocks™ and pAAVS1 with Gibson® assembly cloning

2.7.1 Linearization of pAAVS1 with *Hin*dIII-HF

The pAAVS1_*puro* plasmid vector contains two HindIII-sites; one located just downstream from the 5' homology arm, and one just upstream from the 3' homology arm. Digestion with this enzyme was therefore used to cut out the *PuroR* expression cassette, enabling ligation of the *FOXA1* expression cassette in its place. Digestion mix in **Table 2.3**.

Table 2.3 Restriction digest mix for cleavage of pAAVS1_*puro* plasmid with HindIII-HF

Component	Volume	Final conc/mass
CutSmart Buffer (10X)	30.0 µL	1X
pAAVS1_*puro*	15.96 µL	600ng
HindIII-HF	6.0 µL	400 U/mL
Nuclease free H_2O	248.04 µL	
Total volume	300 µL	

The reaction mix was incubated at 37 °C for 30 minutes, and then at 80 °C for 20 minutes to inactivate the enzyme. In some of the digests, SDS was added to the mix after inactivation to dissolve enzymes still bound to the DNA substrate. SDS (0.1-0.5 %) was then added to the loading buffer, and the samples were incubated with loading buffer for 10 minutes at 65 °C before loading samples on agarose gel.

2.7.2 Linearization of pAAVS1 with PCR

Another method of removing the *PuroR* expression cassette from pAAVS1_*puro* was by PCR amplification using primers annealing just outside the homology arms. The forward primer was designed to anneal just upstream from the 3' homology arm, and the reverse primer just downstream from the 5' homology arm. Amplification was conducted using Q5® High Fidelity polymerase (**Table 2.4**), and the PCR amplicon was a 5569 bp linear pAAVS1 sequence without the *PuroR* expression cassette.

Table 2.4 PCR reaction mix for linearization of pAAVS1

Component	Volume	Final conc/mass
Q5 Reaction Buffer (5X)	10 µL	1X
10 mM dNTPs	1 µL	200 µM
Primer Fw_pAAVS1_linearize (10 µM)	2.5 µL	0.5 µM
Primer Rev_pAAVS1_linearize (10 µM)	2.5 µL	0.5 µM
pAAVS1_*puro* (5 ng/µL)	0.5 µL	2.5 ng
Q5 High Fidelity DNA Polymerase	0.5 µL	20 U/mL
Nuclease free H_2O	33 µL	
Total volume	50 µL	

The reaction was run on an MJ Research PTC-220 DYAD Thermal Cycler with the program shown in **Table 2.5**.

Table 2.5 PCR program for linearization of pAAVS1

Step	Temp	Time
1. Initial denaturation:	98 °C	30 s
2. Denaturation	98 °C	8 s
3. Annealing	63 °C	20 s
4. Extension	72 °C	1 min
5. Repeat step 2-4 for 25 cycles		
6. Final extension	72 °C	2.20 min

2.7.3 Gibson assembly® reaction

The Gibson Assembly Cloning Kit (NEB) includes a master mix that contains all three enzymes required for the assembly in a single buffer. The 5' exonuclease creates 3' overhangs, enabling the overlapping regions of the gBlocks to anneal. DNA polymerase fills in the gaps, and the DNA ligase seals the nicks for a scarless, circular product. The entire reaction is run simultaneously at isothermal conditions (50 °C), and the reaction time can be extended based on the length of the overlapping regions.

The three gBlock fragments were incubated with the plasmid vector linearized either with a HindIII restriction enzyme or by PCR. Both long and short overlapping regions were used for assembly (**3.4.1** Primer design for amplification of gBlocks). Small fragments are advised to be added in large molar excess. The protocol provided by NEB recommended a minimum of 2:1 molar excess of the inserts, but this could be increased up to 10:1 if the insert was very small compared to the vector. The fragments in these assembly reactions were quite large compared to the vector, so molar excesses of 2:1 and 3:1 were used.

The following formulas were used to calculate the volume of gBlocks and vector to be added to the reaction mix:

$$\text{input mass dsDNA (g)} = \text{moles of dsDNA} \cdot ((\text{size of dsDNA (bp)} \cdot 617.96\ \text{g/mol}) + 36.04\ \text{g/mol})$$

$$\text{input volume (µL)} = \frac{\text{input mass DNA (ng)}}{\text{conc DNA (ng/µL)}}$$

Addition of DMSO may be advantageous to reduce formation of 3D structures [74], and some assembly reaction parallels therefore had DMSO (5 %) added to attempt an increase of efficiency. Three different reaction times were tested; one, two and four hours. The reaction conditions of the assemblies are shown in **Table 2.6** and **Table 2.7**.

Table 2.6 Gibson reaction mix for assembly of the pAAVS1 vector and gBlocks with long overlapping regions. Reaction mix was incubated for one hour

Component	Volume	Final conc/mass
Gibson Assembly Master Mix (2x)	10 µL	1X
pAAVS1 (linearized w/*Hin*dIII)	2,03 µL	0.2 pmol
gBlock1	3,37 µL	0.4 pmol
gBlock2	1,56 µL	0.4 pmol
gBlock3	2,34 µL	0.4 pmol
Nuclease free H_2O	0,70 µL	
Total volume	20 µL	

Table 2.7 Gibson reaction mix for assembly of the pAAVS1 vector and gBlock2 and gBlock3 with short overlapping regions. Reaction mix was incubated for two and four hours

Component	Volume	Final conc/mass
Gibson Assembly Master Mix (2x)	5 µL	1X
pAAVS1 (linearized w/PCR)	1.01 µL	0.2 pmol
gBlock2	1.47 µL	0.6 pmol
gBlock3	0.79 µL	0.6 pmol
Nuclease free H_2O	1.73 µL	
Total volume	10 µL	

All Gibson reaction mixes were incubated on an MJ Research PTC-220 DYAD Thermal Cycler at 50 °C, and subsequently stored at -20 °C or kept on ice until further use. The new name given to the assembled plasmid was pAAVS1_*FOXA1*_*eGFP*_T2A_*HygroB* (pAAVS1_*FOXA1*)

2.8 Insertion of gRNAs in pSpCas9n with *Bbs*I

The gRNA oligos were inserted into the pSpCas9n plasmid vector by restriction-ligation cloning. Single stranded oligos had to be phosphorylated and annealed. Then they were incubated with the vector backbone in a cycle program that cuts the plasmid and ligates the oligos with plasmid in cycles. The new ligated plasmid products were given the names pSpCas9n_gRNA1 and pSpCas9n_gRNA2, dependent on which oligo was ligated into the vector.

2.8.1 Phosphorylation and annealing of ssoligos

Ligation of single stranded oligonucleotides into double stranded gRNAs (gRNA1 and gRNA2) was carried out in a phosphorylation/annealing reaction. A polynucleotide kinase (T4 PNK) attaches a phosphate group to the 5' end of the oligos, and denaturing followed by a slow decline in temperature cause the oligo pairs to anneal and form the double stranded gRNA spacers with phosphorylated ends to be inserted into the pSpCas9n vector. gRNA1 is the spacer that will bind the forward strand of the CRISPR target, and gRNA2 is the spacer that will bind the reverse strand

Table 2.8 Reaction mix for phosphorylation and annealing of single stranded gRNA oligos

Component	Volume	Final conc/mass
Oligo1	1 µL	100 µM
Oligo2	1 µL	100 µM
T4 PNK reaction buffer (10X)	1µL	1X
T4 PNK (NEB)	0.5 µL	500 U/mL
Nuclease free H_2O	6.5 µL	
Total volume	10 µL	

Table 2.9 Thermocycler program for phosphorylation and annealing of single stranded gRNA oligos

Step	Temp	Time
1. Phosphorylation	37 °C	30 min
2. Denaturing	95 °C	5 min
3. Ramp down to 25 °C at 0.1 °C/s	25 °C	

2.8.2 Digestion-ligation

The annealed gRNAs were diluted 250-fold before proceeding to digest the pSpCas9n vector with *Bbs*I, and ligate the gRNAs into the vector with T7 DNA ligase. The cleavage and ligation steps were repeated sequentially six times to increase cloning efficiency.

Table 2.10 Reaction mix for combined digestion-ligation of pSpCas9n and gRNAs

Component	Volume	Final conc/mass
pSpCas9n	1 µL	100 ng
gRNA	2 µL	~0.8 µM
T7 ligase reaction buffer (2X, NEB)	10 µL	1X
*Bbs*I (NEB)	1 µL	0.5 U/µL
T7 DNA ligase (NEB)	0.5 µL	75000 U/mL
Nuclease free H_2O	5.5 µL	
Total volume	20 µL	

Table 2.11 Thermocycler program for combined digestion-ligation of pSpCas9n and gRNAs

Step	Temp	Time
1. Restriction digest (*Bbs*I)	37 °C	5 min
2. Ligation (T7 DNA ligase)	23 °C	5 min
3. Repeat step 2-4 for 6 cycles		
4. Keep cooled until further use	4 °C	For ever

2.8.3 Exonuclease digestion to prevent unwanted recombination products

An optional digestion with T5 exonuclease was conducted on the reaction mix to degrade unwanted linear DNA. The exonuclease removes nucleotides from the 5' end of linear DNA, or at gaps or nicks in circular DNA, thus degrading sequences that had been digested but not properly ligated.

Table 2.12 Exonuclease degradation of unwanted DNA from the pSpCas9n_gRNAs cloning reaction

Component	Volume	Final conc/mass
Digestion-ligation reaction	11 µL	~55 ng
NEBuffer 4 (10X)	1.5 µL	1X
T5 Exonuclease (NEB)	1 µL	667 U/mL
Nuclease free H_2O	1.5 µL	
Total volume	15 µL	

The mix was incubated for 30 min at 37 °C, and kept at 4 °C until further use

2.9 Bacterial transformation

Bacterial transformation is a process where bacteria take up exogenous DNA from its surroundings. In the laboratory, bacteria must be competent for transformation to occur, and this can be achieved either chemically or by electrophoration. These treatments alter the cell wall/membrane to make it permeable for DNA. Competent cells are incubated with plasmid DNA on ice before they undergo a heat shock, and this process causes uptake and replication of plasmids. The transformed bacteria are often outgrown for a short time in rich media without antibiotics to allow expression of the antibiotic resistance genes in the plasmid. They are then spread on antibiotic containing LB agar plates for selection of single colonies. Chemically competent 5α *E. coli* cells (NEB) were transformed with the cloned plasmids pSpCas9n_gRNA1 and pSpCas9n_gRNA2.

2.9.1 Transformation protocol

1. Competent cells were thawed on ice for 10 minutes
2. Chilled plasmid DNA (2 µL) was carefully added to the cells while stirring with the pipette tip, and gently mixed by flicking on tubes 4-5 times
3. The mixture was placed on ice for 30 minutes without mixing
4. It was then heat shocked at 42 °C for 30 seconds without mixing
5. The tubes were transferred to ice for 2 minutes
6. Then room-temperature SOC media (950 µL) was added to the tubes
7. The cultures were outgrown by incubation at 37 °C for 30 minutes while being vigorously shaken at 300 rpm
8. Several 1:10 serial dilutions were made, and 100 µL of the cell suspensions were spread onto pre-warmed LB-agar plates
9. The plates were incubated for 16-18 hours at 37 °C

2.9.2 Inoculation of transformed bacteria in liquid LB medium

Autoclaved filter tips from Biosphere® were used to pick colonies. Starter cultures were inoculated in liquid LB media (1mL) with ampicillin (100 µg/mL) in an Eppendorf Thermomixer comfort heat block (37 °C, 300 rpm), and secondary cultures were inoculated in 250 mL Erlenmeyer flasks with liquid LB medium (100 mL) ampicillin (100 µg/mL) and incubated in a Infors HT Ecotron shaker incubator (37 °C, 250 rpm).

2.10 Plasmid Purification with GeneJET Plasmid Midiprep Kit

The GeneJet MidiPrep plasmid purification kit (Thermo Scientific™) is a silica spin-column based plasmid purification kit designed for large scale plasmid isolation from recombinant *E. coli* cultures. The bacteria are lysed by an SDS/alkaline buffer solution to release their DNA, and a neutralization buffer allows plasmids to reanneal while chromosomal DNA and cell debris is precipitated out of solution. The precipitate is pelleted, and supernatant is loaded onto the silica columns. The high salt concentration of the supernatant creates appropriate conditions for plasmid adsorption to the silica column, followed by washing of the columns to remove contaminants. Plasmids are eluted from the columns with Tris-HCl elution buffer. Centrifugation of the tubes was performed in a Sigma 4K15 refrigerated centrifuge. Harvesting of cells by centrifugation was carried out at 4 °C, and the remaining steps at RT.

2.10.1 OD_{600} measurement with Biophotometer

To estimate the amount of bacterial culture to be harvested for plasmid purification, the turbidity (optical density) of the bacterial cultures was measured. In this measurement, the spectrophotometer is set to pass light of 600 nm wavelength through a sample cuvette, and the bacteria will scatter the light. Optical density (OD) measurement of light scattering is proportional to the concentration of bacterial cells. Generally, a 50 mL overnight culture with OD_{600} = 2-3 is optimal for high yield of plasmid DNA from the purification kit. Max culture volume should not be exceeded, as it lowers the quality of the DNA. A formula for calculation of max culture volume was provided with the purification kit:

$$\textit{Max culture volume} = \frac{150}{OD_{600}}$$

Fresh LB medium was used as blanking reagent.

2.10.2 MidiPrep plasmid purification protocol using low speed centrifuge

1. Bacteria were harvested by centrifugation for 10 min at 5000 g, and supernatant was discarded by decanting
2. Cell pellet was resuspended in Resuspension Solution (2 mL) with RNase A by pipetting and vortexing
3. Lysis solution (2 mL) was added, and the tubes were mixed by inverting 4-6 times. Tubes were incubated for exactly 3 minutes at RT
4. Neutralization solution (2 mL) was added, and tubes were mixed by inverting 5-8 times
5. Endotoxin binding reagent (0.5 mL) was added, and tubes were mixed by inverting 5-8 times gently but thoroughly
6. Then 96 % ethanol (3 mL) was added, and the tubes were inverted 5-6 times
7. The tubes were centrifuged for 40 min at 5000 g to pellet cell debris and chromosomal DNA
8. The supernatant was carefully transferred by decanting to a 15 mL tube without disturbing the precipitate, and 96 % ethanol (3 mL) was added immediately and mixed by inverting
9. Part of the sample (approx. 5.5 mL) was transferred to a silica column pre-assembled in a 15 mL collection tube, and centrifuged for 3 min at 2000 g. Flow through was discarded and the column placed back into the tube
10. Step 9 was repeated with the remaining lysate
11. Wash solution I (4 mL) with added isopropanol was added to the column, which was centrifuged for 2 min at 3000 g. Flow through was discarded, and the column was placed back into the tube
12. Wash solution II (4 mL) with added ethanol was added to the column, which was centrifuged for 2 min at 3000 g. Flow through was discarded, and the column was placed back into the tube
13. Step 12 was repeated
14. Residual wash solution was removed by centrifuging the tube for 5 min at 3000 g
15. The column was transferred to a fresh 15 mL collection tube, and plasmid DNA was eluted by adding Elution Buffer (0.5 mL) to the column and incubating for 2 min at RT. The tubes were centrifuged for 5 min at 3000 g, and the column was discarded

2.11 PCR cleanup with Silica bead DNA Gel Extraction Kit

The silica bead DNA gel extraction kit (Thermo Scientific™) is a quick and simple DNA extraction system for purification from PCR reaction mixes or agarose gels. DNA is adsorbed to suspended silica powder (beads) in high salt concentration, and washing of the beads remove unwanted components from the mix. Finally, the DNA is eluted into water or TE-buffer. The kit has an up to 80 % DNA recovery. The PCR products in this project were purified directly from reaction mix to minimize UV DNA damage during gel excision, which means that unspecific PCR products were not removed by the purification process.

2.11.1 DNA extraction protocol from PCR reaction mix using a table top centrifuge

1. A 3:1 volume of Binding Buffer was added to the reaction mixture followed by thorough mixing
2. The appropriate amount of Silica Powder Suspension was added to the mixture. For <2.5 µg of DNA, 5µL was added. For >2.5 µg, 2 µL per µg of DNA was added. The mixture was incubated for 5 min at 55°C, and mixed by vortexing every few minutes
3. The silica powder/DNA complexes were spun down for 1 min at 13000 rpm, and the supernatant was carefully removed and discarded with a pipette
4. Ice cold Washing Buffer (500 µL) with ethanol was added, and the pellet was resuspended by mechanically stirring with the pipette tip, and gentle pipetting. The tubes were spun for 1 minute at 13000 rpm. The procedure was repeated two times. After the supernatant from the last wash had been removed, the tubes were spun again and remaining liquid was removed with a pipette. Samples were air dried for 10-15 minutes
5. The pellet was resuspended in TE buffer (0.1 M EDTA) of the same volume as the silica powder suspension in step 2. The tubes were incubated for 5 min at 55 °C
6. The sample was spun for 1 min at 13000 rpm, and supernatant was carefully transferred to a new tube. Step 5-6 could be repeated optionally for an overall higher yield of DNA, but with a lower concentration
7. For removal of residual silica powder, the eluate was spun for 1 min at 13000 rpm, and the supernatant transferred to a fresh tube

2.12 DNA quantification and purity assessment with Nanodrop

Plasmids and PCR products were quantified on an eight channel NanoDrop 8000l (Thermo Scientific). The NanoDrop is a spectrophotometer that can measure concentration and purity of nucleic acids, peptides, proteins and other compounds in a single drop (0.5-2.0 µL) of sample with no dilution required even for highly concentrated samples. The sample is loaded onto an optical measurement surface, and light of a specific wavelength is passed through it to obtain an absorbance reading that correlates with sample concentration. The Beer-Lamberts law relates the amount of absorbed light to the concentration of DNA, and an average extinction coefficient for DNA is used to calculate the concentration. An absorbance of 1 is equivalent to 50 ng/µL DNA, and this method of calculation is considered valid up to Absorbance = 2.

DNA shows absorbance maximum at 260 nm wavelength, while proteins mainly absorb light at 280 nm. Phenols have an absorbance maximum at 270 nm, while carbohydrates have a maximum at 230 nm. By calculating the ratio between these measurements, it’s possible to

estimate the sample's purity. A 260/280 ratio of ~1.8 is generally accepted as pure for DNA, and lower ratios indicate contamination by proteins. A 260/230 ratio below 2.0 indicate contamination by sugars, salts or organic compounds.

2.13 PCR with *Taq* polymerase

Amplification with *Taq* DNA Polymerase Master Mix Kit (VWR) was conducted after transformation of NEB 5-alpha Competent *E. coli* cells with plasmid vectors pAAVS1_*FOXA1* and pSpCas9n_gRNA1 and 2. This was to identify bacterial cultures that had taken up the correct vector before larger scale culturing and plasmid isolation.

Taq polymerase was also used to test new primers in various assays, and the reaction setup in these PCR assays was equal except that DNA (100 pg) was used as template instead of 1µL bacterial culture.

Table 2.13 PCR reaction mix for colony screening

Component	Volume	Final conc/mass
Taq Master Mix	18 µL	1X
Primer Fw	1 µL	0.4 µM
Primer Rev	1µL	0.4 µM
Bacterial culture	1 µL	
Total volume	21	

The reaction was run on an MJ Research PTC-220 DYAD Thermal Cycler with the program shown in **Table 2.14**.

Table 2.14 PCR program for colony screening (*annealing temperature was optimized for each set of primers, and extension time depended on the length of the amplicons)

Step	Temp	Time
1. Initial denaturation:	95 °C	5 min
2. Denaturation	95 °C	15 s
3. Annealing	50-54 °C*	30 s
4. Extension	72 °C	0:20-5:45 min*
5. Repeat step 2-4 for 26 cycles		
6. Final extension	72 °C	5-7 min

2.14 Agarose gel electrophoresis

All PCR products and isolated plasmids were run on Lonza SeaKem® GTG® Agarose gel in TAE buffer (**Appendix A**) to check for correct size. PCR products from amplification of gBlocks, colony screening of Gibson assembly and isolated supercoiled plasmids (794-8506 bp) were run on 1% agarose, and the electrophoresis was carried out at 85 V for 60 minutes. PCR products from colony screening of pSpCas9n_gRNA plasmids inserts (271-341 bp) were run on 2% agarose, and the electrophoresis was carried out at 150 V for 30 min.

GelRed (1.5 µL, Biotium) was added to the gel for UV visualization of electrophoresis bands. Several of the samples were diluted in MQ-water before loading to ensure that the DNA mass was appropriate (20-80 ng). 2 µL loading dye was added for every 5µL PCR product.

Electrophoresis bands of expected size between 1000-10,000 bp were size verified using a 1kb ladder (NEB), while bands of expected size between 100-1000 bp were size verified using a 100bp ladder (NEB). Plasmids purified from bacterial cultures were size verified using a supercoiled ladder (NEB) for plasmids ranging between 2000-10,000 bp.

2.15 Sequencing

The pSpCas9n_gRNA1 and pSpCas9n_gRNA2 plasmids were sent to Eurofins Genomics (Ebersberg, Germany) for value read sequencing. Plasmid and sequencing primer was diluted in nuclease free H_20 and mixed in 1.5 mL safe lock tubes according to the sample submission guide provided by Eurofins (**Table 2.15**). [75]

Table 2.15 Premixed sample for sequencing of pSpCas9n_gRNAs

Component	Volume
pSpCas9n_gRNA (75 ng/µL)	15 µL
Sequencing Primer Fw (10 µM)	2 µL
Total volume	17 µL

2.15.1 Design of sequencing primers

Primers for sequencing of pAAVS1_*FOXA1* and pSpCas9n_gRNAs were designed with IDT's PrimerQuest software, according to the criteria for optimum sequencing primer conditions, provided by Eurofins. As the sequencing was to be conducted on purified plasmid DNA, alignment search with 5α *E. coli* chromosomal DNA was unnecessary. Primer sequences are listed in **Appendix B**.

Optimum Primer Conditions [75]:

- Primers must not contain phosphorylation or fluorescent dyes
- The optimum primer length is between 16-25 bases
- The primer melting temperature (Tm) should be 50 - 62°C
- The GC content of the primer should be 35-60%
- Ideally one G or C should be located at the 3' primer end
- The number of Gs or Cs in the 3' end of the sequencing primer should not exceed 2
- If possible, avoid >3 identical bases in a row in the sequence

2.15.2 Sequence reads

Sequence reads were received from Eurofins Genomics in FASTA format, and sequence alignment with the predicted plasmid sequences was interpreted with Jalview Muscle nucleotide alignment software with preset parameters. Sequence reads are presented in **Appendix D**.

3 Results

3.1 *FOXA1* repair template construct design

At the start of this project, the commercially available systems for knock-in by CRISPR-Cas9 were examined. Several companies are offering CRISPR-Cas9 systems for different purposes (**3.2.1** - Choice of CRISPR vector), but the only available complete knock-in kit was Genome-CRISP^{TM} human AAVS1 safe harbor gene knock-in kit by GeneCopoeia. A gene of interest from their ORF knock-in clone library could be delivered in a complete repair template construct. The kit came with a CRISPR-Cas9 expression plasmid, a donor plasmid with the repair template construct flanked by AAVS1 homology arms, and primers required for selection of correct knock-in. The GeneCopoeia repair construct, with gene of interest already cloned in, consists of two expression cassettes. A CMV promoter and a bGHpA terminator are used for expression of the gene of interest. The reporter gene *GFP* and puromycin resistance (*PuroR*) selection gene are both expressed in one transcript by an Ef1a promoter and terminated by a SV40pA. Upon translation, a T2A peptide bridge cleaves GFP and PuroR into functional proteins.

This kit would be suitable for *FOXA1* knock-in, were it not for the PuroR selection marker. The HBECs in the *in vitro* transformation model already possess resistance to puromycin due to the immortalization process, when they were transfected by retroviral constructs containing *PuroR* [29]. The construct offered in GeneCopoeia was nevertheless a useful template for creating a knock-in repair construct.

3.1.1 Choice of promoters

With a goal of stable expression of the *FOXA1* gene in HBEC cells, as well as strong expression of the reporter/selection genes, literature searches were performed to identify suitable promoters for the construct. Quin *et al*. (2010) compared six common constitutive promoters (SV40, CMV, UBC, Ef1a, PGK and CAGG) for expression levels in a variety of different mammalian cell types. The study found that UBC and PGK were consistently weak in all the cell types, and Ef1A and CAGG was consistently strong in all cell types. SV40 had mid to high stable expression levels, while CMV was the most variable promoter, being very strong in certain cell types and weak in others. [76]

Based on this study, the CAGG and Ef1a promoters would be the best options, especially for the reporter/selection genes. Unfortunately, synthesis of gBlocks and subsequent Gibson assembly held restrictions to the sequence. These promoters were large and contained repetitive elements that were not compatible with gBlock synthesis. The choice eventually landed on CMV as promoter for the *FOXA1* cassette, and the stable mid to high expression level SV40 promoter for the reporter/selection genes.

3.1.2 Choice of reporter/selection markers

The green fluorescent protein (GFP) emits bright green fluorescence (510 nm) when it's exited by UV light (395 nm). This makes it a suitable reporter for knock-in genes, as transfected cells can easily be distinguished from other cells in a fluorescence microscope. An enhanced version of the GFP (eGFP) was chosen as a reporter in this project as it fluoresces with an approximate 100-fold higher intensity than the wild type GFP. [77] For an easier clonal isolation and visual quantification of single clones, a nuclear localization signal (NLS) was attached to the 3' end of *eGFP*, which would ideally result in green fluorescence of the nuclei of transfected cells. The gene was codon optimized for gBlock synthesis.

An antibiotic resistance gene was also important to include in the construct, as a fluorescence cell sorter was not available for this project. Hygromycin B is an antibiotic that inhibits protein synthesis in prokaryotes and eukaryotes, and the *HygroB* resistance gene is widely used for selective culturing [78]. *HygroB* was therefore included in the *FOXA1* construct.

3.1.3 Choice of separator for bicistronic reporter/selection marker cassette

Several strategies exist for multicistronic expression. When expressing multiple genes in one cassette, double or bidirectional promoters is an option, or internal ribosomal entry sites (IRES) situated between the genes. The drawbacks to these strategies is unreliable results; for example when using IRES-coexpression, where the first gene in the cassette is expressed at a much higher level than the second [79]. A more promising method is the 2A or 2A-like peptide bridges, short sequences in picornaviruses that separate viral proteins expressed by the same promoter by cleaving them into discrete proteins during expression. A highly conserved 18 a.a. sequence motif is common to these peptides. [80] A study by Kim *et al*. compared cleavage efficiency of four different 2A-peptides (P2A, T2A, E2A and F2A), and found that the P2A petide derived from porcine teschovirus-1 had the highest efficiency in human cells, but that T2A from *Thosea asigna* virus was also efficient [81]. A different study by Tang *et al*. has shown that T2A gives efficient and reliable coexpression of two or more genes from a single promoter [80].

In this project, the T2A peptide was chosen to separate eGFP and HygroB.

3.1.4 Choice of safe harbor

As explained in the introduction, a genomic safe harbor (GSH) is an insertion site that confers stable and non-malignant knock-in of an expression cassette. A promising candidate matching these criteria is AAVS1, with several studies having shown that heterologous expression at these sites is non-pathogenic to the target cell (described in chapter **1.4**). Based on these studies, the AAVS1 locus was determined as the best site for *FOXA1* insertion.

3.1.5 gBlock™ DNA fragments

The construct was to be synthesized as gBlock fragments, and as the maximum size of each gBlock was 2000 bp, thus the ~4000 bp construct was split into three suitable fragments with overlapping regions to be assembled with Gibson assembly cloning. The fragments were named gBlock1 (1888 bp), gBlock2 (872 bp) and gBlock3 (1766 bp). **Figure 3.2** shows which parts of the construct the respective gBlocks contain, and also a schematic of the Gibson assembly with the pAAVS1 plasmid vector. A plasmid map is shown in **Figure 3.3**. Length of

gBlocks, and design of primers are described in chapter **3.4.1** Primer design for amplification of gBlocks.

3.1.6 Choice of donor plasmid

For site-specific insertion into the AAVS1 locus, a donor plasmid containing homology arms was needed. pAAVS1_*puro* plasmid (AddGene ID #22075) was chosen because it contained 800 bp AAVS1 5' and 3' homology arms, and its 992 bp *PuroR* expression cassette could easily be removed by single digestion with *Hin*dIII.

3.1.7 The final construct

After extensive literature searches and database/software analyses, a final design of the construct to be assembled was made. This construct has the features of a codon optimized *FOXA1* gene within a CMV + bGHpA cassette, and a codon optimized *eGFP* reporter gene with a nuclear localization signal and *HygroB* resistance gene within a SV40 + SV40pA cassette, separated by a T2A cleaving peptide (**Figure 3.1**). The construct was to be cloned into the pAAVS1 vector so that it would be flanked by two 800 bp homology arms. The sequences within the homology arms matching the target protospacers and PAM were mutated (single nucleotide substitution) to avoid the repair template being cleaved by Cas9 upon transfection. Sequences used in the construct, where they were obtained, and which modifications have been done to them is listed in **Appendix C**.

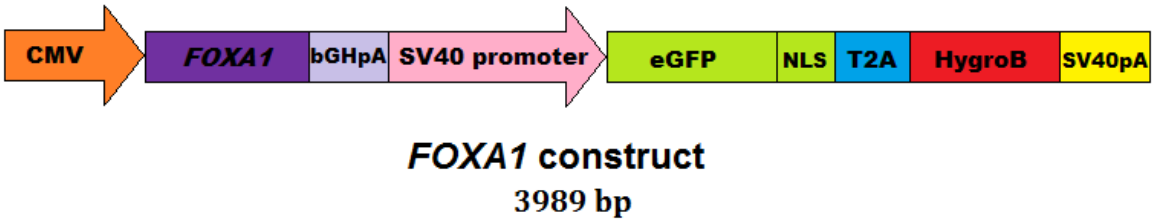


Figure 3.1 Schematic presentation of components in the *FOXA1* construct cassette. The construct is a polycistronic expression cassette that contains a codon optimized *FOXA1*-gene, a codon optimized eGFP and Hygromycin B selection marker. *FOXA1* has a constitutive mid level expression promoter and polyA-tail. eGFP and HygroB shares a mid to high level expression promoter and polyA tail, are therefore separated by a T2A cleaving peptide which splits the peptide sequence after translation. eGFP has an added nuclear localization signal at its 3' end, which will visualize the nuclei of the transfected cells.

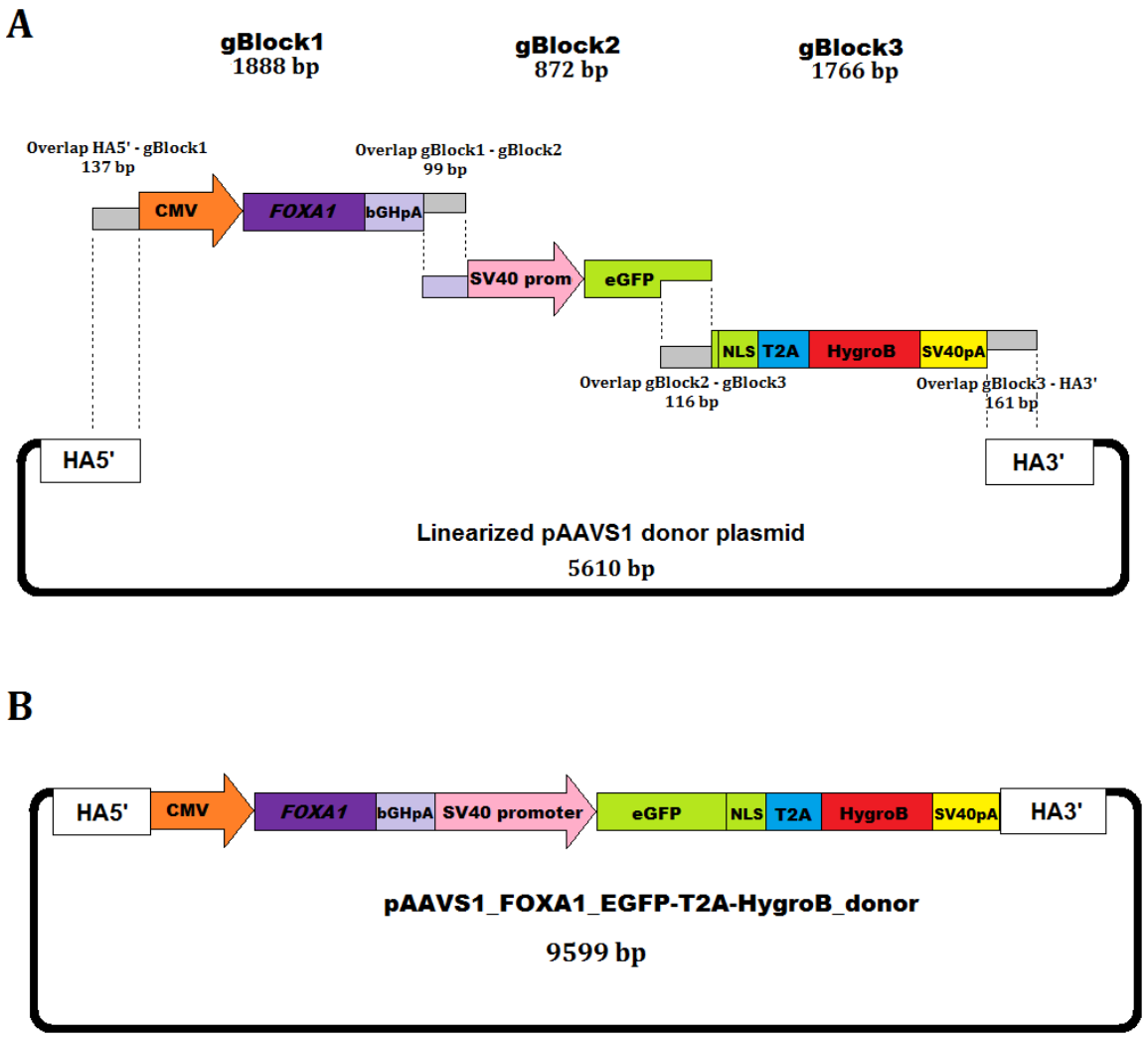


Figure 3.2 A: Schematic of the pAAVS1 vector and the *FOXA1* construct divided into three gBlocks with overlapping regions between them that facilitate Gibson® assembly. B: The complete assembled pAAVS1_*FOXA1* plasmid

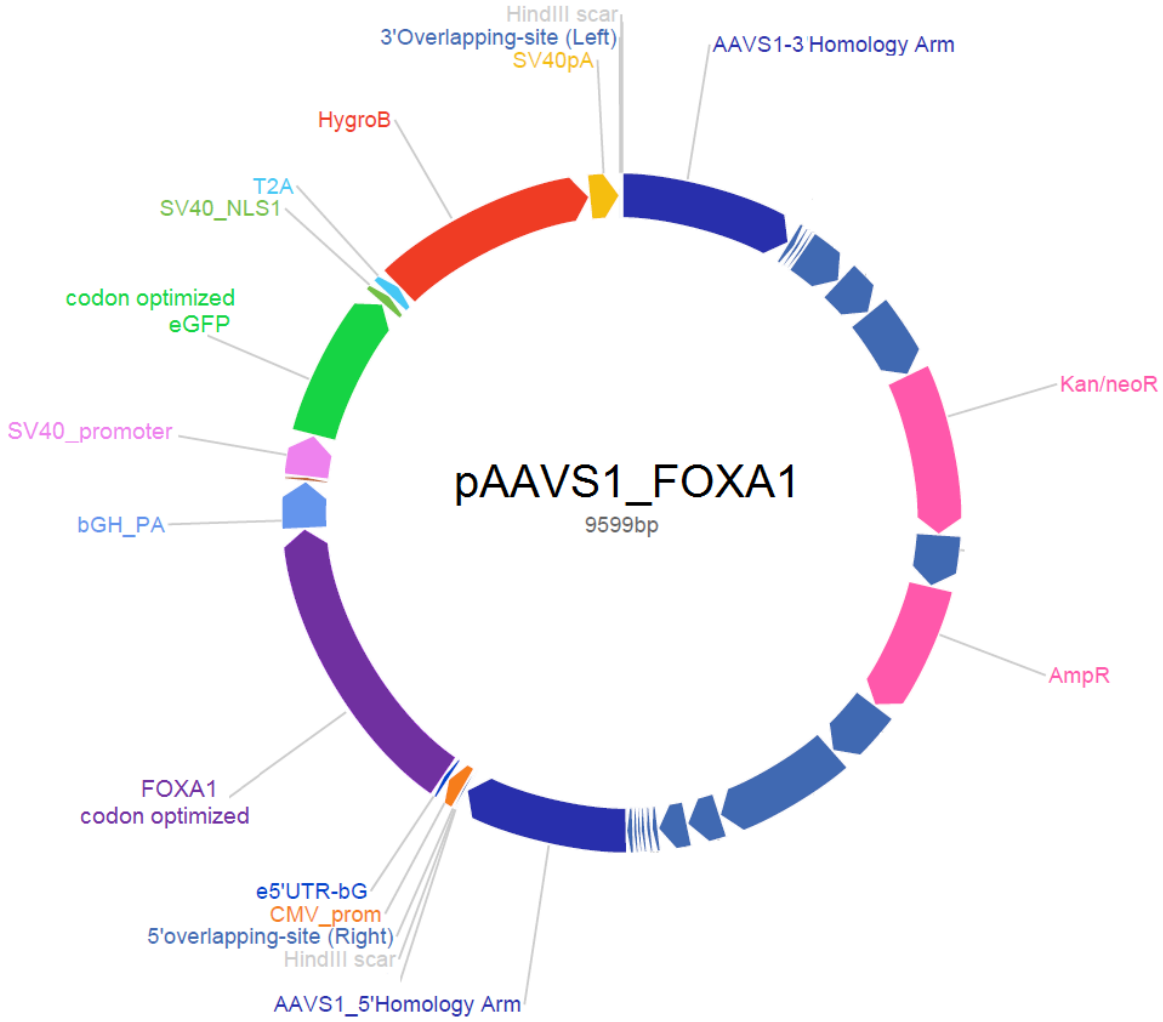


Figure 3.3 Plasmid map of the finished pAAVS1_*FOXA1* plasmid after Gibson Assembly cloning.

3.2 Guide RNA-design

3.2.1 Choice of CRISPR vector

Various companies offer different delivery systems for CRISPR-Cas9 genome editing that include i) all-in-one solutions where one sgRNA and Cas9 is expressed from the same plasmid vector, which confers a higher efficiency than expressing them in separate vectors. ii) all-in-one solutions for double nicking, where two tandem sgRNAs are expressed from the same plasmid. iii) sgRNA delivered as RNA, iii) Cas9 as mRNA, iv) Cas9 as purified protein, v) sgRNA and Cas9 delivered in a lentiviral vector for insertion into target genome, which is practical in cell types where expression would otherwise be difficult, and for high efficiency knockout screens.

For this project, Cas9 expression was only required transiently, long enough to create a DSB and allow the cell to perform a HDR using the constructed *FOXA1* repair template. There

was no desire to insert the Cas9n gene into the genome of the HBEC cells, as the goal was to insert *FOXA1*, and avoid genome editing apart from at the AAVS1 safe harbor locus. An all-in-one nickase system for transient transfection was determined to be the most suitable solution for this project. AddGene, a non-profit plasmid repository, offers many plasmids for transient Cas9 expression. The pSpCas9n(BB) plasmid (PX460) was chosen for the project, as it had been used with success in previous studies [44, 48, 52] . This is an all-in-one gRNA + SpCas9n plasmid, but it could only express one sgRNA, which meant that two plasmids would have to be transfected simultaneously along with the repair template construct. In this system, insertion of the gRNA into the vector must be done manually. An advantage to this is the versatility of being able to clone in any synthesized gRNA for future purposes.

Restriction cloning with *Bbs*I, a type IIs restriction enzyme, enabled the gRNA to be inserted into the vector. Enzymes of this type recognize asymmetric sequences, and cleave them at a defined distance from the restriction seat[82]. *Bbs*I cleaves six bp upstream on one strand, and two bp upstream on the other (Figure 3.4).

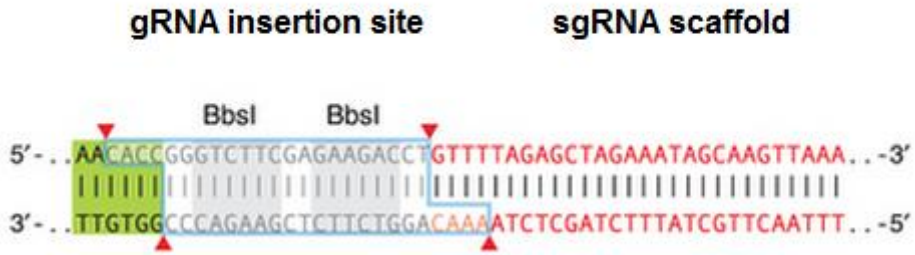


Figure 3.4 The double *Bbs*I recognition sites (grey rectangles) enable cleavage by BbsI, and the sites are removed (cut-out outlined with blue). Specific overhangs are generated which enable insertion of a DNA fragment with complementary overhangs. Modified from Ran *et al*. [48]

The digest leaves an overhang specific to the plasmid backbone, and by designing complementary overhangs at the gRNA oligo sequences, the gRNA duplex can be ligated into the plasmid (**Figure 3.5**). Two *Bbs*I recognition sequences adjacent to each other will effectively cut themselves out when cleaved by *Bbs*I. The gRNA oligos were designed so that they did not contain a *Bbs*I recognition site, and that enabled the restriction-ligation reaction to be repeated in several cycles for the purpose of increasing efficiency. The assembled product would be safe from cleavage by *Bbs*I, and since the overhangs in the plasmid did not match, it could not be ligated on itself without an insert.

3.2.2 Design of gRNA sequences with *Bbs*I-restriction overhangs.

To obtain suitable gRNA sequences, a 114 bp section of the first intron of the *PPP1R12C* gene (AAVS1 site) was plotted into the gRNA design software, and the option for nickase analysis was selected. The software provided two gRNAs, one on each DNA strand, with an offset of 9 bp.

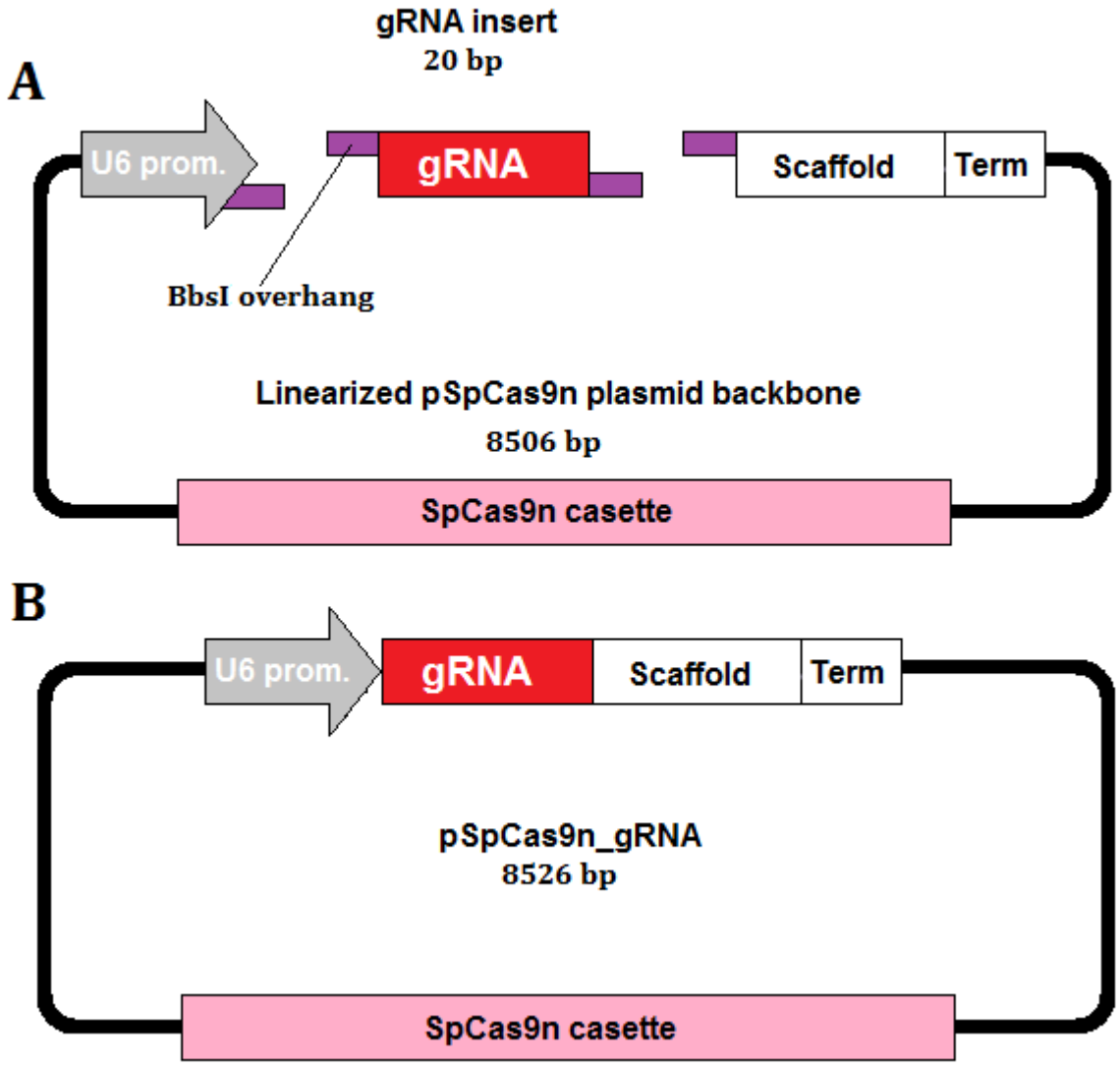


Figure 3.5 Schematic for scarless restriction-digestion cloning of gRNA into the pSpCas9n plasmid containin the sgRNA scaffold and Cas9n. A: The pSpCas9n plasmid is digested by *Bbs*I, and the gRNA oligo duplex have 4 nt overhangs that is complementary to the overhangs created by *Bbs*I digestion. B: Ligation of the gRNA into the insertion site provides a scarless product. When combined with the RNA scaffold, the gRNA sequence can be transcribed in a functional sgRNA by the U6 promoter.

This offset was within the recommended limit, thus the gRNA pair should be able to mediate a DSB. The software predicted no off-target activity from the pair. An additional gRNA design tool was used, plotting the same exact sequence. This software suggested the same gRNA sequences with no predicted off-target effects, and this consistency was considered a verification of the results. Single stranded sense and antisense oligos were synthesized for both gRNAs (IDT), and these would later be annealed to create a double stranded gRNA to be cloned into the pSpCas9n vector. Before synthesis of the gRNAs, 4 nt overhangs were added to the 5' ends of the single stranded oligos; CACC on the sense oligos, and AAAC on the antisense oligos. These overhangs would ensure scarless insert of the gRNAs into the vector. The gRNA sequences are listed in **Appendix C**.

3.3 Amplification of pAAVS1_*puro* and pSpCas9n from stab culture

The bacteria containing the plasmids were streaked on agar plates, and single colonies were picked for inoculation in liquid medium the following day. After secondary culturing, the plasmids were extracted using a silica spin column kit. The appropriate volume of bacterial culture to harvest was determined by OD_{600} measurements and calculation of the max culture volume that the spin columns had capacity for (**Table 3.1**). For both plasmids, the lowest dilution (1:1000) were picked for harvesting because these had the closest to optimal OD_{600} values (2-3, highlighted in bold in the table). The culture volumes harvested for purification was 80 mL for pAAVS1_*puro* and 60 mL for pSpCas9n.

Table 3.1 OD_{600}-measurements of cultures to determine maximum culture volume for plasmid purification. Several dilutions were made from the starter cultures, ranging from 1:1000 to 1:10000, before growing in larger scale. This was to ensure an optimal concentration of bacteria for the plasmid purification.

	pAAVS1_*puro*:		pSpCas9n:	
Dilution	OD_{600}	Max culture vol	OD_{600}	Max culture vol
1:1000	**1.687**	**88.9 mL**	**2.166**	**69.3 mL**
1:2500	1.658	89.3 mL	2.043	73.4 mL
1:7500	1.504	99.7 mL	2.053	73.0 mL
1:10000	1.400	107.1 mL	2.090	71.8 mL

The plasmids were run on an electrophoresis gel after isolation to verify the correct size. As shown on the gel image in **Figure 3.6**, the isolated plasmids appeared to be of the expected sizes of 6602 (pAAVS1_*puro*) and 8506 (pSpCas9n). DNA concentration was measured, and total mass was calculated, showing that there was a high DNA yield for both plasmids (**Table 3.2**).

Table 3.2 DNA yield after plasmid purification from bacterial culture

Plasmid	Concentration	Size	Mass	260/280	260/230
pAAVS1_*puro*	376.5 ng/µL	6602 bp	112.5 µg	1.89	2.07
pSpCas9n	305.1 ng/µL	8506 bp	91.53 µg	1.89	2.13

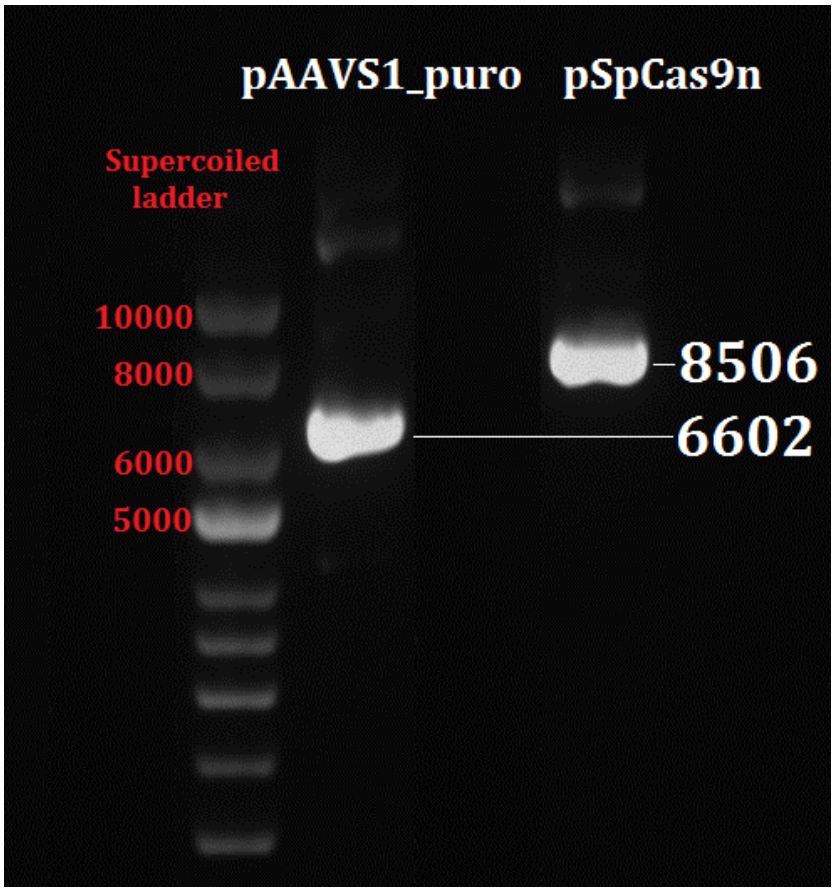


Figure 3.6 Gel electrophoresis of the purified plasmids. Supercoiled ladder was used for size comparison. Bands of approximately 7000 bp and 9000 bp were apparent after purification of the pAAVS1_*puro* and pSpCas9n plasmids respectively, which is in accordance with the expected band sizes (6602 bp and 8506 bp).

3.4 Amplification of gBlocks™

3.4.1 Primer design for amplification of gBlocks™

Primers for PCR amplification of gBlocks were designed using IDT's software tool, PrimerQuest. The whole construct was plotted, along with the desired approximate length of the blocks, and primers were suggested by the software. The overlapping region's size ranged from 99 to 161 bp. Primers creating shorter overlapping regions were also designed, to reduce possible formation of 3D structures. An alternative reverse primer was designed for gBlock1, forward and reverse for gBlock 2, and forward for gBlock3, making the overlapping regions from 54-83 bp in length. The relative positions and amplicons are shown in **Figure 3.7**, and sequences of all primers can be found in **Appendix B**. The primers are also highlighted in the nucleotide sequence of the pAAVS1_*FOXA1* plasmid in **Appendix C**.

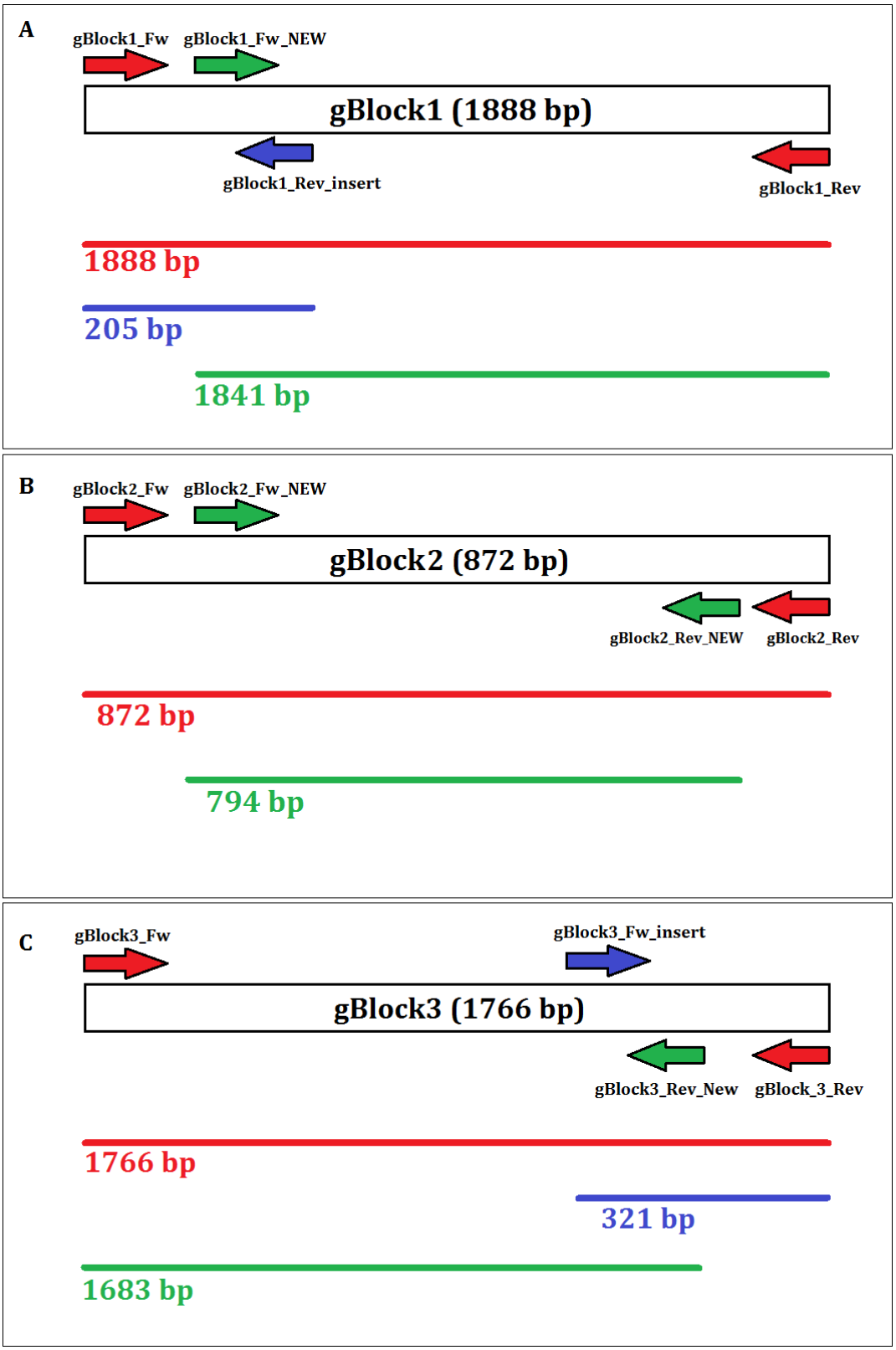


Figure 3.7 Relative positions of primers used for amplification of the three gBlocks. The red arrows represent the primers for the long overlapping regions, and the green arrows represent the primers for short overlapping regions. The blue arrow represents the primers used to test the integrity of gBlock1 and 3. The colored lines represent the expected amplicons created by the primers.

3.4.2 PCR of gBlocks™ with long overlapping regions

All three gBlocks were amplified with Q5 high fidelity polymerase (**Figure 3.8**). gBlock1 showed many unspecific bands of varying strength after amplification, including two strong bands; one just below 200 bp, and one at just above 1500 bp. The expected product of 1888 bp could not be amplified. Amplification of the correct band and elimination of the unspecific bands was attempted through systematically tweaking the PCR reaction conditions (**Appendix E**), but apart from eliminating one of the weaker unspecific bands, the optimization attempts were unsuccessful.

Amplification of gBlock2 gave the expected band of 872 bp, but unfortunately it was accompanied by several unspecific bands all ranging below 500 bp. Non-template controls ruled out that any of the bands were caused by primer dimers. Optimization of the PCR assay was conducted to eliminate the unspecific bands, but the attempts were unsuccessful (**Appendix E**).

Amplification of gBlock3 gave a clear band at just below 2000 bp, corresponding to the expected length of 1766 bp.

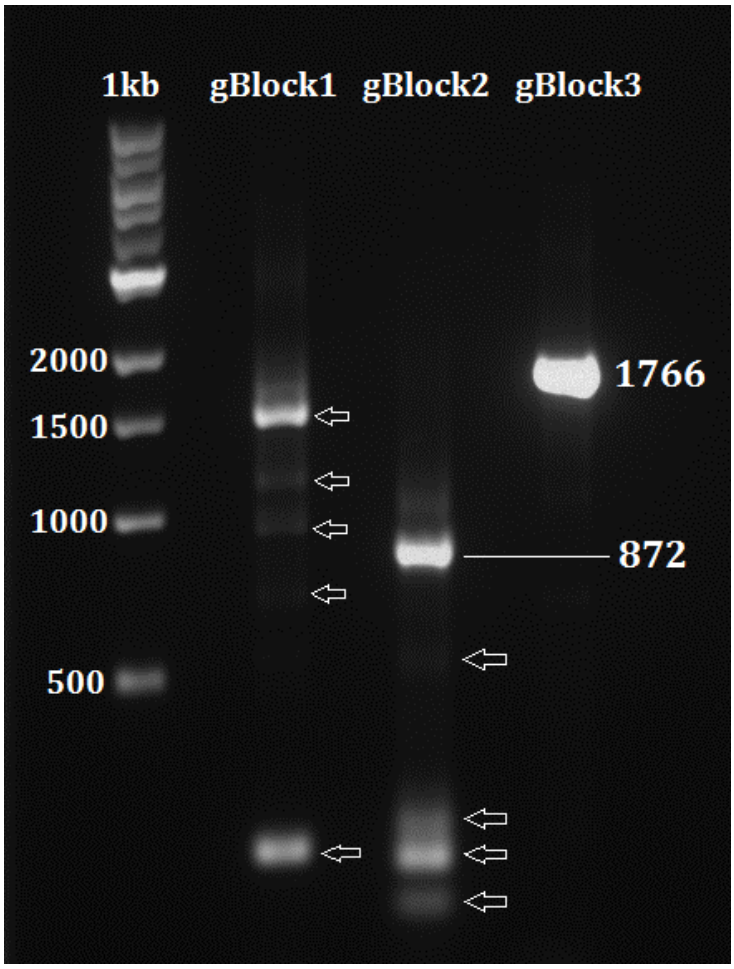


Figure 3.8 gBlocks amplified with Q5 polymerase for long overlapping regions. gBlock1 shows several unspecific bands of various sizes (indicated with arrows) and not the expected band of 1888 bp. gBlock2 shows a strong band corresponding to the expected size off 872, but also several unspecific bands of various sizes (indicated with arrows). gBlock3 only displays one strong band at the expected size of 1766.

3.4.3 PCR of gBlocks™ with short overlapping regions

All three gBlocks were amplified with new sets of primers giving somewhat smaller amplicons, and thus shorter overlapping regions for the subsequent Gibson Assembly (see **Figure 3.7**). The overall results with the new primers were similar to what was observed using the original primers; gBlock1 displayed only unspecific bands of equivalent sizes, gBlock2 showed the expected band as well as small, unspecific bands that were not a result of primer dimers, and gBlock3 only showed one bright band corresponding to the expected size (**Figure 3.9**). Optimization reactions were attempted on the assay for gBlock2, but they were unfortunately unsuccessful at eliminating the unspecific bands (**Appendix E**). No further attempts at optimization for gBlock1 was conducted, as the template showed evidence of being fragmented (presented in chapter **3.4.5**)

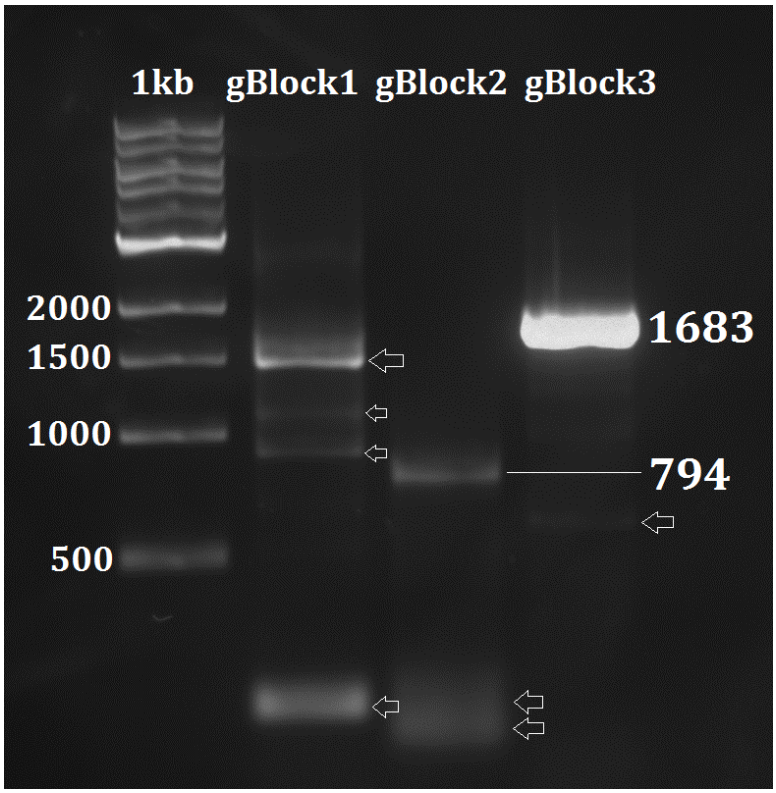


Figure 3.9 gBlocks amplified with Q5 polymerase for short overlapping regions. gBlock1 shows several unspecific bands of various sizes (indicated with arrows) and not the expected band of 1841 bp. gBlock2 shows a weak band corresponding to the expected size off 794, but also several unspecific bands of various sizes (indicated with arrows). gBlock3 displays one strong band at the expected size of 1683. There is also a very weak unspecific band at approx. 600 bp.

3.4.4 DNA concentration and purity of gBlocks™

After amplification, the samples were purified using silica beed purification kit. DNA quantification and purity measures are presented in **Table 3.3** and **Table 3.4**

Table 3.3 DNA-yield after amplification with long overlapping regions. (*gBlock1 was not amplified correctly, so the size of the purified product is not the expected size of 1888 bp)

Fragment	Concentration	Size	Total mass	260/280	260/230
gBlock1	129,7 ng/µL	1888 bp*	1,556 µg	1.63	0.81
gBlock2	138,5 ng/µL	872 bp	1,662 µg	1.71	1.02
gBlock3	199,6 ng/µL	1766 bp	2,395 µg	1.87	1.76

Table 3.4 DNA-yield after amplification with short overlapping regions

Fragment	Concentration	Size	Total mass	260/280	260/230
gBlock1	Not amplified	1841 bp	-	-	-
gBlock2	340.6 ng/µL	794 bp	3.406 µg	1.82	0.32
gBlock3	395.6 ng/µL	1683 bp	3.956 µg	1.66	0.68

3.4.5 Agarose gel electrophoresis of unamplified gBlock™ templates

When the optimization reactions proved ineffective at providing optimal amplicons of gBlock1 and gBlock2, the unamplified templates were tested by running them on an agarose electrophoresis gel (**Figure 3.10**). The template of gBlock1 showed evidence of being fragmented, and there was weak, if any evidence that the complete gBlock1 fragment was present in the sample. The template of gBlock2 appeared to be correct and intact.

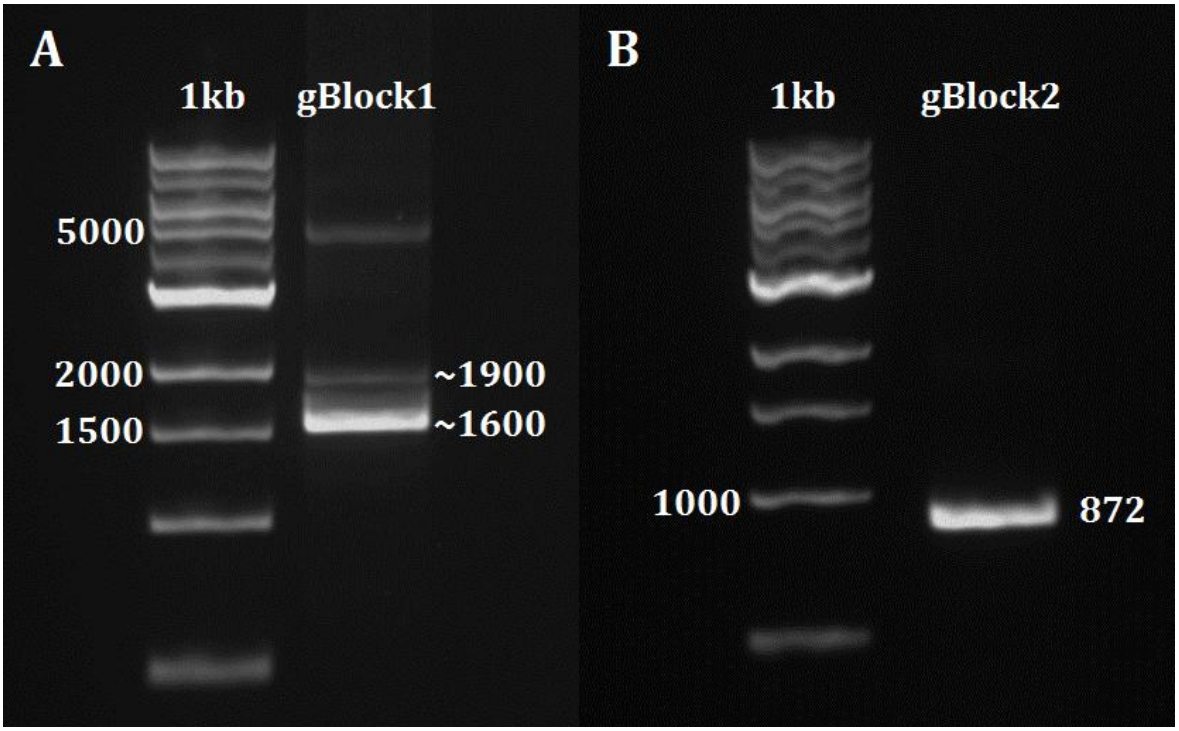


Figure 3.10 Gel picture of unamplified template gBlocks run on agarose gel electrophoresis. A: gBlock1 looks smeared on the gel suggesting that the template is fragmented. The brightest part of the smear is just above 1500

bp. There is a weak band at approximately 1900 bp, which means that gblock1 (1888 bp) may represent a fraction of the sample. Additionally there is a band at approx. 5000 bp. Many of the unspecific bands seen after PCR are not visible in the template. B: gBlock2 has a clear band at just below 1000 bp that corresponds with the expected size of 872 bp There are no unspecific bands visible in the template, indicating that they arise during PCR.

3.5 Linearization of pAAVS1

3.5.1 Restriction digest with *Hin*dIII

To linearize the pAAVS1 plasmid and simultaneously excise the 992 bp *PuroR* cassette, the pAAVS1_*puro* plasmid was digested with *Hin*dIII (**Figure 3.11**). The digest resulted in two fragments of the expected sizes (5610 bp and 992 bp). The digested plasmid was then purified from the digestion mix and quantified (**Table 3.5**). The *PuroR* cut-out was not removed, as the UV-light required for gel extraction of pAAVS1 could cause unnecessary DNA damage to the plasmid.

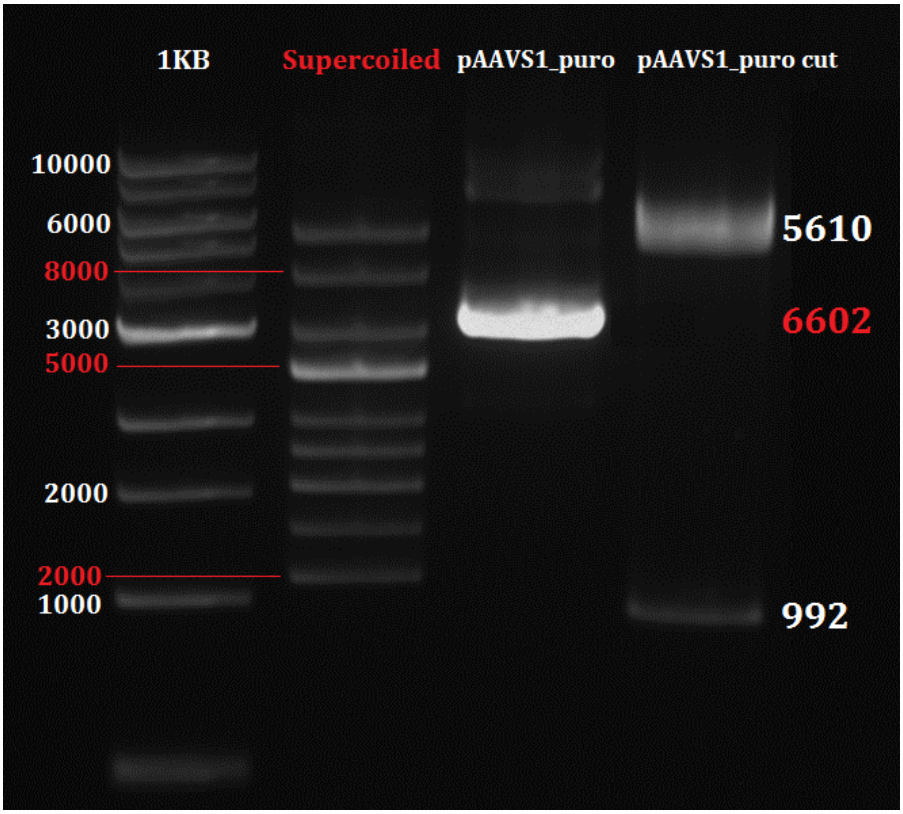


Figure 3.11 pAAVS1_*puro* digested with HindIII. An undigested pAAVS1_*puro* is run simultaneously on the gel for comparison. The size indications are shown in white for 1kb ladder and red for supercoiled ladder. The digested plasmid (pAAVS1_*puro* cut) shows two bands (approximately 1000 and 6000 bp) that correspond to the expected lengths of the 992 bp cut-out and 5610 bp linearized pAAVS1 plasmid.

Table 3.5 DNA yield of restriction digested pAAVS1 after PCR cleanup

Plasmid	Concentration	Size	Mass
pAAVS1_*puro*_cut	401.7 ng/µL	5610 + 995 bp	4.42 µg

3.5.2 Linearization by PCR amplification

In order to reduce the chance the pAAVS1_*puro* vector being recovered after linearization, the pAAVS1 plasmid was linearized using high fidelity PCR in addition to restriction digest. Both approaches have advantages and disadvantages discussed in chapter **4.1**. Primers were designed to anneal as close as possible to the *Hin*dIII site, leaving a blunt end amplicon. This ensured that the whole plasmid sequence was intact, while preventing the plasmid from ligating on itself without the *FOXA1* insert, which was a possible outcome for the *Hin*dIII linearized plasmid. The plasmid was amplified, and showed one clear band at the expected size (**Figure 3.12**)

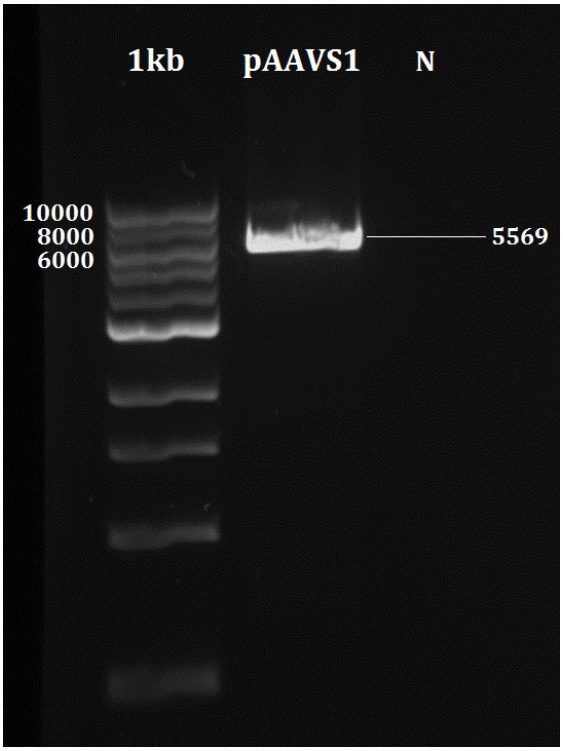


Figure 3.12 Electrophoresis gel of PCR linearized plasmid. The expected amplicon is visible on the gel as a clear band at approximately 6000 bp corresponding to the 5569 bp amplicon.

The PCR product was purified and quantified for subsequent use in Gibson assembly (**Table 3.6**)

Table 3.6 DNA-yield of PCR linearized pAAVS1 vector after PCR cleanup

Fragment	Concentration	Size	Total mass	260/280	260/230
pAAVS1	340.6 ng/µL	5569 bp	3.4 µg	1.62	0.75

3.6 Gibson assembly® cloning

3.6.1 Gibson assembly® cloning using long overlapping regions and restriction digested vector

The *Hin*dIII linearized pAAVS1 plasmid and the three amplified gBlocks were attempted cloned together by Gibson Assembly. Unfortunately a gel electrophoresis of the Gibson-mix showed no signs of any of the components being assembled to each other (**Figure 3.13**). The gel displayed a fragment at the size of the pAAVS1_*puro* (~6500 bp), bands corresponding to all three gBlocks, and also the 992 bp cut-out representing the *PuroR* expression cassette. A positive control that was provided in the kit was run along with the samples. The control consisted of a pUC19 linearized vector and a DNA fragment with overlapping regions, and after assembly, it was used to transform competent bacteria. The agar plates had rich growth of single colonies.

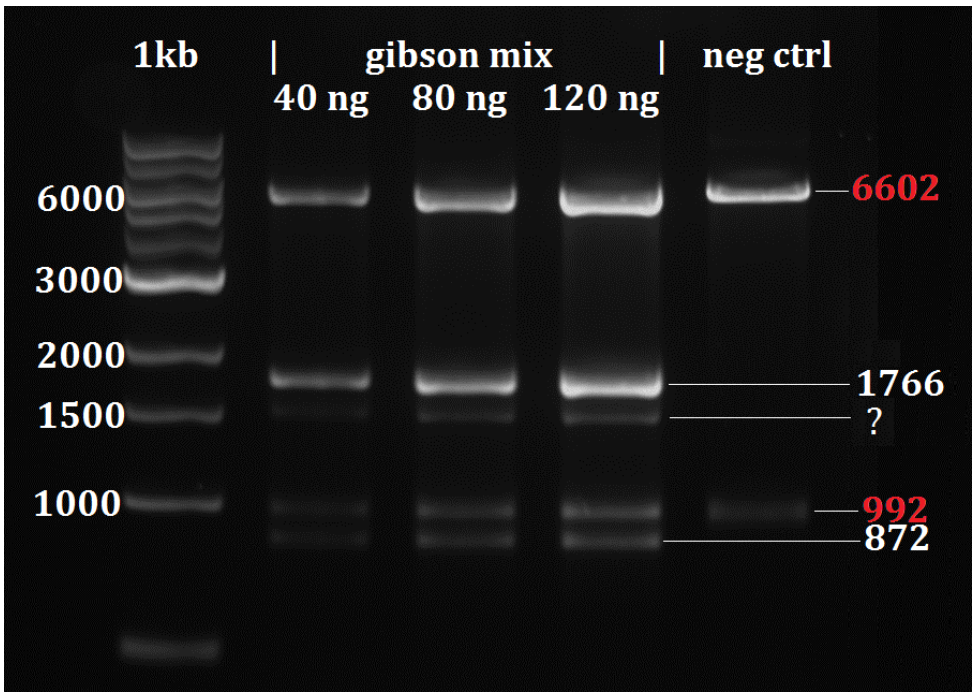


Figure 3.13 Electrophoresis gel of Gibson assembly reaction mix. The figure shows bands representing the ligated plasmid and the three singular gBlocks, plus the cut out piece of pAAVS1_*puro* plasmid. Perhaps the small bands from the plasmid were added in molar excess, and that's why we see a band on 992, but not 5610. Different mass of DNA from the reaction mix was loaded on the gel to visualize possible weaker bands.

To determine whether the ~6500 bp band was the pAAVS1_*puro* vector that had ligated back to its original form, the plasmid was used to transform competent 5-alpha cells, purified and digested with *Hin*dIII. Although the digestion was not complete, it was clear that two bands of the expected size appeared, confirming that the 992 bp cut-out *PuroR* cassette had been ligated into the vector again (**Figure 3.14**).

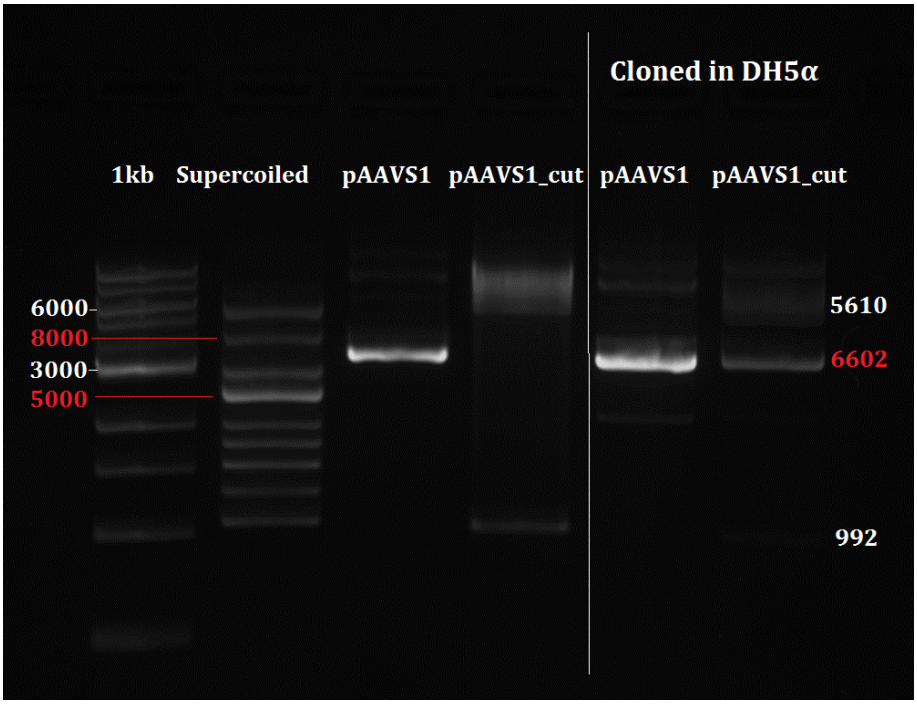


Figure 3.14 A restriction digest was performed on the purified plasmid from the Gibson Assembly reaction, and compared with the original supercoiled and digested pAAVS1 plasmids. The gel electrophoresis displayed bands of ~6000 bp and ~1000 bp corresponding to the two digested fragments, and one band at ~6000 (supercoiled) that indicates an incomplete digestion

3.6.2 Gibson assembly® cloning using long overlapping regions without vector

When the vector proved difficult to assemble with Gibson because of its *Hin*dIII overhang, an assembly was attempted without the vector, to see if the gBlocks alone could be combined into one construct. If this strategy would work, perhaps the construct could be cloned into the vector by restriction digestion/ligation. The reaction mix was set up identically to the first attempt at assembly, with the volume of the vector replaced by nuclease free H_2O. Subsequent gel electrophoresis of the reaction mix unfortunately showed that there appeared to be no assembly of any of the gBlocks, as the bands on the gel corresponded to the bands seen after amplification of the gBlocks (**Figure 3.15**). A parallel reaction mix was run with DMSO (5 %) to eliminate possible 3D structures, but this did not result in any observable difference in band pattern.

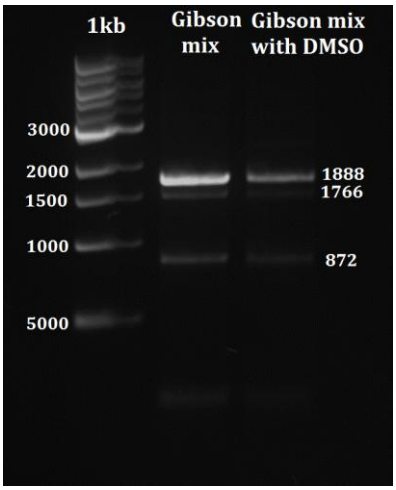


Figure 3.15 Electrophoresis gel of Gibson assembly reaction mix with just gBlocks. The figure shows only bands that correspond to the PCR product of each gBlock separated, and no assembled products.

3.6.3 Gibson assembly® cloning using short overlapping regions and PCR linearized vector

Lastly, the Gibson assembly was repeated using the PCR linearized pAAVS1 vector and gBlock 2 and 3 with short overlapping regions. The PCR-linearized pAAVS1 should not have any problems with self-ligation, and shorter overlaps should give less potential problems with 3D structures. gBlock1 was omitted from this assembly because of the problems with its amplification. Four parallels were made, two incubated for two hours, and two incubated for four hours (**Figure 3.16**). The longer incubation time was to ensure that the whole sequence of the overlapping regions would be exposed by the endonuclease. DMSO (5 %) was added to one of the parallels at each incubation time. None of the fragments showed signs of being assembled when the reaction mixes was run on an electrophoresis gel.

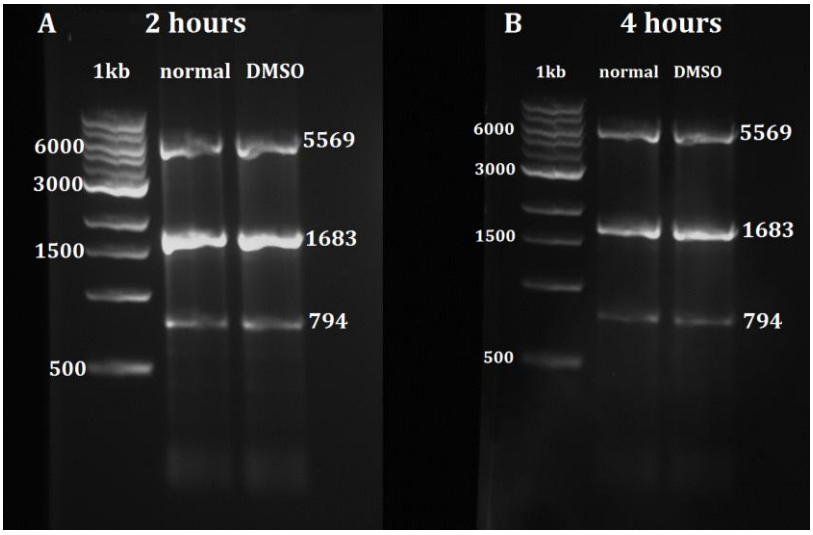


Figure 3.16 Gibson assembly of pAAVS1, gBlock2 and gBlock3 with short overlapping regions. Only bands corresponding to the added fragments were visible (794, 1683 and 5569 bp), which indicated no assembly either after 2 or 4 hour incubation.

3.7 Transformation of DH5α with pAAVS1_*FOXA1*

3.7.1 Primer design for colony screening

The primers for colony screening of pAAVS1_*FOXA1* were designed to give PCR products at different size, depending on whether the plasmids contained the correct inserts. They were therefore made to anneal in or outside of the homology arms. Two sets of primers with these qualities were designed; one set that annealed outside of the homology arms (primer pair 1), and one set within them (primer pair 2). Because the PCR was to be performed with bacterial cultures as template, the primers were tested for, and showed no predicted complementarity to the DH5α *E. coli* genome. All four primers were designed with equal Tm so they could be used interchangeably, and expected amplicon sizes from amplification of pAAVS1 with and without cassette inserts is listed in **Table 3.7**.

Table 3.7 Expected size of colony screening amplicons

Plasmid	Primer pair 1	Primer pair 2	Fw1 + Rev2	Fw2 + Rev1
pAAVS1_*FOXA1*	5759 bp	4733 bp	5199 bp	5293 bp
pAAVS1_*puro*	2738 bp	1712 bp	2178 bp	2272 bp
pAAVS1	1746 bp	720 bp	1186 bp	1280 bp

3.7.2 Testing primers for colony screening

Pair 1 of colony screening primers was tested by *Taq* PCR with pAAVS1_*puro* as template (**Figure 3.17**), but the assay did not result in electrophoresis bands of the expected size of 2738 bp. There were however several unspecific bands of varying sizes present, as well as primer dimers.

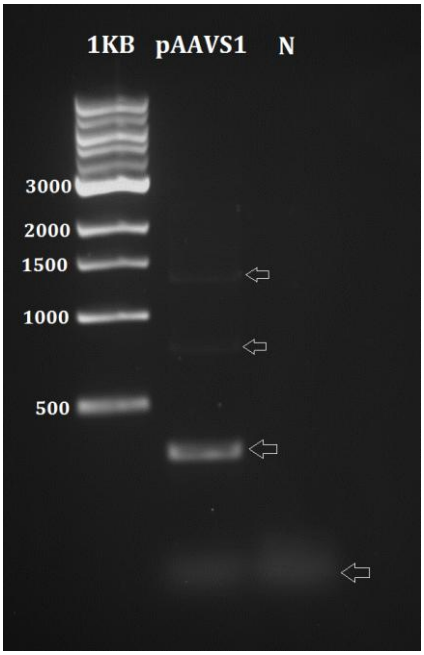


Figure 3.17 Test PCR of colony screening with primer pair 1. The expected band of 2738 bp was not present after the amplification, but unspecific bands were observed at approx. 400, 800 and 1400 bp. A weak band of small size could be observed in sample and non-template control, indicating primer dimerization.

Both pairs of primers were then tested with *Taq* PCR using pAAVS1_*puro* as template (**Figure 3.18**). In addition to testing them in their intended pairs, the forward and reverse from pair 1 and 2 were combined with each other as alternative pairs (Fw1 + Rev2 and Fw2 + Rev1). As shown in the first assay, primer pair 1 only displayed unspecific bands and primer dimers. The first and second primer pair were both unable to create amplicons of the expected sizes, and neither did combinations of the two pairs. All primers formed primer dimers in the negative control, and the dimers were also present in all samples, although not all dimers are seen clearly in the figure.

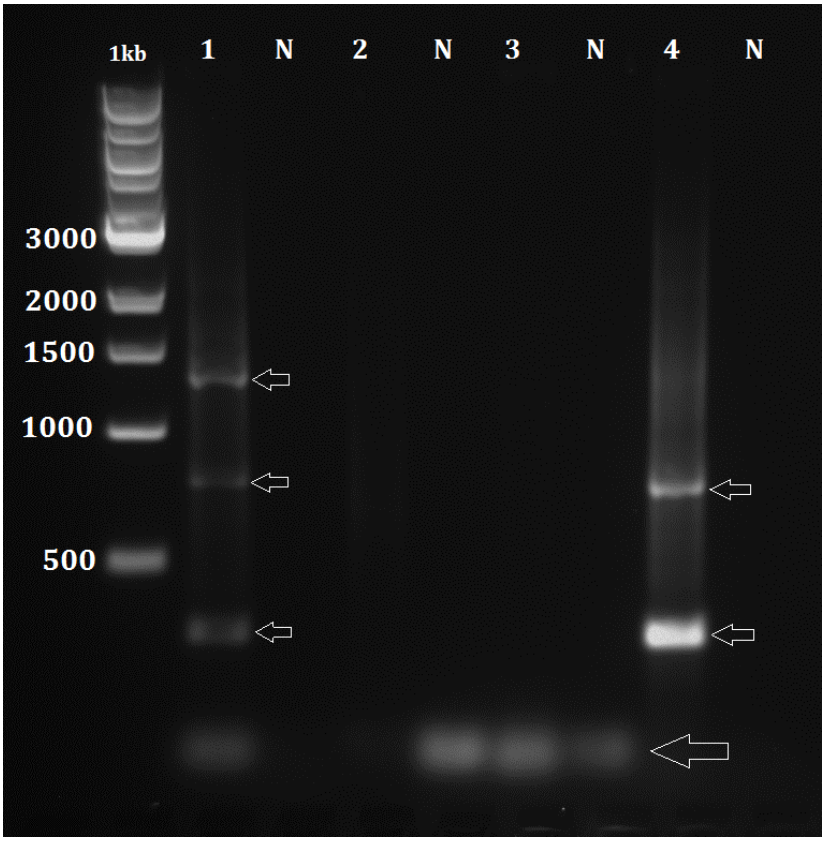


Figure 3.18 Test PCR colony screening of pAAVS1_*puro* with both primer pairs.1: Primer pair 1, 2: Primer pair 2, 3: Fw1 + Rev2, 4:Fw2 + Rev1. Primer pair 1 does not show a band at the expected size (2738 bp), but unspecific bands can be observed at approx. 400, 800 and 1400 bp, and a small band indicates primer dimerization. Primer pair 2 does not display any bands (expected size: 1712 bp), but non-template control shows a band corresponding to primer dimers. The Fw1 + Rev2 pair also shows a primer dimer-band, but no other bands (expected size: 2178 bp). The Fw2 + Rev1 pair shows two clear bands at approximately 400 and 800 bp, which match some of the unspecific bands seen with primer pair 1. The expected band (2272 bp) is not present.

Further optimization of these primers is required to obtain stable PCR amplification, as it is imperial to be able to screen for colonies containing the correct insert, and avoid unnecessary isolation and sequencing of plasmids from negative cultures. If further optimization proves unsuccessful, new primers must be designed.

3.8 Transformation of DH5α with pSpCas9n_gRNAs

3.8.1 Design of primers for colony screening

Primers for the pSpCas9n_gRNA1 and pSpCas9n_gRNA2 vectors were designed to anneal inside the inserted gRNA sequences, and in the plasmid backbone creating amplicons of 271 and 341 bp. These primers were therefore specific to the inserted gRNA sequence and could only anneal if the respective gRNAs were inserted correctly.

3.8.2 Insertion of gRNA into pSpCas9n

Both gRNA1 and gRNA2 was attempted cloned into the pSpCas9n vector using *Bbs*I digestion and ligation in cycles. This method should give minimal background growth of negative colonies (vector without insert) because the *Bbs*I recognition site is cut out of the vector during digestion. Competent bacteria were then transformed with the restriction ligation mixes. There was only one colony of each plasmid growing on agar after transformation, and they were picked for subsequent culturing. Colony screening of the pSpCas9n_gRNA1 plasmid was negative, indicating that gRNA1 had not been cloned into the vector. The colony screening of pSpCas9n_gRNA2 showed two bands at the approximately correct size (341 bp), an indication that gRNA had been successfully cloned into the vector. (**Figure 3.19**)

Since the primers were insert-specific, they could not be tested prior to the colony screening. Both plasmids were therefore purified from culture, even though colony PCR had indicated that gRNA1 was not inserted into the vector, because it was impossible to determine if the result was a false negative.

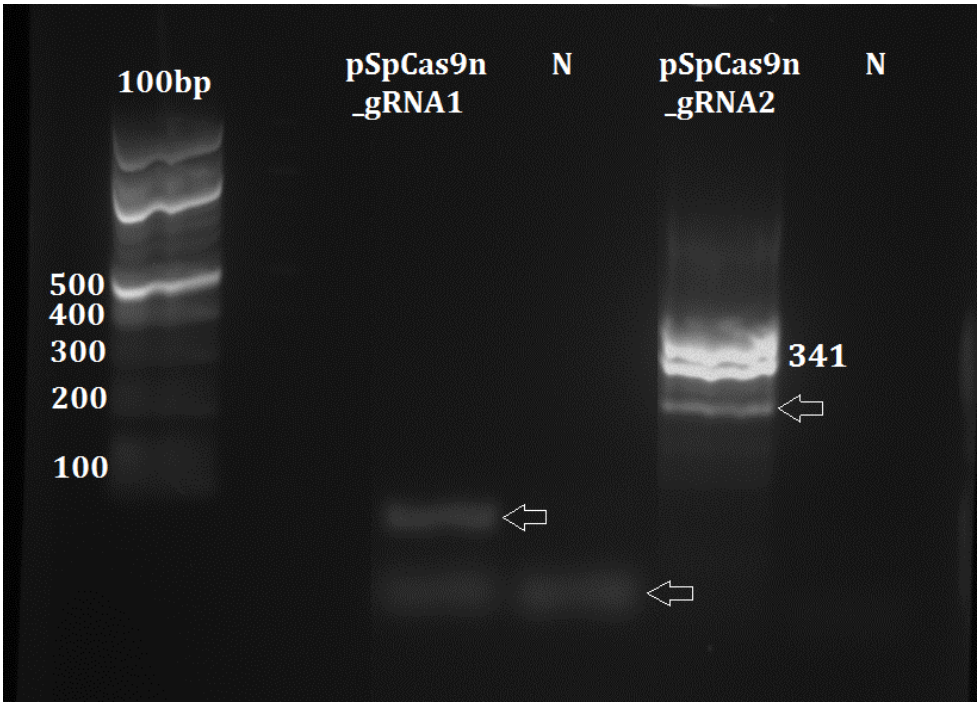


Figure 3.19 Colony screening of the attempted cloned pSpCas9n_gRNA1 and pSpCas9n_gRNA2: The amplicon size for pSpCas9n_gRNA1's is 271 bp, but this band is not present on the gel. There is one unspecific band visible at ~90 bp, and one smaller band that is also present in the non-template control, indicating formation of primers dimers. For pSpCas9n_gRNA2 there are two strong bands visible at around 300-400 bp, and one of them might be the expected product (341 bp) A weak band of ~250 bp is visible as well.

The PCR assay was repeated after the plasmids were isollated from culture. In addition to testing the primer pairs on their matching templates, they were both tested on each of the isolated plasmids as well as the original pSpCas9n vector. The gRNA1 primer pair assay did not show any significant difference between the three plasmids, while the gRNA2 primer pair assay showed an evident difference between the plasmids suggesting that the gRNA2 had been cloned successfully into the vector (**Figure 3.20**).

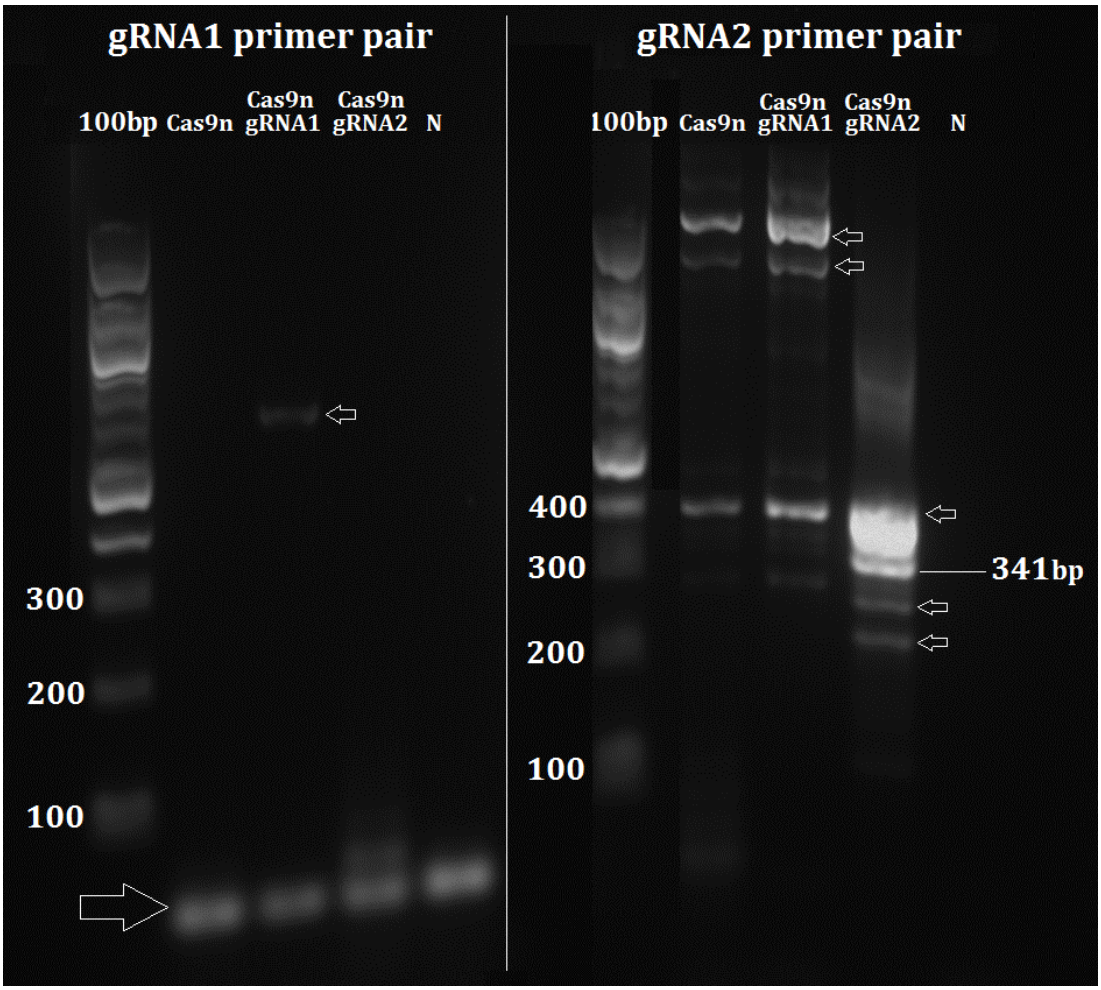


Figure 3.20 PCR on purified plasmids to test primers on different templates. All three plasmids were tested using the primer pair for gRNA1 and the primer pair for gRNA2. The gRNA1 primer pair does not give any amplicons from either of the plasmids, but an unspecific band can be seen from pSpCas9n_gRNA1 at approximately 800 bp. The primers formed dimers in all reactions. The gRNA2 primer pair displays several unspecific bands from all plasmids, but while the pSpCas9n vector without insert and with possible gRNA1 insert has an identical pattern, the plasmid with gRNA2 insert had a ~350 bp band corresponding to the predicted amplicon size, as well as two other, weaker bands at ~200-300 bp that is not seen in the other two plasmids. The other two plasmids show two unspecific bands at >1500 bp, and these bands are not present when using pSpCas9n_gRNA2 as a template.

The plasmids were both run on a gel for size verification after plasmid isolation, and pSpCas9n was run along with them for comparison (**Figure 3.21**). All plasmids had the correct size.

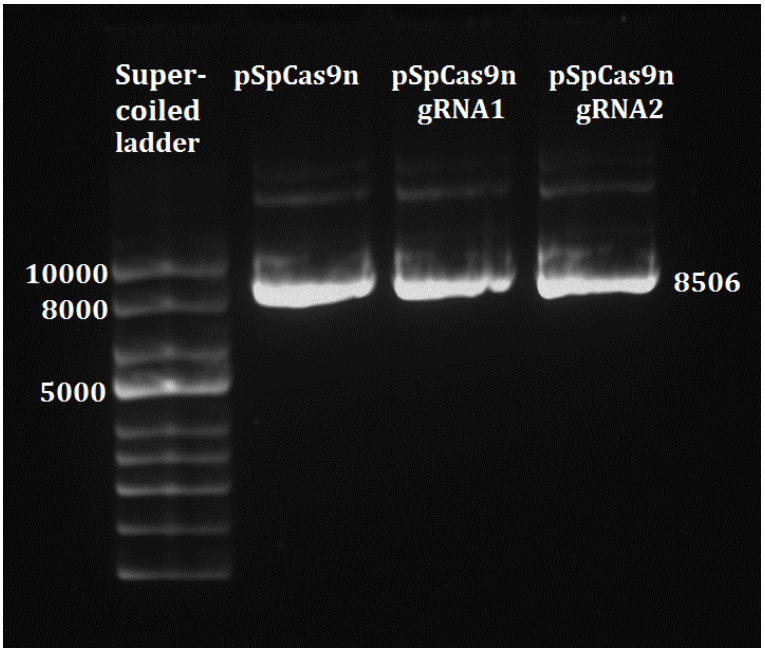


Figure 3.21 Cloned plasmids run on a gel after plasmid purification, and compared with original pSpCas9n. The gel displays bands corresponding to the expected sizes of 8506 (pSpCas9n) and 8508 (pSpCas9n_gRNAs).

Table 3.8 DNA yield after plasmid purification of cloned pSpCas9n_gRNAs

Fragment	Conc	Size	Total mass	260/280	260/230
pSpCas9n_gRNA1	317.3 ng/µL	8508 bp	95.19 µg	1.9	2.07
pSpCas9n_gRNA2	296.7 ng/µL	8508 bp	89.01 µg	1.9	3.58

3.9 Sequencing

Before proceeding with co-transfection of HBECs with the pAAVS1_*FOXA1* donor plasmid and the pSpCas9n_gRNA plasmids, the sequence of the plasmids had to be verified. A value read from Eurofins Genomics has an expected reading length of 500-800 bp, and a maximum of 1100 bp with optimal conditions. The first ~100 base pairs immediately following the primer are not expected to give a clear sequence read.

Sequencing primers for the pAAVS1_*FOXA1* construct were designed, but as the assembly of the plasmid was not successful, sequence reads of the construct could not be performed.

3.9.1 Primer design for sequencing of pAAVS1_*FOXA1* construct

The sequence primers were designed to anneal approximately 500 bp apart from each other, with alternating forward and reverse primers (**Figure 3.22**). The alternating orientations of the primers should give some length of overlapping reads, and increase the chance of having a complete combined sequence read of the whole construct (~5000 bp). The sequence primers are listed in **Appendix C**.

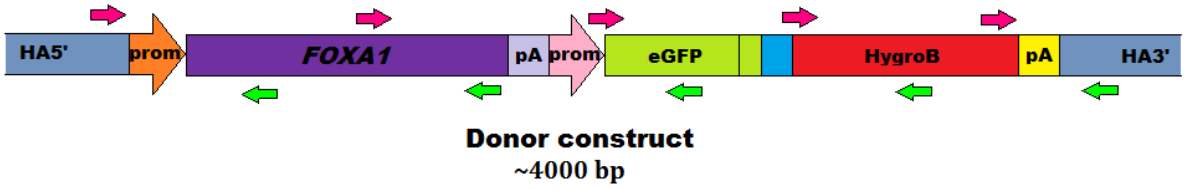


Figure 3.22 Relative positions of the sequencing primers in the *FOXA1* construct (cloned into the pAAVS1 plasmid). Pink arrows indicate positions of the forward primers, and green arrows indicate the reverse primers.

3.9.2 Primer design for sequencing of pSpCas9n_gRNAs

The size of the inserted gRNA sequences were 20 bp, meaning that one sequencing primer in an appropriate length from the gRNA inserts would be sufficient for a sequence read. A forward sequencing primer was designed, annealing at an approximately 250 bp distance to insert. The sequence primer is listed in **Appendix B**.

3.9.3 Sequence reads of pSpCas9_gRNAs

Sequence reads of the pSpCas9n_gRNAs are presented in **Appendix D**.

The sequence read for the isolated pSpCas9n_gRNA1 matched the sequence for the original pSpCas9n plasmid, suggesting that the cloning of gRNA1 into the vector was unsuccessful.

The pSpCas9n_gRNA2 sequence was verified, suggesting that the cloning of gRNA2 into the vector was successful.

4 Discussion

Aberrant expression of *FOXA1* is associated with cancer development, and there are indications that steroid pathways regulated by this transcription factor exhibit cross-talk with carcinogen metabolism pathways [24, 25]. At STAMI, *FOXA1* has been subjected to experimental studies related to EMT in HBEC cells, and long-term exposure experiments with chemical carcinogens such as B[*a*]P, CSC and MNU has resulted in EMT and downregulation of *FOXA1* [30, 31].

The aims of this study was to develop a vector system for site-specific integration of *FOXA1* in CSC exposed HBEC cells, and to determine the role of *FOXA1* in maintaining their transformed phenotype. To achieve site-specific integration of *FOXA1*, the CRISPR-Cas9 system was used to create a DSB in the safe harbor locus AAVS1, and a repair template containing *FOXA1* was used for homologous recombination at the DSB, thus inserting the gene in place. First, a brief summary of the results obtained:

A *FOXA1* repair construct was designed and synthesized into three fragments (gBlocks), which were amplified with Q5 polymerase. gBlock1 could not be amplified and gBlock2 was amplified with the appearance of unspecific PCR products. Only gBlock3 was amplified correctly without any detectable unspecific products. The pAAVS1 donor vector for insertion of construct was linearized by both restriction digestion and PCR amplification, and both were tested for Gibson Assembly with the amplified gBlocks. No assembly products were generated, and the restriction digested vector ligated upon itself. Primers for colony screening of pAAVS1 were designed, but initial PCR with *Taq* polymerase showed that these primers did not give the expected amplicons from pAAVS1. gRNA sequences (gRNA1 and gRNA2) were designed to guide Cas9n to double nick within the AAVS1 safe harbor. The gRNA sequences were received as single stranded oligos, annealed, and inserted into the scaffold sequence of the pSpCas9n vector using BbsI restriction cutting. gRNA2 was successfully cloned into the vector and sequence verified, but gRNA1 was not.

4.1 *FOXA1* construct design, synthesis and Gibson® assembly

4.1.1 DNA purification methods used in this work

Using pure DNA is important in any cloning reaction, but perhaps even more so in Gibson Assembly cloning, that relies on relatively long overlapping regions between fragments of which any unspecific binding would be inhibitory. Two different kits were used to purify DNA in this project:

Plasmid DNA from bacterial cultures was isolated and purified with a column based plasmid MidiPrep kit designed for high plasmid yield. The kit provided a high yield of all plasmids (pAAVS1_*puro*, pSpCas9n, pSpCas9n_gRNA1 and pSpCas9n_gRNA2), and 260/280 nm and 260/230 nm ratios were above 1.89 and 2.0 respectively for all samples. This is well within the levels considered to be pure, and indicates minimal contamination of proteins, phenols, salts and polysaccharides.

PCR products and restriction digested plasmids were purified with a silica bead PCR cleanup kit that was applicable either on DNA excised from agarose gel, or directly from the reaction mix. Gel extraction was not used in this project, as it was decided that the potential damage

of UV light to the DNA had be avoided. The drawback of purifying directly from reaction mix is that all DNA is recovered, including unspecific products. Additionally, any proteins bound to DNA will be recovered, as there is no phenol extraction involved in this kit. An up to 80 % recovery of DNA could be obtained by using this kit. Unfortunately, the samples purified in this kit were generally not very pure. The 260/280 ratio was between 1.6-1.89, meaning that none of the samples had much protein contamination. The 260/230 ratios were all below 1.6 with one exception, and for many of the samples, the ratio was below 1. This indicates the presence of significant contamination absorbing light at 230 nm, most likely caused by excess silica beads. The elution volume is very small (~5 µL), and carry-over from pipetting the supernatant off pelleted silica beads is not uncommon. A way to circumvent this could be to amplify more DNA, and collect them in one purification reaction, as the kit instructs to scale up the elution volume with the DNA mass purified. Generally, column based kits are better for avoidance of this type of contamination, as the silica is in a solid phase.

For future work, purifying gBlock PCR products by gel extraction should be considered, to remove the unspecific products generated during the amplification. There are measures that can be done to minimize UV-induced DNA damage; which include using long-wavelength UV light and exposing the gel for a minimum amount of time.

4.1.2 Construct design

Various software tools were used to aid the design of the *FOXA1* construct. These tools allowed for visualizing and combining sequences, identifying features, aligning sequences and perform sequence edits. The commercially available Genome-CRISP™ knock-in kit from GeneCopoeia provided a repair template construct containing an expression cassette with a gene of interest and one expression cassette with GFP and Puromycin as selection markers. The idea of combining two cassettes in one construct flanked by AAVS1 homology arms was used as an inspiration for design of the *FOXA1* knock-in construct. The expression of one fluorescent marker and one antibiotic resistance gene from the same strong promoter was enabled by separating the genes with a 2A cleaving peptide [81].

The sequences used to build the construct were obtained from various sources, listed in **Appendix C**. The final *in silico* design contained a *FOXA1* expression cassette and an *eGFP*-T2A-*HygroB* bicistronic expression cassette. *FOXA1* and *eGFP* were codon optimized for ease of chemical synthesis, and a nuclear localization signal was attached to *eGFP* to obtain fluorescent nuclei of transfected cells. Literature studies were conducted to select suitable promoters for expression. Even though comparative studies had shown that promoters such as CAGG and Ef1a were the strongest and most stable promoters in various mammalian cell types [76], they were both large and high in repetitive elements that likely could prove difficult in both synthesis and subsequent Gibson Assembly. Therefore the mid to high expression promoters CMV and SV40 were chosen. Strong expression of the reporter/selection marker cassette was considered to be of highest priority, to enable efficient enrichment of transfected cells by culturing in medium with HygromycinB. The more stable SV40 promoter was thus chosen for expression of the reporter/selection marker cassette.

Whether choices in construct components would yield desired expression levels of *FOXA1* and selection markers in HBECs remains to be seen, as successful assembly was not achieved during the course of this thesis. In the following section possible reasons for this will be discussed.

4.1.3 gBlock™ design and amplification

The final *FOXA1* construct was ~4000 bp long, and the 2000 bp size limit for gBlock fragments synthesized by IDT meant that the construct had to be divided into three fragments (sized 1888 bp, 872 bp and 1766 bp). The gBlocks had overlaps (99-161 bp) with each other and the plasmid vector, to enable Gibson Assembly into a circular plasmid containing the construct. The gBlocks were attempted amplified using Q5 high fidelity polymerase, but the results were not optimal. gBlock1 could not be amplified, and exhibited many unspecific bands on agarose gel. gBlock2 was amplified, albeit not strongly, and it was accompanied with multiple unspecific products apparent on agarose gel. Only gBlock3 was amplified successfully with no unspecific products. The possible reasons for unspecific bands could be many, e.g. unspecific annealing of the primers, too many PCR cycles, too much template, and too much primer. These possible reasons were investigated by running a series of optimizations of the PCR assays for gBlock1 and gBlock2, tweaking both reaction mix components and thermocycler program settings. Unfortunately, gBlock1 could not be amplified despite all efforts, and the unspecific bands remained in both assays. New primers were designed, generating shorter overlaps (54-83 bp) between the gBlocks. The problems of gBlock1 and gBlock2 remained after amplification with these primers, even after efforts to optimize the assay for gBlock2.

When the unamplified gBlock templates were run on a gel, it became apparent that gBlock1 did not have the expected integrity. It was fragmented, and did not appear to contain the expected band of 1888 bp. The template of gBlock2 however appeared to be intact. IDT, the manufacturers of gBlocks were at this point consulted about the difficulties in amplification. The advice given by IDT was to avoid PCR amplification of fragments larger than 1000 bp due to preferential amplification of lower level contaminants, and they recommended maximum 10-12 cycles if attempted.

It was unclear why the gBlock1 template had been fragmented, as IDTs initial capillary electrophoresis showed that it had the expected size of ~1900 bp. DNA shearing/fragmenting can occur by mechanical stress and repeated freeze/thaw cycles. However, all gBlocks were treated equally, and the sample was not subjected to any excessive mechanical stress (vortexing). They were not subjected to repeated freeze thaw cycles, as dilutions were made and used in the PCR reactions. The unamplified gBlock run on the gel was not diluted. The unspecific bands seen in the amplified gBlock2 was proposed to stem from unspecific primer binding or lower level contaminants, so to eliminate unspecific products, IDT recommended gel purification of the correctly sized band.

If PCR amplification of the gBlocks is to be attempted in the future, gBlock1 must firstly be re-synthesized. The thermocycler program should not exceed 10-12 cycles, and after amplification, the correctly sized bands should be excised and gel purified in a column based purification kit. Even gBlock3, which displayed no visible unspecific bands, should be gel extracted, as there is a possibility of unspecific products in a lower concentration that are not detected on the gel.

4.1.4 Linearization of pAAVS1

Two methods were used to linearize the AAVS1 donor vector before assembly with the *FOXA1* construct: Restriction digest with *Hin*dIII and PCR amplification using Q5 polymerase.

*Hin*dIII digestion would cleave the plasmid in two places, cutting out the *puro* cassette, but leaving the donor backbone with the AAVS1 homology arms. A drawback of using *Hin*dIII for linearization is the 3 nt overhangs created upon digestion. These sticky ends may ligate to each other when subjected to DNA ligase, which is a component of the Gibson Assembly mix, and reassemble the original plasmid. This was exactly what happened when Gibson Assembly was attempted using the pAAVS1_*puro* vector. A way to circumvent this problem could be treatment of the digested vector with phosphatase removing the 5' phosphate group, as phosphorylated 5' ends are required for ligase to connect two fragments.

DNA amplified by PCR does not have phosphorylated ends, so the probability of self-ligation is very low. There are, however, other drawbacks to linearization by PCR. For one, PCR amplification always has a certain risk of errors. As the error rate of Q5 polymerase was ~1/1,000,000, the chance of errors was negligible for this purpose. Another drawback is recovery of the template plasmid, which also confers antibiotic resistance. Although template concentration was low (50 pg/µL) and would be further diluted during a Gibson Assembly reaction, there is still a chance that the template plasmid can be taken up and amplified by the bacteria. As the correctly assembled pAAVS1_*FOXA1* plasmid (9599 bp) was quite a lot larger than the original pAAVS1_*puro* plasmid (6602 bp), the replication efficiency would also be higher for the template plasmid. This problem can be circumvented by treating the template plasmid with *Dpn*I restriction enzyme, which only digests the naturally methylated *Dpn*I site, and ignores the PCR amplified plasmid [83]. pAAVS1 was successfully linearized using PCR, but whether the undesired template plasmid would be recovered during transformation of bacteria remains to be seen, as the vector was not used to transform bacterial cells in the course of this thesis.

4.1.5 Problems with Gibson Assembly®

The advantages of using Gibson Assembly for cloning is that multiple fragments principally can be joined together in a one step reaction, without leaving any scars or seams, which are commonly generated when using restriction cloning. Another advantage is that no unwanted sequences have to be added to the junctions for assembly to occur, and in theory any two sequences can be assembled. In practice there are some restrictions regarding sequence composition. Highly repetitive and GC-rich sequences are difficult to assemble because of 3D structures that may form when single stranded DNA is revealed by exonuclease.

In this thesis, overlapping regions of varying lengths were used; the initial long overlaps (99-161 bp) and the short overlaps generated when amplifying the gBlocks with new sets of primers (54-83 bp). Although referred to as short in this thesis, these may also be considered relatively long overlaps. Overlaps of these lengths were used to increase homology between the fragments, in order to increase the chance of assembly. Common overlapping regions in Gibson Assembly are 15-40 bp, and increased size also increase the possibility of 3D structure formation. The extended manual of the NEB master mix kit recommends using overlaps up to 40 bp, and notes that incubation time increases with the size of the overlaps as it will take longer for the exonuclease to chew back the 5' ends.

One of the downsides to Gibson Assembly is that it can be difficult to troubleshoot. Several fragments are being assembled simultaneously, and pinpointing what could have gone wrong can be challenging. In this project, no assembly products were generated, at least not efficiently enough to be visible on a gel. The assembly problems could be caused by one or several of these reasons: i) inhibition of overlap annealing caused by unspecific bands in the

silica bead-purified samples, ii) inhibition of overlap annealing caused by contamination of silica beads, iii) 3D structure formation of the overlapping regions, iv) incomplete exonuclease digestion of overlapping regions, making it impossible for the overlaps to anneal, v) the Gibson Assembly master mix was not functioning properly. Point i and ii can be solved by choosing a DNA purification method using gel extraction and silica columns instead of suspended silica beads, while iii and iv can be solved by using shorter overlapping regions (15-40 bp). There are also tools available to predict formation of 3D structures, their Tm, and the energy required for them to be denatured. This can prove a valuable tool to investigate the probability that the overlaps used in this project formed 3D structures, and also to predict in advance if newly designed, shorter overlaps will form 3D structures. Point v is unlikely, as the positive control yielded rich growth on agar when used to transform competent cells, indicating that the assembly had been efficient.

Another possibility that can reduce Gibson efficiency is the assembly of incompatible intermediate products. For example, when assembling three inserts (1, 2, 3) and a plasmid (P), one can achieve e.g. two assembly products of P(5')-1-2 and 2-3-P(3'). These intermediates would not be possible to join together, and the fragments would be wasted. A way to avoid this is to add the fragments sequentially. The three inserts may be assembled together initially, and after incubation, the plasmid vector can be added to form the finished circular product. Additionally, the molar amounts of the fragments can be regulated. If e.g. fragment 1, 2 and 3 is added in a 1:0.8:1 ratio, there is an increased chance that fragment 2 will assemble with both 1 and 3, and that incompatible intermediates of the inserts (1-2 and 2-3) can be avoided. In this thesis, there were no signs of assembly either by the complete circular products, or intermediates, but upon future attempts at Gibson Assembly, stepwise addition should be attempted.

If these measures should fail to generate the complete plasmid, subcloning of each fragment into the vector is an option. By adding oligonucleotides to the 5' end of primers, different ends can be added to the fragments during PCR, and they can be added to the plasmid vector one by one. In this stepwise approach, the first fragment, gBlock1 will have overlaps complementary to the plasmid flanking both sides, and is assembled with Gibson. PCR amplification can then be used to linearize the assembled vector + gBlock1. gBlock2, with overlaps complementary to the plasmid and gBlock1, may then be added for a second Gibson assembly creating a vector with gBlock1 and 2. The procedure can be repeated with gBlock3 resulting in a vector with all three gBlocks inserted. By subcloning, the specific fragments/overlaps causing the difficulties may be identified.

For future attempts at Gibson Assembly of the pAAVS1_*FOXA1* plasmid, the problems with amplification of gBlocks discussed earlier must first be solved, or the assembly must be done directly with unamplified gBlocks. Overlaps between gBlocks and the pAAVS1 vector should be reduced to 15-40 bp either by redesigning and synthesizing the fragment (gBlock1), or by low-cycle PCR-amplification and gel-extraction of the fragments with new primers that give shorter amplicons (gBlock2, gBlock3 and pAAVS1). The overlaps should also be analyzed with software such as OligoAnalyzer (IDT) predicting 3D structures before attempting the assembly.

4.2 CRISPR-Cas9 plasmid and restriction-ligation cloning of gRNA oligos

4.2.1 CRISPR-Cas9 plasmid delivery

The CRISPR-Cas9 delivery system chosen for this project was by a plasmid vector containing both sgRNA and Cas9n, with Type IIs restriction cloning for insertion of the custom gRNA into the RNA scaffold. The advantage of this delivery system is that combined expression increases the cleavage efficiency. The double nicking strategy was also chosen for this project, using a nickase mutated Cas9 and two sgRNAs that bind in close proximity at opposite strands. Double nicking was used in this project since it strongly reduces off-target effects. As mentioned in chapter **3.2.1**, the other possible delivery systems; Cas9 in the form of mRNA or purified protein, sgRNA as RNA and lentiviral delivery of Cas9 and sgRNA; was not suitable for this project. Delivering Cas9 and sgRNA in the form of RNA and proteins is useful when genetically engineering zygotes to generate transgenic animal disease models. Transcription and translation is suppressed in the zygote [84], and microinjection of *in vitro* transcribed or translated sgRNA and Cas9 before the first cell division can avoid genetic mosaicism of the animal [60]. In this *in vitro* project, antibiotic selection can enrich growth of transfected cells. Lentiviral delivery is useful for genome-wide functional screens, and enables genetic alterations of thousands of elements in parallel [60]. For the purpose of stable knock-in by one expression cassette *in vitro*, only transient transfection with a plasmid vector is necessary.

4.2.2 Design and insertion of gRNAs

The two custom-designed 20 nt gRNA sequences (gRNA1 and gRNA2) were synthesized as single stranded oligos, and annealed and phosphorylated with T4 PNK before ligation into the pSpCas9n vector. Vector-specific 4 nt overhangs were added to the oligos. A restriction-ligation reaction was then set up to cleave the vector and create the overhangs complementary to the gRNA overhangs, and ligate the vector with insert. The reaction was set up with alternating optimal temperatures for *Bbs*I and ligase in six cycles, increasing the cloning efficiency drastically compared to performing a single digestion and ligation. However, the cloning efficiency was not as high expected. Sequencing of the isolated plasmids showed that gRNA2 had been successfully inserted into the pSpCas9n plasmid.

The most probable reason for the low cloning efficiency is that the gRNAs were not phosphorylated properly. T4 PNK requires ATP for efficient phosphorylation, and the T4 PNK buffer provided with the kit does not contain ATP. For increased efficiency, a buffer containing ATP should be used. According to NEB, the T4 ligase buffer (NEB) has the optimal ATP concentration for T4 PNK.

Successful cloning of gRNA2 was achieved, and the remaining steps to enable testing of CRISPR efficiency is to additionally clone gRNA1 into the plasmid using T4 ligase buffer when phospohorylating the 5’ ends, and to sequence verify the clones.

4.3 Transfection of HBECs

Because the repair template construct containing *FOXA1* was not assembled correctly, and gRNA1 was not cloned into the pSpCas9n vector, transfection of HBECs could not be attempted at the course of this thesis. Some work remains in design and assembly of the repair construct, as discussed previously, but a simple cloning procedure can be done with gRNA1. With both gRNAs inserted into respective vectors, the efficiency of the sgRNAs can be tested by transfecting HBECs without the repair template. The formation of NHEJ repair events occurring may then be analyzed. Testing the efficiency of the sgRNAs is practical, because the transfection assay may then be optimized for the specific plasmid and HBECs. Screening of transfected cells may be done by PCR amplifying the target site, denaturing and reannealing the amplicons to create heteroduplexes (mutated amplicon annealed with wild type amplicon), and digestion with T7 or Surveyor endonucleases, which detect and cleave mismatching sequences [85, 86]. Gel electrophoresis of the fragments will then reveal if an in/del mutation has occurred.

If the sgRNA pair is unable to induce DSBs at the target, the sgRNAs may be tested separately, by cloning each gRNA into a wild type Cas9 vector, and transfecting them separately. The mismatch nuclease assays may be used as explained above, and thus efficiency of each sgRNA can be determined. Separate testing is not as feasible though, as this would require using a different vector than the one used in this thesis.

When the *FOXA1* construct is assembled correctly, and can be used as repair template in co-transfection with the CRISPR system, there are a few things that should be considered. NHEJ occurs more readily than HDR, and to achieve repair by the HDR pathway, it is imperative that the repair template is available in the S-phase of the cell cycle. Efficiency of HDR repair can be increased by chemically induced cell cycle synchronization [87], and by inhibition of the NHEJ pathway [88, 89].

5 Conclusion

The two major aims of this thesis was to 1) develop a site-specific knock-in system for insertion of a gene of interest into HBEC cells, and 2) to use this system to integrate the *FOXA1* gene into the AAVS1 safe harbor and investigate its role in EMT. Specific hypotheses regarding *FOXA1's* role in EMT were formulated, but proving or disproving them depended on the success of the first aim.

Still some work remains for successful establishment of the knock-in system. The repair template construct containing *FOXA1* must be redesigned and assembled successfully, and gRNA1 must be cloned into the pSpCas9n. For completion of the *FOXA1* construct, this likely involves a new design, and possible stepwise subcloning of the fragments. gRNA1 on the other hand, should be cloned into its vector, isolated and sequence verified fairly easily. The ability of the sgRNA pair to facilitate a Cas9-mediated double strand break can and should be investigated on its own, without repair template co-transfection. If a double strand break is successfully created, then all that remains to having established the knock-in system is a repair template.

5.1 Future work

To fulfill the second aim of this project, investigating the role of FOXA1 in EMT, the following work remains: i) Co-transfection of HBEC cells with pAAVS1_*FOXA1* and pSpCas9n_gRNAs, ii) Clonal selection of transgenic cells with Hygromycin B enrichment and dilution cloning, iii) Gene expression analysis of *FOXA1* and relevant EMT markers with reverse transcriptase-quantitative PCR, iv) Analysis of colony formation in soft agar assay, v) Analysis of *in vitro* invasive and migratory properties, vi) Protein expression analysis of *FOXA1* and relevant EMT markers with confocal microscopy to study cellular localization and western blot.

These assays can elucidate the role of the FOXA1 factor on EMT marker expression, if it can decrease the number of colonies formed in soft agar, and if it can reduce invasive properties.

Once the CRISPR-mediated AAVS1 site-specific vector system is established, other interesting genes that may be of importance in lung carcinogenicity (i.e. tumor suppressor genes) can be knocked in and ectopically expressed by designing repair template constructs containing the gene of interest. *FOXA2* is another candidate for ectopic expression in the HBEC *in vitro* transformation model. This gene is, like *FOXA1*, downregulated after long time exposure to CSC, and its role in EMT should be investigated.

The background for developing this site-specific knock-in system was to elucidate mechanisms in lung carcinogenicity by restoring expression of genes downregulated in cancer-like HBECs. As the AAVS1 site is initially present in all human cells, the purposes of the knock-in system can be extended beyond lung carcinogenesis. In principle, any type of human *in vitro* model may be knocked in with a gene of interest using this system. Thus, its development sets the stage for many exiting and intriguing future knock-in studies.

References

References

*Credits for the cover photo: Image courtesy of Stephen Dixon and Feng Zhang Reprinted with permission from AAAS. The image relates to a paper that appeared in Jan. 3, 2013, issue of *Science* Express, published by AAAS. [44].

1. Jemal, A., F. Bray, M.M. Center, J. Ferlay, E. Ward, and D. Forman, *Global cancer statistics.* CA Cancer J Clin, 2011. **61**(2): p. 69-90.
2. Zang, E.A. and E.L. Wynder, *Differences in lung cancer risk between men and women: examination of the evidence.* J Natl Cancer Inst, 1996. **88**(3-4): p. 183-92.
3. Mollerup, S., G. Berge, R. Baera, V. Skaug, A. Hewer, D.H. Phillips, L. Stangeland, and A. Haugen, *Sex differences in risk of lung cancer: Expression of genes in the PAH bioactivation pathway in relation to smoking and bulky DNA adducts.* Int J Cancer, 2006. **119**(4): p. 741-4.
4. Hanahan, D. and Robert A. Weinberg, *Hallmarks of Cancer: The Next Generation.* Cell, 2008. **144**(5): p. 646-674.
5. Hecht, S.S., *Lung Carcinogenesis by Tobacco Smoke.* International journal of cancer. Journal international du cancer, 2012. **131**(12): p. 2724-2732.
6. Hecht, S.S., *Tobacco carcinogens, their biomarkers and tobacco-induced cancer.* Nat Rev Cancer, 2003. **3**(10): p. 733-44.
7. Moorthy, B., C. Chu, and D.J. Carlin, *Polycyclic aromatic hydrocarbons: from metabolism to lung cancer.* Toxicol Sci, 2015. **145**(1): p. 5-15.
8. Sagredo, C., S. Mollerup, K.J. Cole, D.H. Phillips, H. Uppstad, and S. Ovrebo, *Biotransformation of benzo[a]pyrene in Ahr knockout mice is dependent on time and route of exposure.* Chem Res Toxicol, 2009. **22**(3): p. 584-91.
9. Shimizu, Y., Y. Nakatsuru, M. Ichinose, Y. Takahashi, H. Kume, J. Mimura, Y. Fujii-Kuriyama, and T. Ishikawa, *Benzo[a]pyrene carcinogenicity is lost in mice lacking the aryl hydrocarbon receptor.* Proc Natl Acad Sci U S A, 2000. **97**(2): p. 779-82.
10. Yoshino, I., T. Kometani, F. Shoji, A. Osoegawa, T. Ohba, H. Kouso, T. Takenaka, T. Yohena, and Y. Maehara, *Induction of epithelial-mesenchymal transition-related genes by benzo[a]pyrene in lung cancer cells.* Cancer, 2007. **110**(2): p. 369-74.
11. Alberts, B., *Molecular biology of the cell: reference edition*. 2008, New York: Garland Science. XXXIII, 1601 s. : ill. + 1 plate (DVD-ROM).
12. Kalluri, R. and R.A. Weinberg, *The basics of epithelial-mesenchymal transition.* The Journal of Clinical Investigation, 2009. **119**(6): p. 1420-1428.
13. Steinestel, K., S. Eder, A.J. Schrader, and J. Steinestel, *Clinical significance of epithelial-mesenchymal transition.* Clin Transl Med, 2014. **3**: p. 17.
14. Mani, S.A., W. Guo, M.J. Liao, E.N. Eaton, A. Ayyanan, A.Y. Zhou, M. Brooks, F. Reinhard, C.C. Zhang, M. Shipitsin, L.L. Campbell, K. Polyak, C. Brisken, J. Yang, and R.A. Weinberg, *The epithelial-mesenchymal transition generates cells with properties of stem cells.* Cell, 2008. **133**(4): p. 704-15.
15. Chui, M.H., *Insights into cancer metastasis from a clinicopathologic perspective: Epithelial-Mesenchymal Transition is not a necessary step.* Int J Cancer, 2013. **132**(7): p. 1487-95.
16. Dave, B., V. Mittal, N.M. Tan, and J.C. Chang, *Epithelial-mesenchymal transition, cancer stem cells and treatment resistance.* Breast Cancer Res, 2012. **14**(1): p. 202.

17. Palena, C., R.I. Fernando, and D.H. Hamilton, *An immunotherapeutic intervention against tumor progression: Targeting a driver of the epithelial-to-mesenchymal transition.* Oncoimmunology, 2014. **3**(1): p. e27220.
18. Hazan, R.B., R. Qiao, R. Keren, I. Badano, and K. Suyama, *Cadherin switch in tumor progression.* Ann N Y Acad Sci, 2004. **1014**: p. 155-63.
19. Li, Z., G. Tuteja, J. Schug, and K.H. Kaestner, *Foxa1 and Foxa2 are essential for sexual dimorphism in liver cancer.* Cell, 2012. **148**(1-2): p. 72-83.
20. Bernardo, G.M. and R.A. Keri, *FOXA1: a transcription factor with parallel functions in development and cancer.* Biosci Rep, 2012. **32**(2): p. 113-30.
21. Wang, H., C.A. Meyer, T. Fei, G. Wang, F. Zhang, and X.S. Liu, *A systematic approach identifies FOXA1 as a key factor in the loss of epithelial traits during the epithelial-to-mesenchymal transition in lung cancer.* BMC Genomics, 2013. **14**: p. 680.
22. Tang, Y., G. Shu, X. Yuan, N. Jing, and J. Song, *FOXA2 functions as a suppressor of tumor metastasis by inhibition of epithelial-to-mesenchymal transition in human lung cancers.* Cell Res, 2011. **21**(2): p. 316-26.
23. Katoh, M., M. Igarashi, H. Fukuda, H. Nakagama, and M. Katoh, *Cancer genetics and genomics of human FOX family genes.* Cancer Lett, 2013. **328**(2): p. 198-206.
24. Caruso, J.A., D.W. Laird, and G. Batist, *Role of HSP90 in mediating cross-talk between the estrogen receptor and the Ah receptor signal transduction pathways.* Biochem Pharmacol, 1999. **58**(9): p. 1395-403.
25. Thomsen, J.S., X. Wang, R.N. Hines, and S. Safe, *Restoration of aryl hydrocarbon (Ah) responsiveness in MDA-MB-231 human breast cancer cells by transient expression of the estrogen receptor.* Carcinogenesis, 1994. **15**(5): p. 933-7.
26. Olivo-Marston, S.E., L.E. Mechanic, S. Mollerup, E.D. Bowman, A.T. Remaley, M.R. Forman, V. Skaug, Y.L. Zheng, A. Haugen, and C.C. Harris, *Serum estrogen and tumor-positive estrogen receptor-alpha are strong prognostic classifiers of non-small-cell lung cancer survival in both men and women.* Carcinogenesis, 2010. **31**(10): p. 1778-86.
27. Uppstad, H., G.H. Osnes, K.J. Cole, D.H. Phillips, A. Haugen, and S. Mollerup, *Sex differences in susceptibility to PAHs is an intrinsic property of human lung adenocarcinoma cells.* Lung Cancer, 2011. **71**(3): p. 264-70.
28. Damiani, L.A., C.M. Yingling, S. Leng, P.E. Romo, J. Nakamura, and S.A. Belinsky, *Carcinogen-induced gene promoter hypermethylation is mediated by DNMT1 and causal for transformation of immortalized bronchial epithelial cells.* Cancer Res, 2008. **68**(21): p. 9005-14.
29. Ramirez, R.D., S. Sheridan, L. Girard, M. Sato, Y. Kim, J. Pollack, M. Peyton, Y. Zou, J.M. Kurie, J.M. Dimaio, S. Milchgrub, A.L. Smith, R.F. Souza, L. Gilbey, X. Zhang, K. Gandia, M.B. Vaughan, W.E. Wright, A.F. Gazdar, J.W. Shay, and J.D. Minna, *Immortalization of human bronchial epithelial cells in the absence of viral oncoproteins.* Cancer Res, 2004. **64**(24): p. 9027-34.
30. Sjøberg, M., *Lung Carcinogenesis - Chemically induced in vitro transformation of human bronchial epithelial cells*, in *Departement of Toxicology*. 2012, University of Oslo: Oslo. p. 99.
31. Bersaas, A., M. Sjøberg, Y. Arnoldussen, A. Haugen, and S. Mollerup, *Manuscript in preparation: FOXA1/2 involved in tobacco-induced EMT during transformation of human bronchial cells*.

32. Hacein-Bey-Abina, S., C. Von Kalle, M. Schmidt, M.P. McCormack, N. Wulffraat, P. Leboulch, A. Lim, C.S. Osborne, R. Pawliuk, E. Morillon, R. Sorensen, A. Forster, P. Fraser, J.I. Cohen, G. de Saint Basile, I. Alexander, U. Wintergerst, T. Frebourg, A. Aurias, D. Stoppa-Lyonnet, S. Romana, I. Radford-Weiss, F. Gross, F. Valensi, E. Delabesse, E. Macintyre, F. Sigaux, J. Soulier, L.E. Leiva, M. Wissler, C. Prinz, T.H. Rabbitts, F. Le Deist, A. Fischer, and M. Cavazzana-Calvo, *LMO2-associated clonal T cell proliferation in two patients after gene therapy for SCID-X1.* Science, 2003. **302**(5644): p. 415-9.
33. Huang, S., Y. Kawabe, A. Ito, and M. Kamihira, *Adeno-associated virus Rep-mediated targeting of integrase-defective retroviral vector DNA circles into human chromosome 19.* Biochem Biophys Res Commun, 2012. **417**(1): p. 78-83.
34. Lombardo, A., D. Cesana, P. Genovese, B. Di Stefano, E. Provasi, D.F. Colombo, M. Neri, Z. Magnani, A. Cantore, P. Lo Riso, M. Damo, O.M. Pello, M.C. Holmes, P.D. Gregory, A. Gritti, V. Broccoli, C. Bonini, and L. Naldini, *Site-specific integration and tailoring of cassette design for sustainable gene transfer.* Nat Methods, 2011. **8**(10): p. 861-9.
35. Sadelain, M., E.P. Papapetrou, and F.D. Bushman, *Safe harbours for the integration of new DNA in the human genome.* Nat Rev Cancer, 2012. **12**(1): p. 51-8.
36. Raitano, S., L. Ordovas, L. De Muynck, W. Guo, I. Espuny-Camacho, M. Geraerts, S. Khurana, K. Vanuytsel, B.I. Toth, T. Voets, R. Vandenberghe, T. Cathomen, L. Van Den Bosch, P. Vanderhaeghen, P. Van Damme, and C.M. Verfaillie, *Restoration of progranulin expression rescues cortical neuron generation in an induced pluripotent stem cell model of frontotemporal dementia.* Stem Cell Reports, 2015. **4**(1): p. 16-24.
37. Merling, R.K., C.L. Sweeney, J. Chu, A. Bodansky, U. Choi, D.L. Priel, D.B. Kuhns, H. Wang, S. Vasilevsky, S.S. De Ravin, T. Winkler, C.E. Dunbar, J. Zou, K.A. Zarember, J.I. Gallin, S.M. Holland, and H.L. Malech, *An AAVS1-targeted minigene platform for correction of iPSCs from all five types of chronic granulomatous disease.* Mol Ther, 2015. **23**(1): p. 147-57.
38. Tiyaboonchai, A., H. Mac, R. Shamsedeen, J.A. Mills, S. Kishore, D.L. French, and P. Gadue, *Utilization of the AAVS1 safe harbor locus for hematopoietic specific transgene expression and gene knockdown in human ES cells.* Stem Cell Res, 2014. **12**(3): p. 630-7.
39. Smith, J.R., S. Maguire, L.A. Davis, M. Alexander, F. Yang, S. Chandran, C. ffrench-Constant, and R.A. Pedersen, *Robust, persistent transgene expression in human embryonic stem cells is achieved with AAVS1-targeted integration.* Stem Cells, 2008. **26**(2): p. 496-504.
40. Wiedenheft, B., S.H. Sternberg, and J.A. Doudna, *RNA-guided genetic silencing systems in bacteria and archaea.* Nature, 2012. **482**(7385): p. 331-338.
41. Mali, P., K.M. Esvelt, and G.M. Church, *Cas9 as a versatile tool for engineering biology.* Nat Meth, 2013. **10**(10): p. 957-963.
42. Deltcheva, E., K. Chylinski, C.M. Sharma, K. Gonzales, Y. Chao, Z.A. Pirzada, M.R. Eckert, J. Vogel, and E. Charpentier, *CRISPR RNA maturation by trans-encoded small RNA and host factor RNase III.* Nature, 2011. **471**(7340): p. 602-7.
43. Sapranauskas, R., G. Gasiunas, C. Fremaux, R. Barrangou, P. Horvath, and V. Siksnys, *The Streptococcus thermophilus CRISPR/Cas system provides immunity in Escherichia coli.* Nucleic Acids Res, 2011. **39**(21): p. 9275-82.
44. Cong, L., F.A. Ran, D. Cox, S. Lin, R. Barretto, N. Habib, P.D. Hsu, X. Wu, W. Jiang, L.A. Marraffini, and F. Zhang, *Multiplex genome engineering using CRISPR/Cas systems.* Science, 2013. **339**(6121): p. 819-23.

45. Sternberg, S.H., S. Redding, M. Jinek, E.C. Greene, and J.A. Doudna, *DNA interrogation by the CRISPR RNA-guided endonuclease Cas9.* Nature, 2014. **507**(7490): p. 62-7.
46. Jinek, M., K. Chylinski, I. Fonfara, M. Hauer, J.A. Doudna, and E. Charpentier, *A programmable dual-RNA-guided DNA endonuclease in adaptive bacterial immunity.* Science, 2012. **337**(6096): p. 816-21.
47. Deveau, H., R. Barrangou, J.E. Garneau, J. Labonte, C. Fremaux, P. Boyaval, D.A. Romero, P. Horvath, and S. Moineau, *Phage response to CRISPR-encoded resistance in Streptococcus thermophilus.* J Bacteriol, 2008. **190**(4): p. 1390-400.
48. Ran, F.A., P.D. Hsu, J. Wright, V. Agarwala, D.A. Scott, and F. Zhang, *Genome engineering using the CRISPR-Cas9 system.* Nat Protoc, 2013. **8**(11): p. 2281-308.
49. DiCarlo, J.E., J.E. Norville, P. Mali, X. Rios, J. Aach, and G.M. Church, *Genome engineering in Saccharomyces cerevisiae using CRISPR-Cas systems.* Nucleic Acids Res, 2013. **41**(7): p. 4336-43.
50. Saleh-Gohari, N. and T. Helleday, *Conservative homologous recombination preferentially repairs DNA double-strand breaks in the S phase of the cell cycle in human cells.* Nucleic Acids Res, 2004. **32**(12): p. 3683-8.
51. Fu, Y., J.A. Foden, C. Khayter, M.L. Maeder, D. Reyon, J.K. Joung, and J.D. Sander, *High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells.* Nat Biotechnol, 2013. **31**(9): p. 822-6.
52. Ran, F.A., Patrick D. Hsu, C.-Y. Lin, Jonathan S. Gootenberg, S. Konermann, A.E. Trevino, David A. Scott, A. Inoue, S. Matoba, Y. Zhang, and F. Zhang, *Double Nicking by RNA-Guided CRISPR Cas9 for Enhanced Genome Editing Specificity.* Cell, 2013. **154**(6): p. 1380-1389.
53. Fu, Y., J.D. Sander, D. Reyon, V.M. Cascio, and J.K. Joung, *Improving CRISPR-Cas nuclease specificity using truncated guide RNAs.* Nature biotechnology, 2014. **32**(3): p. 279-284.
54. Sanchez-Rivera, F.J. and T. Jacks, *Applications of the CRISPR-Cas9 system in cancer biology.* Nat Rev Cancer, 2015. **advance online publication**.
55. Lu, X.-J., X. Qi, D.-H. Zheng, and L.-J. Ji, *Modeling cancer processes with CRISPR-Cas9.* Trends in Biotechnology, 2015. **33**(6): p. 317-319.
56. Drost, J., R.H. van Jaarsveld, B. Ponsioen, C. Zimberlin, R. van Boxtel, A. Buijs, N. Sachs, R.M. Overmeer, G.J. Offerhaus, H. Begthel, J. Korving, M. van de Wetering, G. Schwank, M. Logtenberg, E. Cuppen, H.J. Snippert, J.P. Medema, G.J.P.L. Kops, and H. Clevers, *Sequential cancer mutations in cultured human intestinal stem cells.* Nature, 2015. **521**(7550): p. 43-47.
57. Blasco, Rafael B., E. Karaca, C. Ambrogio, T.-C. Cheong, E. Karayol, Valerio G. Minero, C. Voena, and R. Chiarle, *Simple and Rapid In Vivo Generation of Chromosomal Rearrangements using CRISPR/Cas9 Technology.* Cell Reports, 2014. **9**(4): p. 1219-1227.
58. Park, A., S.T. Won, M. Pentecost, W. Bartkowski, and B. Lee, *CRISPR/Cas9 Allows Efficient and Complete Knock-In of a Destabilization Domain-Tagged Essential Protein in a Human Cell Line, Allowing Rapid Knockdown of Protein Function.* PLoS ONE, 2014. **9**(4): p. e95101.
59. Qi, L.S., M.H. Larson, L.A. Gilbert, J.A. Doudna, J.S. Weissman, A.P. Arkin, and W.A. Lim, *Repurposing CRISPR as an RNA-guided platform for sequence-specific control of gene expression.* Cell, 2013. **152**(5): p. 1173-83.
60. Hsu, P.D., E.S. Lander, and F. Zhang, *Development and applications of CRISPR-Cas9 for genome engineering.* Cell, 2014. **157**(6): p. 1262-78.
61. Tebas, P., D. Stein, W.W. Tang, I. Frank, S.Q. Wang, G. Lee, S.K. Spratt, R.T. Surosky, M.A. Giedlin, G. Nichol, M.C. Holmes, P.D. Gregory, D.G. Ando, M. Kalos, R.G. Collman, G. Binder-

Scholl, G. Plesa, W.T. Hwang, B.L. Levine, and C.H. June, *Gene editing of CCR5 in autologous CD4 T cells of persons infected with HIV.* N Engl J Med, 2014. **370**(10): p. 901-10.

62. Bikard, D., C. Euler, W. Jiang, P.M. Nussenzweig, G.W. Goldberg, X. Duportet, V.A. Fischetti, and L.A. Marraffini, *Development of sequence-specific antimicrobials based on programmable CRISPR-Cas nucleases.* Nature biotechnology, 2014. **32**(11): p. 1146-1150.
63. Alberts, B., *Molecular biology of the cell*. 2008, New York: Garland Science. XXXIII, 1268, [90] s. : ill.
64. Gibson, D.G., L. Young, R.Y. Chuang, J.C. Venter, C.A. Hutchison, 3rd, and H.O. Smith, *Enzymatic assembly of DNA molecules up to several hundred kilobases.* Nat Methods, 2009. **6**(5): p. 343-5.
65. Gibson, D.G., G.A. Benders, C. Andrews-Pfannkoch, E.A. Denisova, H. Baden-Tillson, J. Zaveri, T.B. Stockwell, A. Brownley, D.W. Thomas, M.A. Algire, C. Merryman, L. Young, V.N. Noskov, J.I. Glass, J.C. Venter, C.A. Hutchison, 3rd, and H.O. Smith, *Complete chemical synthesis, assembly, and cloning of a Mycoplasma genitalium genome.* Science, 2008. **319**(5867): p. 1215-20.
66. Villalobos, A., J.E. Ness, C. Gustafsson, J. Minshull, and S. Govindarajan, *Gene Designer: a synthetic biology tool for constructing artificial DNA segments.* BMC Bioinformatics, 2006. **7**: p. 285.
67. Clamp, M., J. Cuff, S.M. Searle, and G.J. Barton, *The Jalview Java alignment editor.* Bioinformatics, 2004. **20**(3): p. 426-7.
68. Waterhouse, A.M., J.B. Procter, D.M. Martin, M. Clamp, and G.J. Barton, *Jalview Version 2--a multiple sequence alignment editor and analysis workbench.* Bioinformatics, 2009. **25**(9): p. 1189-91.
69. Zhang, L. *CRISPR design tool*. 2013; Available from: https://www.dna20.com/products/crispr#4.
70. DNA2.0. *gRNA Design Tool*. Available from: https://www.dna20.com/products/crispr#4.
71. Integrated DNA Technologies, I. *PrimerQuest*. 2012; Available from: http://eu.idtdna.com/PrimerQuest/Home/Index.
72. NCBI. *Primer-BLAST*. 2009; Available from: http://www.ncbi.nlm.nih.gov/tools/primer-blast/.
73. Hockemeyer, D., F. Soldner, C. Beard, Q. Gao, M. Mitalipova, R.C. DeKelver, G.E. Katibah, R. Amora, E.A. Boydston, B. Zeitler, X. Meng, J.C. Miller, L. Zhang, E.J. Rebar, P.D. Gregory, F.D. Urnov, and R. Jaenisch, *Efficient targeting of expressed and silent genes in human ESCs and iPSCs using zinc-finger nucleases.* Nat Biotechnol, 2009. **27**(9): p. 851-7.
74. Jensen, M.A., M. Fukushima, and R.W. Davis, *DMSO and Betaine Greatly Improve Amplification of GC-Rich Constructs in <italic>De Novo</italic> Synthesis.* PLoS ONE, 2010. **5**(6): p. e11024.
75. Genomics, E. *Value Read - Trusted And Reliable DNA Sequencing*. 2015; Available from: http://www.eurofinsgenomics.eu/en/custom-dna-sequencing/single-read-sequencing/value-read.aspx.
76. Qin, J.Y., L. Zhang, K.L. Clift, I. Hulur, A.P. Xiang, B.Z. Ren, and B.T. Lahn, *Systematic comparison of constitutive promoters and the doxycycline-inducible promoter.* PLoS One, 2010. **5**(5): p. e10611.
77. Cormack, B.P., R.H. Valdivia, and S. Falkow, *FACS-optimized mutants of the green fluorescent protein (GFP).* Gene, 1996. **173**(1 Spec No): p. 33-8.

78. Rao, R.N., N.E. Allen, J.N. Hobbs, Jr., W.E. Alborn, Jr., H.A. Kirst, and J.W. Paschal, *Genetic and enzymatic basis of hygromycin B resistance in Escherichia coli.* Antimicrob Agents Chemother, 1983. **24**(5): p. 689-95.
79. Hennecke, M., M. Kwissa, K. Metzger, A. Oumard, A. Kroger, R. Schirmbeck, J. Reimann, and H. Hauser, *Composition and arrangement of genes define the strength of IRES-driven translation in bicistronic mRNAs.* Nucleic Acids Res, 2001. **29**(16): p. 3327-34.
80. Tang, W., I. Ehrlich, S.B. Wolff, A.M. Michalski, S. Wolfl, M.T. Hasan, A. Luthi, and R. Sprengel, *Faithful expression of multiple proteins via 2A-peptide self-processing: a versatile and reliable method for manipulating brain circuits.* J Neurosci, 2009. **29**(27): p. 8621-9.
81. Kim, J.H., S.R. Lee, L.H. Li, H.J. Park, J.H. Park, K.Y. Lee, M.K. Kim, B.A. Shin, and S.Y. Choi, *High cleavage efficiency of a 2A peptide derived from porcine teschovirus-1 in human cell lines, zebrafish and mice.* PLoS One, 2011. **6**(4): p. e18556.
82. Szybalski, W., S.C. Kim, N. Hasan, and A.J. Podhajska, *Class-IIS restriction enzymes--a review.* Gene, 1991. **100**: p. 13-26.
83. Weiner, M.P., G.L. Costa, W. Schoettlin, J. Cline, E. Mathur, and J.C. Bauer, *Site-directed mutagenesis of double-stranded DNA by the polymerase chain reaction.* Gene, 1994. **151**(1-2): p. 119-23.
84. Oh, B., S. Hwang, J. McLaughlin, D. Solter, and B.B. Knowles, *Timely translation during the mouse oocyte-to-embryo transition.* Development, 2000. **127**(17): p. 3795-803.
85. Cho, S.W., S. Kim, Y. Kim, J. Kweon, H.S. Kim, S. Bae, and J.S. Kim, *Analysis of off-target effects of CRISPR/Cas-derived RNA-guided endonucleases and nickases.* Genome Res, 2014. **24**(1): p. 132-41.
86. Qiu, P., H. Shandilya, J.M. D'Alessio, K. O'Connor, J. Durocher, and G.F. Gerard, *Mutation detection using Surveyor nuclease.* Biotechniques, 2004. **36**(4): p. 702-7.
87. Lin, S., B.T. Staahl, R.K. Alla, and J.A. Doudna, *Enhanced homology-directed human genome engineering by controlled timing of CRISPR/Cas9 delivery.* Elife, 2014. **3**: p. e04766.
88. Chu, V.T., T. Weber, B. Wefers, W. Wurst, S. Sander, K. Rajewsky, and R. Kuhn, *Increasing the efficiency of homology-directed repair for CRISPR-Cas9-induced precise gene editing in mammalian cells.* Nat Biotech, 2015. **advance online publication**.
89. Maruyama, T., S.K. Dougan, M.C. Truttmann, A.M. Bilate, J.R. Ingram, and H.L. Ploegh, *Increasing the efficiency of precise genome editing with CRISPR-Cas9 by inhibition of nonhomologous end joining.* Nat Biotechnol, 2015. **33**(5): p. 538-42.

Appendices

A. Materials
B. List of primers
C. List of sequences
D. Sequence reads of pSpCas9n_gRNAs
E. Optimization of PCR gBlock1 and gBlock2

Appendix A
Materials

Enzymes

Enzyme	Company	Buffer
*Hin*dIII-HF	New England Biolabs	CutSmart® Buffer (10X)
*Bbs*I	New England Biolabs	CutSmart® Buffer (10X)
*Pme*I	New England Biolabs	CutSmart® Buffer (10X)

Restriction enzymes

Enzyme	Company	Buffer
Q5® High-Fidelity DNA polymerase	New England Biolabs	Q5 Reaction Buffer (5X)
Taq polymerase	VWR	Master mix
T4 polynucleotide kinase (PNK)	New England Biolabs	T4 PNK reaction buffer (10X)/T4 ligase reaction buffer
T7 DNA Ligase	New England Biolabs	T7 DNA Ligase Reaction Buffer (2X)
T5 Exonuclease	New England Biolabs	NEBuffer 4 (10X)

Bacteria

NEB 5-alpha Competent E. coli (DH5α™ derivative)

Genotype: *fhuA2 Δ(argF-lacZ)U169 phoA glnV44 Φ80 Δ(lacZ)M15 gyrA96 recA1 relA1 endA1 thi-1 hsdR17*

Kits

Gibson Assembly® Cloning Kit, #E5510 (New England Biolabs)

GeneJET Plasmid Midiprep Kit, #K0482 (Thermo Scientific)

Silica Bead DNA Gel Extraction Kit, #K0513 (Thermo Scientific)

Taq DNA Polymerase Master Mix Kit #733-1314 (VWR)

Instruments

Autoclave (Systec DX-90)

Biophotometer – Eppendorf Biophotometer

Haerus® Biofuge® Pico table top centrifuge

Infors HT Ecotron shaker incubator for bacterial cultures

LAF bench (OAS LAF)?

Mini-Sub cell GT cell (BioRad)

MJ Research PTC-220 DYAD Thermal Cycler (2 x 96 wells)

NanoDrop 8000 UV-Vis Spectrometer

Scale – DeltaRange® PG503

Sigma 4K15 Refrigerated Centrifuge

Thermomoxer – Eppendorf Thermomixer comfort

Termax B8000 Bacteriological chamber

Computer Software

Artemis	Sanger Institutes
Codon optimization tool	Integrated DNA Technologies
CRISPR design tool	Zhang Labs (MIT)
CRISPR/Cas9 gRNA tool	DNA 2.0
GeneDesigner	DNA 2.0
Gibson entry tool	Integrated DNA Technologies
Jalview Workbench	Waterhouse et. Al
jEdit Programmer's Editor	jEdit
OligoAnalyzer	Integrated DNA Technologies
PrimerQuest	Integrated DNA Technologies
Primer-BLAST	NCBI
SerialCloner 2.6.1	Frank Perez (SerialBasics)

Chemicals

Agar	Invitrogen
Ampicillin sodium salt	Amresco
DMSO	Sigma

DNA-ladders	New England Biolabs
dNTP mix	New England Biolabs
GelRed	Biotium
Glycerol 99.5 %	Sigma
Kanamycin	Gibco
Propane-2-ol (isopropanol)	Merck
SeaKem GTG agarose	Lonza

Solutions
(all solutions made with ddH_2O and sterile filtered or autoclaved before use)

LB-agar (Merck + Invitrogen)
10 g/L Peptone from casein
5 g/L Yeast extract
10 g/L Sodium Chloride
15 g/L Agar (Invitrogen)

LB-media (Merck)
10 g/L Peptone from casein
5 g/L Yeast extract
10 g/L Sodium Chloride

Loading buffer
3 mL glycerol (30%)
25 mg bromophenol blue (0.25%)
dH_2O to 10mL

TAE (50 X) 1 L solution
252 g Tris base
57.1 mL acetate
100 mL 0.5 M EDTA, pH 8
dH20 to 1 L

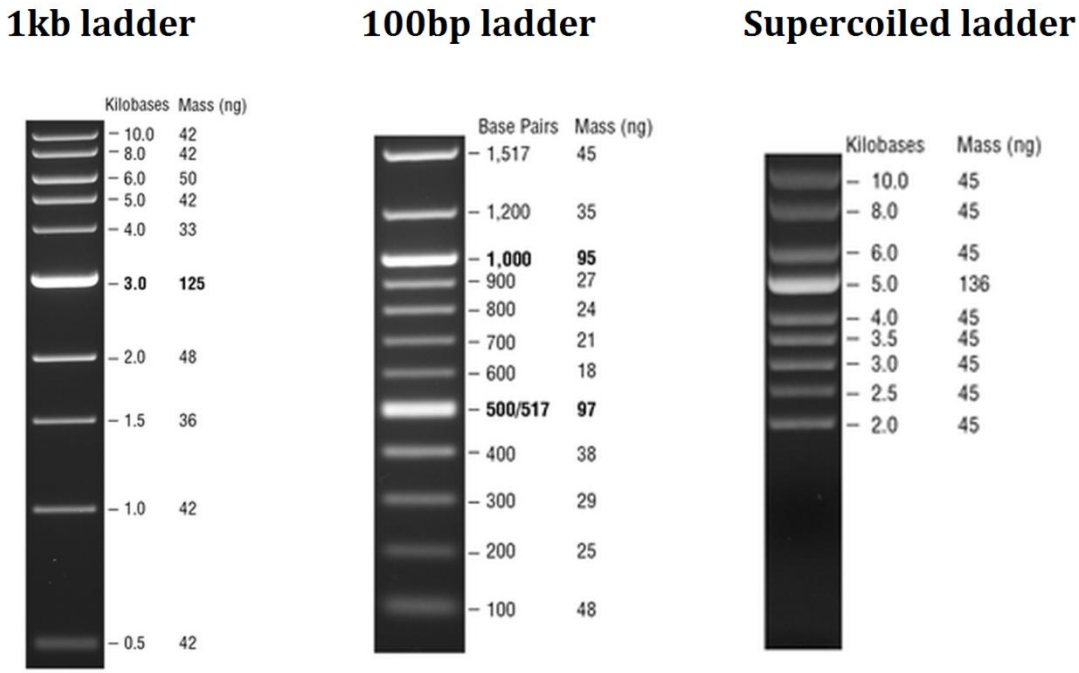
Tris-EDTA (TE, 1X, low EDTA) 30 mL solution:
300 µL Tris-buffer (1 M, pH 8.0) (final conc. 10 mM)
6 µL EDTA (0.5 M, pH 8.0) (final conc. 0.1 mM)
29,69 mL dH_2O

SOC media (NEB)
2% Vegetable Peptone
0.5% Yeast Extract
10 mM NaCl
2.5 mM KCl
10 mM MgCl2

10 mM MgSO4
20 mM Glucose

DNA-ladders

The ladders are all visualized by loading 0.5 µg/lane. 1kb and supercoiled are run on a 0.8 % agarose gel, while 100bp is run on a 1.3 % agarose gel.



Supplementary Figure A Ladders used in this thesis

Appendix B
List of Primers

All sequences are shown in 5’ -> 3’ direction unless otherwise noted

Amplification of gBlocks

AAVS1-gBlock1-fwd	CCACTTCAGGACAGCATGTTT
AAVS1-gBlock1_Fw-New	GAGCTGGGACCACCTTATATTC
AAVS1-gBlock1-rev	CCAGCATGCCTGCTATTGT
AAVS1-gBlock2-fwd	TGCATCGCATTGTCTGAGTAG
AAVS1-gBlock2-rev	CACTTCCGTCCTCGATGTTATG
AAVS1-gBlock2_Fw-New	AGTAGGTGTCATTCTATTCT
AAVS1-gBlock2-Rev-New	TCGGCCATTATATACACATT
AAVS1-gBlock3-fwd	GGGCCACAAGTTGGAGTATAA
AAVS1-gBlock3-rev	GTTCTGGCAAGGAGAGAGATG
AAVS1-gBlock3-rev-New	ACCCAATATCAGGAGACTAGGA

Linearization of pAAVS1

Fw_pAAVS1_linearize	GGACAGGATTGGTGACAGAA
Rev_pAAVS1_linearize	AAAGTACCCAGAACCAGAGC

Conony screening of pAAVS1_*FOXA1*

Fw_HA5’_to_insert	CTTTCTCTGACCAGCATTCTCT
Rev_HA3’_to_insert	TTTCCCAGTCACGACGTT
Fw_pAAVS1_colony_screen	CTCTCCATCCTCTTGCTTTCTT
Rev_pAAVS1_colony_screen	TCCAAACTGCTTCTCCTCTTG

Colony screening of pSpCas9n_gRNA1

gRNA1_Fw	GGGCCTATTTCCCATGATTCCTT
gRNA1_Rev	AAACAAGGATGGGGCTTTTCTGTC

Colony screening of pSpCas9n_gRNA2

gRNA2_Fw	GGACGAAACACCGTCCCTAGT
gRNA2_Rev	CCAAGTGGGCAGTTTACCGT

Sequencing of pAAVS1_*FOXA1*

Seq_Fwd_1	GAGCTGGGACCACCTTATATTC
Seq_Fwd_2	CCGACAGAAGAGGTTCAAAT
Seq_Fwd_3	GGTCGACAGTACTATGCTTTAC
Seq_Fwd_4	AGGAGAATCCTGGACCTATG
Seq_Fwd_5	CTATCAGAGCTTGGTTGACG
Seq_Rev_1	CACTGGTGGTCATTGTGTTC
Seq_Rev_2	GCTGCTCGCTAGATGACATA
Seq_Rev_3	CTCTATCCTATTGACCAGTGTATCC
Seq_Rev_4	GCCATGTAGTGTATTGACCGA
Seq_Rev_5	AGAGATGGCTCCAGGAAATG

Sequencing of pSpCas9n_gRNA1/2

Seq_Fw_sgRNA	TTCTTGGGTAGTTTGCAGTT

Junction PCR of *FOXA1* construct in transfected HBEC cells

Fw_PPP1R12C_to_insert	CTTTGAGCTCTACTGGCTTCTG
Rev_insert_to_HA5’	ATATAGACCTCCCACCGTACAC
Rev_PPP1R12C_to_insert	CACGTAACCTGAGAAGGGAATC
Fw_insert_to_HA3’	CAAAGGAACCATGGCCCAACTT

Appendix C
List of sequences

All sequences are shown in 5’ -> 3’ direction unless otherwise noted

Restriction sites

*Hin*dIII AAGCTT

*Bbs*I GAAGAC

*Pme*I GTTTAAAC

FOXA1 repair template construct

Sequences included in design of repair template construct, where they were obtained, and which modifications were made to the sequences. The sequences are listed in the order they appear in the construct in 5’->3’ direction.

Sequence	Origin	Modifications
CMV promoter	AddGene	
5’ UTR	Gene Designer	
FOXA1	NCBI (NM_004496)	Codon optimized for gBlocks
bGHpA terminator	AddGene	
SV40 promoter	AddGene	
eGFP	AddGene	Codon optimized for gBlocks
SV40 nls	Ran *et al*. [48]	
T2A	AddGene	
Hygro B	AddGene	
SV40pA terminator	AddGene	
AAVS1 Homology arms	AddGene	Silent point mutations of PAM and gRNA spacer in overlaps between gBlock1 and 3 that overlaps with homology arms

>pAAVS1_SA-2A-puro-pA donor (pAAVS1_puro)

agcgcccaatacgcaaaccgcctctccccgcgcgttggccgattcattaatgcagctggcacgacaggtttcccgactgg
aaagcgggcagtgagcgcaacgcaattaatgtgagttagctcactcattaggcaccccaggctttacactttatgcttcc
ggctcgtatgttgtgtggaattgtgagcggataacaatttcacacaggaaacagctatgaccatgattacgccaagctca
gaattaaccctcactaaagggactagtcctgcaggtttaaacgaattcgcccttgctttctctgaccagcattctctcc
cctgggcctgtgccgctttctgtctgcagcttgtggcctgggtcacctctacggctggcccagatccttccctgccgcct
ccttcaggttccgtcttcctccactccctcttcccccttgctctctgctgtgttgctgcccaaggatgctctttccggagc
acttccttctcggcgctgcaccacgtgatgtcctctgagcggatcctccccgtgtctgggtcctctccgggcatctctcc
tccctcacccaaccccatgccgtcttcactcgctgggttccctttttccttctccttctggggcctgtgccatctctcgtt

tcttaggatggccttctccgacggatgtctcccttgcgtcccgcctcccccttcttgtaggcctgcatcatcaccgttttt
ctggacaaccccaaagtaccccgtctccctggctttagccacctctccatcctcttgctttcttgcctggacaccccgt
tctcctgtggattcgggtcacctctcactcctttcatttgggcagctcccctaccccccttacctctctagtctgtgcta
gctcttccagcccccctgtcatggcatcttccaggggtccgagagctcagctagtcttcttcctccaaccccgggcccctat
gtccacttcaggacagcatgtttgctgcctccagggatcctgtgtccccgagctgggaccaccttatattcccagggccg
gttaatgtggctctggttctgggtacttttatctgtcccctccaccccacagtggggcaagcttctgacctcttctcttc
ctcccacagggcctcgagagatctggcagcggagagggcagaggaagtcttctaacatgcggtgacgtggaggagaatcc
cggccctaggctcgagatgaccgagtacaagcccacggtgcgcctcgccacccgcgacgacgtccccagggccgtacgca
ccctcgccgccgcgttcgccgactaccccgccacgcgccacaccgtcgatccggaccgccacatcgagcgggtcaccgag
ctgcaagaactcttcctcacgcgcgtcgggctcgacatcggcaaggtgtgggtcgcggacgacggcgccgcggtggcggt
ctggaccacgccggagagcgtcgaagcgggggcggtgttcgccgagatcggcccgcgcatggccgagttgagcggttccc
ggctggccgcgcagcaacagatggaaggcctcctggcgccgcaccggcccaaggagcccgcgtggttcctggccaccgtc
ggcgtctcgcccgaccaccagggcaagggtctgggcagcgccgtcgtgctccccggagtggaggcggccgagcgcgccgg
ggtgcccgccttcctggagacctccgcgccccgcaacctccccttctacgagcggctcggcttcaccgtcaccgccgacg
tcgaggtgcccgaaggaccgcgcacctggtgcatgacccgcaagcccggtgcctgatctagagggcccgtttaaacccgc
tgatcagcctcgactgtgccttctagttgccagccatctgttgtttgcccctcccccgtgccttccttgaccctggaagg
tgccactcccactgtcctttcctaataaaatgaggaaattgcatcgcattgtctgagtaggtgtcattctattctggggg
gtggggtggggcaggacagcaagggggaggattgggaagacaatagcaggcatgctggggatgcggtgggctctatgggt
cgacagtactaagcttactagggacaggattggtgacagaaaagccccatccttaggcctcctccttcctagtctcctg
atattgggtctaacccccacctcctgttaggcagattccttatctggtgacacacccccatttcctggagccatctctct
ccttgccagaacctctaaggtttgcttacgatggagccagagaggatcctgggagggagagcttggcaggggggtgggagg
gaaggggggggatgcgtgacctgcccggttctcagtggccaccctgcgctaccctctcccagaacctgagctgctctgacg
cggctgtctggtgcgtttcactgatcctggtgctgcagcttccttacacttcccaagaggagaagcagtttggaaaaaca
aaatcagaataagttggtcctgagttctaactttggctcttcacctttctagtccccaatttatattgttcctccgtgcg
tcagttttacctgtgagataaggccagtagccagccccgtcctggcagggctgtggtgaggaggggggtgtccgtgtgga
aaactccctttgtgagaatggtgcgtcctaggtgttcaccaggtcgtggccgcctctactccctttctctttctccatcc
ttctttccttaaagagtccccagtgctatctgggacatattcctccgcccagagcagggtcccgcttccctaaggccctg
ctctgggcttctgggtttgagtccttggcaagcccaggagaggcgctcaggcttccctgtccccttcctcgtccaccat
ctcatgcccctggctctcctgccccttccctacaggggttcctggctctgctctaagggcaagggcgaattcgcggccgc
taaattcaattcgccctatagtgagtcgtattacaattcactggccgtcgttttacaacgtcgtgactgggaaaaccctg
gcgttacccaacttaatcgccttgcagcacatccccctttcgccagctggcgtaatagcgaagaggcccgcaccgatcgc
ccttcccaacagttgcgcagcctatacgtacggcagtttaaggtttacacctataaaagagagagccgttatcgtctgtt
tgtggatgtacagagtgatattattgacacgccggggcgacggatggtgatccccctggccagtgcacgtctgctgtcag
ataaagtctcccgtgaactttacccggtggtgcatatcggggatgaaagctggcgcatgatgaccaccgatatggccagt
gtgccggtctccgttatcggggaagaagtggctgatctcagccaccgcgaaaatgacatcaaaaacgccattaacctgat
gttctggggaatataaatgtcaggcatgagattatcaaaaaggatcttcacctagatccttttcacgtagaaagccagtc
cgcagaaacggtgctgaccccggatgaatgtcagctactgggctatctggacaagggaaaacgcaagcgcaaagagaaag
caggtagcttgcagtgggcttacatggcgatagctagactgggcggttttatggacagcaagcgaaccggaattgccagc
tggggcgccctctggtaaggttgggaagccctgcaaagtaaactggatggctttcttgccgccaaggatctgatggcgca
ggggatcaagctctgatcaagagacaggatgaggatcgtttcgcatgattgaacaagatggattgcacgcaggttctccg
gccgcttgggtggagaggctattcggctatgactgggcacaacagacaatcggctgctctgatgccgccgtgttccggct
gtcagcgcaggggcgcccggttctttttgtcaagaccgacctgtccggtgccctgaatgaactgcaagacgaggcagcgc
ggctatcgtggctggccacgacgggcgttccttgcgcagctgtgctcgacgttgtcactgaagcgggaagggactggctg
ctattgggcgaagtgccggggcaggatctcctgtcatctcaccttgctcctgccgagaaagtatccatcatggctgatgc
aatgcggcggctgcatacgcttgatccggctacctgcccattcgaccaccaagcgaaacatcgcatcgagcgagcacgta
ctcggatggaagccggtcttgtcgatcaggatgatctggacgaagagcatcaggggctcgcgccagccgaactgttcgcc
aggctcaaggcgagcatgcccgacggcgaggatctcgtcgtgacccatggcgatgcctgcttgccgaatatcatggtgga
aaatggccgcttttctggattcatcgactgtggccggctgggtgtggcggaccgctatcaggacatagcgttggctaccc
gtgatattgctgaagagcttggcggcgaatgggctgaccgcttcctcgtgctttacggtatcgccgctcccgattcgcag
cgcatcgccttctatcgccttcttgacgagttcttctgaattattaacgcttacaatttcctgatgcggtattttctcct
tacgcatctgtgcggtatttcacaccgcatcaggtggcacttttcggggaaatgtgcgcggaacccctatttgtttattt
ttctaaatacattcaaatatgtatccgctcatgagattatcaaaaaggatcttcacctagatccttttaaattaaaaatg
aagttttaaatcaatctaaagtatatatgagtaaacttggtctgacagttaccaatgcttaatcagtgaggcacctatct
cagcgatctgtctatttcgttcatccatagttgcctgactccccgtcgtgtagataactacgatacgggagggcttacca
tctggccccagtgctgcaatgataccgcgagacccacgctcaccggctccagatttatcagcaataaaccagccagccgg
aagggccgagcgcagaagtggtcctgcaactttatccgcctccatccagtctattaattgttgccgggaagctagagtaa
gtagttcgccagttaatagtttgcgcaacgttgttgccattgctacaggcatcgtggtgtcacgctcgtcgtttggtatg
gcttcattcagctccggttcccaacgatcaaggcgagttacatgatcccccatgttgtgcaaaaaagcggttagctcctt
cggtcctccgatcgttgtcagaagtaagttggccgcagtgttatcactcatggttatggcagcactgcataattctctta

ctgtcatgccatccgtaagatgcttttctgtgactggtgagtactcaaccaagtcattctgagaatagtgtatgcggcga
ccgagttgctcttgcccggcgtcaatacgggataataccgcgccacatagcagaactttaaaagtgctcatcattggaaa
acgttcttcggggcgaaaactctcaaggatcttaccgctgttgagatccagttcgatgtaacccactcgtgcacccaact
gatcttcagcatcttttactttcaccagcgtttctgggtgagcaaaaacaggaaggcaaaatgccgcaaaaaagggaata
agggcgacacggaaatgttgaatactcatactcttccttttttcaatattattgaagcatttatcagggttattgtctcat
gaccaaaatcccttaacgtgagttttcgttccactgagcgtcagaccccgtagaaaagatcaaaggatcttcttgagatc
ctttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcggtggtttgtttgccggatcaagag
ctaccaactctttttccgaaggtaactggcttcagcagagcgcagataccaaatactgttcttctagtgtagccgtagtt
aggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcctgttaccagtggctgctgccagtg
gcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcgcagcggtcgggctgaacgggggt
tcgtgcacacagcccagcttggagcgaacgacctacaccgaactgagatacctacagcgtgagctatgagaaagcgccac
gcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacaggagagcgcacgagggagcttccag
ggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgagcgtcgatttttgtgatgctcgtca
ggggggcggagcctatggaaaaacgccagcaacgcggccttttacggttcctggccttttgctggccttttgctcacat
gttctttcctgcgttatcccctgattctgtggataaccgtattaccgcctttgagtgagctgataccgctcgccgcagcc
gaacgaccgagcgcagcgagtcagtgagcgaggaagcggaag

*Bbs*I restriction site

Primers for colony screening

Homology arm 5’

Primers for linearization of pAAVS1

*Hin*dIII restriction site

SA-2A-*puro*-pA-casette

Homology arm 3’

Kan/Neo^{r}

Amp^{r}

> pAAVS1_FOXA1_eGFP_T2A_HygroB (pAAVS1_FOXA1)

AAACGAATTCGCCCTTTGCTTTCTCTGACCAGCATTCTCTCCCCTGGGCCTGTGCCGCTTTCTGTCTGCAGCTTGTGGCC
TGGGTCACCTCTACGGCTGGCCCAGATCCTTCCCTGCCGCCTCCTTCAGGTTCCGTCTTCCTCCACTCCCTCTTCCCCTT
GCTCTCTGCTGTGTTGCTGCCCAAGGATGCTCTTTCCGGAGCACTTCCTTCTCGGCGCTGCACCACGTGATGTCCTCTGA
GCGGATCCTCCCCGTGTCTGGGTCCTCTCCGGGCATCTCTCCTCCCTCACCCAACCCCATGCCGTCTTCACTCGCTGGGT
TCCCTTTTCCTTCTCCTTCTGGGGCCTGTGCCATCTCTCGTTTCTTAGGATGGCCTTCTCCGACGGATGTCTCCCTTGCG
TCCCGCCTCCCCTTCTTGTAGGCCTGCATCATCACCGTTTTTCTGGACAACCCCAAAGTACCCCGTCTCCCTGGCTTTAG
CCACCTCTCCATCCTCTTGCTTTCTTTGCCTGGACACCCCGTTCTCCTGTGGATTCGGGTCACCTCTCACTCCTTTCATT
TGGGCAGCTCCCCTACCCCCCTTACCTCTCTAGTCTGTGCTAGCTCTTCCAGCCCCCTGTCATGGCATCTTCCAGGGGTC
CGAGAGCTCAGCTAGTCTTCTTCCTCCAACCCGGGCCCCTATGTCCACTTCAGGACAGCATGTTTGCTGCCTCCAGGGAT
CCTGTGTCCCCGAGCTGGGACCACCTTATATTCCCAGGGCCGGTTAATGTGGCTCTGGTTCTGGGTACTTTTATCTGTCC
CCTCCACCCCACAGTGGGGCAAGCTTAATTAACTGACCCGCCCCGTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGA
GGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGAATACAAGCTACTTGTTCTTTTTGCACATGCTCGGTACA
GTCAAAATGGAGGGCCACGAGACTTCAGATTGGAATTCCTATTACGCTGACACACAGGAGGCTTACTCAAGTGTGCCTGT
GAGTAATATGAATTCTGGCCTCGGATCAATGAACAGCATGAACACTTATATGACCATGAACACAATGACCACCAGTGGCA
ATATGACTCCCGCTAGCTTCAATATGTCATATGCAAATCCAGGACTCGGTGCCGGCCTTAGTCCTGGAGCTGTGGCTGGC
ATGCCAGGAGGCAGTGCCGGAGCCATGAACTCTATGACCGCCGCAGGTGTGACAGCGATGGGAACCGCCCTTTCACCCTC
CGGGATGGGAGCTATGGGCGCTCAGCAGGCCGCCTCCATGAATGGACTGGGTCCGTACGCCGCCGCCATGAATCCCTGCA
TGAGCCCTATGGCTTATGCACCAAGTAATCTGGGCCGGTCCAGAGCAGGTGGCGGCGGGGATGCTAAGACTTTTAAACGC

TCATATCCTCATGCTAAGCCGCCATACTCTTACATTTCACTCATTACCATGGCTATACAGCAGGCTCCTTCAAAGATGCT
GACACTTTCTGAAATATACCAGTGGATTATGGATCTGTTCCCCTATTACCGGCAAAACCAGCAGCGGTGGCAGAATTCAA
TACGACACAGCCTGAGCTTTAATGACTGTTTTGTGAAGGTGGCTCGATCTCCCGATAAGCCTGGAAAAGGATCCTATTGG
ACTTTGCACCCTGACTCCGGCAACATGTTCGAGAATGGATGCTACCTTCGCCGACAGAAGAGGTTCAAATGCGAAAAACA
GCCAGGCGCAGGTGGCGGCGGAGGTTCTGGTTCCGGGGGGTCTGGAGCCAAAGGCGGCCCTGAAAGTAGGAAAGATCCAA
GCGGGGCCAGTAATCCGAGCGCTGACTCACCGCTCCACAGAGGCGTCCACGGCAAAACAGGTCAGCTCGAAGGCGCCCCA
GCTCCCGGTCCCGCGGCAAGCCCTCAGACTCTGGATCACAGTGGGGCCACGGCTACTGGTGGCGCCTCAGAACTGAAAAC
CCCAGCTAGCAGCACCGCTCCTCCCATCTCTTCCGGGCCTGGCGCACTTGCCAGCGTGCCCGCTAGCCATCCAGCCCACG
GCCTCGCGCCCCATGAATCTCAACTCCACCTCAAGGGAGATCCCCACTATAGCTTTAACCACCCCTTCTCTATCAACAAT
CTTATGTCATCTAGCGAGCAGCAGCACAAACTGGATTTCAAGGCATACGAACAGGCACTTCAGTACAGCCCTTACGGAAG
CACCCTCCCTGCCAGTCTGCCTCTGGGGTCAGCGAGCGTCACCACGCGGTCACCGATTGAACCCTCCGCACTGGAACCCG
CCTACTACCAGGGTGTCTATTCTCGGCCCGTGCTTAATACATCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTT
GCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCG
CATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAG
CAGGCATGCTGGGGATGCGGTGGGCTCTATGGGTCGACAGTACTATGCTTTACTGCATCTCAATTAGTCAGCAACCATAG
TCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACTAATTTTT
TTTATTTATGCAGAGGCCGAGGCCGCCTCGGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAGGCCTAGG
CTTTTGCAAAAAGCTTATGGTGTCTAAGGGAGAAGAATTGTTCACTGGGGTAGTGCCAATCCTGGTGGAGCTGGATGGTG
ATGTGAACGGGCACAAATTTTCCGTTTCTGGAGAAGGGGAAGGCGACGCCACCTACGGAAAACTCACACTCAAATTTATT
TGCACAACTGGAAAGCTGCCGGTCCCGTGGCCCACACTCGTGACAACATTGACATATGGCGTTCAGTGCTTCAGCCGGTA
CCCTGATCATATGAAACAGCACGATTTCTTTAAGTCCGCAATGCCAGAAGGCTACGTGCAGGAGCGCACTATATTTTTTA
AGGATGATGGCAATTATAAAACCCGGGCCGAAGTCAAATTCGAAGGGGATACACTGGTCAATAGGATAGAGTTGAAAGGC
ATCGATTTTAAGGAAGACGGGAATATCCTGGGCCACAAGTTTGGAGTATAACTACAACAGCCACAATGTGTATATAATGGC
CGATAAACAAAAGAATGGCATTAAGGTCAACTTCAAAATTAGACATAACATCGAGGACGGAAGTGTCCAGCTTGCTGACC
ACTACCAGCAGAATACTCCCATAGGTGACGGACCTGTTCTGCTCCCCGATAATCACTACCTGTCAACCCAGTCAGCGCTG
AGCAAGGATCCTAACGAAAAGCGAGATCATATGGTGCTGCTCGAGTTCGTTACCGCCGCAGGCATTACTTTGGGGATGGA
TGAGTTGTATAAGTGAATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGGAAGCGGAGAGG
GCAGAGGAAGTCTGCTAACATGCGGTGACGTCGAGGAGAATCCTGGACCTATGGATAGATCCGGAAAGCCTGAACTCACC
GCGACGTCTGTCGAGAAGTTTCTGATCGAAAAGTTCGACAGCGTCTCCGACCTGATGCAGCTCTCGGAGGGCGAAGAATC
TCGTGCTTTCAGCTTCGATGTAGGAGGGCGTGGATATGTCCTGCGGGTAAATAGCTGCGCCGATGGTTTCTACAAAGATC
GTTATGTTTATCGGCACTTTGCATCGGCCGCGCTCCCGATTCCGGAAGTGCTTGACATTGGGGAATTCAGCGAGAGCCTG
ACCTATTGCATCTCCCGCCGTGCACAGGGTGTCACGTTGCAAGACCTGCCTGAAACCGAACTGCCCGCTGTTCTGCAGCC
GGTCGCGGAGGCCATGGATGCGATCGCTGCGGCCGATCTTAGCCAGACGAGCGGGTTCGGCCCATTCGGACCGCAAGGAA
TCGGTCAATACACTACATGGCGTGATTTCATATGCGCGATTGCTGATCCCCATGTGTATCACTGGCAAACTGTGATGGAC
GACACCGTCAGTGCGTCCGTCGCGCAGGCTCTCGATGAGCTGATGCTTTGGGCCGAGGACTGCCCCGAAGTCCGGCACCT
CGTGCACGCGGATTTCGGCTCCAACAATGTCCTGACGGACAATGGCCGCATAACAGCGGTCATTGACTGGAGCGAGGCGA
TGTTCGGGGATTCCCAATACGAGGTCGCCAACATCTTCTTCTGGAGGCCGTGGTTGGCTTGTATGGAGCAGCAGACGCGC
TACTTCGAGCGGAGGCATCCGGAGCTTGCAGGATCGCCGCGGCTCCGGGCGTATATGCTCCGCATTGGTCTTGACCAACT
CTATCAGAGCTTGGTTGACGGCAATTTCGATGATGCAGCTTGGGCGCAGGGTCGATGCGACGCAATCGTCCGATCCGGAG
CCGGGACTGTCGGGCGTACACAAATCGCCCGCAGAAGCGCGGCCGTCTGGACCGATGGCTGTGTAGAAGTACTCGCCGAT
AGTGGAAACCGACGCCCCAGCACTCGTCCGAGGGCAAAGGAACCATGGCCCAACTTGTTTATTGCAGCTTATAATGGTTA
CAAATAAAGCAATAGCATCACAAATTTCACAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCA
TCAATGTATCTTATCATGTCTGGATCTCGTACTAAGCTTTACTAGGGACAGGATTGGTGACAGAAAAGCCCCATCCTTAG
GCCTCCTCCTTCCTAGTCTCCTGATATTGGGTCTAACCCCCACCTCCTGTTAGGCAGATTCCTTATCTGGTGACACACCC
CCATTTCCTGGAGCCATCTCTCTCCTTGCCAGAACCTCTAAGGTTTGCTTACGATGGAGCCAGAGAGGATCCTGGGAGGG
AGAGCTTGGCAGGGGGTGGGAGGGAAGGGGGGGATGCGTGACCTGCCCGGTTCTCAGTGGCCACCCTGCGCTACCCTCTC
CCAGAACCTGAGCTGCTCTGACGCGGCTGTCTGGTGCGTTTCACTGATCCTGGTGCTGCAGCTTCCTTACACTTCCCAAG
AGGAGAAGCAGTTTGGAAAAACAAAATCAGAATAAGTTGGTCCTGAGTTCTAACTTTGGCTCTTCACCTTTCTAGTCCCC
AATTTATATTGTTCCTCCGTGCGTCAGTTTTACCTGTGAGATAAGGCCAGTAGCCAGCCCCGTCCTGGCAGGGCTGTGGT
GAGGAGGGGGGTGTCCGTGTGGAAAACTCCCTTTGTGAGAATGGTGCGTCCTAGGTGTTCACCAGGTCGTGGCCGCCTCT
ACTCCCTTTCTCTTTCTCCATCCTTCTTTCCTTAAAGAGTCCCCAGTGCTATCTGGGACATATTCCTCCGCCCAGAGCAG
GGTCCCGCTTCCCTAAGGCCCTGCTCTGGGCTTCTGGGTTTGAGTCCTTGGCAAGCCCAGGAGAGGCGCTCAGGCTTCCC
TGTCCCCCTTCCTCGTCCACCATCTCATGCCCCTGGCTCTCCTGCCCCTTCCCTACAGGGGTTCCTGGCTCTGCTCTAAG
GGCAAGGGCGAATTCGCGGCCGCTAAATTCAATTCGCCCTATAGTGAGTCGTATTACAATTCACTGGCCGTCGTTTTACA
ACGTCGTGACTGGGAAAACCCTGGCGTTACCCAACTTAATCGCCTTGCAGCACATCCCCCTTTCGCCAGCTGGCGTAATA
GCGAAGAGGCCCGCACCGATCGCCCTTCCCAACAGTTGCGCAGCCTATACGTACGGCAGTTTAAGGTTTACACCTATAAA
AGAGAGAGCCGTTATCGTCTGTTTGTGGATGTACAGAGTGATATTATTGACACGCCGGGGCGACGGATGGTGATCCCCCT
GGCCAGTGCACGTCTGCTGTCAGATAAAGTCTCCCGTGAACTTTACCCGGTGGTGCATATCGGGGATGAAAGCTGGCGCA
TGATGACCACCGATATGGCCAGTGTGCCGGTCTCCGTTATCGGGGAAGAAGTGGCTGATCTCAGCCACCGCGAAAATGAC

ATCAAAAACGCCATTAACCTGATGTTCTGGGGAATATAAATGTCAGGCATGAGATTATCAAAAAGGATCTTCACCTAGAT
CCTTTTCACGTAGAAAGCCAGTCCGCAGAAACGGTGCTGACCCCGGATGAATGTCAGCTACTGGGCTATCTGGACAAGGG
AAAACGCAAGCGCAAAGAGAAAGCAGGTAGCTTGCAGTGGGCTTACATGGCGATAGCTAGACTGGGCGGTTTTATGGACA
GCAAGCGAACCGGAATTGCCAGCTGGGGCGCCCTCTGGTAAGGTTGGGAAGCCCTGCAAAGTAAACTGGATGGCTTTCTT
GCCGCCAAGGATCTGATGGCGCAGGGGATCAAGCTCTGATCAAGAGACAGGATGAGGATCGTTTCGCATGATTGAACAAG
ATGGATTGCACGCAGGTTCTCCGGCCGCTTGGGTGGAGAGGCTATTCGGCTATGACTGGGCACAACAGACAATCGGCTGC
TCTGATGCCGCCGTGTTCCGGCTGTCAGCGCAGGGGCGCCCGGTTCTTTTTGTCAAGACCGACCTGTCCGGTGCCCTGAA
TGAACTGCAAGACGAGGCAGCGCGGCTATCGTGGCTGGCCACGACGGGCGTTCCTTGCGCAGCTGTGCTCGACGTTGTCA
CTGAAGCGGGAAGGGACTGGCTGCTATTGGGCGAAGTGCCGGGGCAGGATCTCCTGTCATCTCACCTTGCTCCTGCCGAG
AAAGTATCCATCATGGCTGATGCAATGCGGCGGCTGCATACGCTTGATCCGGCTACCTGCCCATTCGACCACCAAGCGAA
ACATCGCATCGAGCGAGCACGTACTCGGATGGAAGCCGGTCTTGTCGATCAGGATGATCTGGACGAAGAGCATCAGGGGC
TCGCGCCAGCCGAACTGTTCGCCAGGCTCAAGGCGAGCATGCCCGACGGCGAGGATCTCGTCGTGACCCATGGCGATGCC
TGCTTGCCGAATATCATGGTGGAAAATGGCCGCTTTTCTGGATTCATCGACTGTGGCCGGCTGGGTGTGGCGGACCGCTA
TCAGGACATAGCGTTGGCTACCCGTGATATTGCTGAAGAGCTTGGCGGCGAATGGGCTGACCGCTTCCTCGTGCTTTACG
GTATCGCCGCTCCCGATTCGCAGCGCATCGCCTTCTATCGCCTTCTTGACGAGTTCTTCTGAATTATTAACGCTTACAAT
TTCCTGATGCGGTATTTTCTCCTTACGCATCTGTGCGGTATTTCACACCGCATCAGGTGGCACTTTTCGGGGAAATGTGC
GCGGAACCCCTATTTGTTTATTTTTCTAAATACATTCAAATATGTATCCGCTCATGAGATTATCAAAAAGGATCTTCACC
TAGATCCTTTTAAATTAAAAATGAAGTTTTAAATCAATCTAAAGTATATATGAGTAAACTTGGTCTGACAGTTACAATGC
TTAATCAGTGAGGCACCTATCTCAGCGATCTGTCTATTTCGTTCATCCATAGTTGCCTGACTCCCGTCGTGTAGATAACT
ACGATACGGGAGGCTTACCATCTGGCCCCAGTGCTGCAATGATACCGCGAGACCCACGCTCACCGGCTCCAGATTTATCA
GCAATAAACCAGCCAGCCGGAAGGGCCGAGCGCAGAAGTGGTCCTGCAACTTTATCCGCCTCCATCAGTCTATTAATTGT
TGCCGGGAAGCTAGAGTAAGTAGTTCGCCAGTTAATAGTTTGCGCAACGTTGTTGCATTGCTCAGGCATCGTGGTGTCAC
CTCGTCGTTTGGTATGGCTTCATTCAGCTCCGGTTCCCAACGATCAAGGCGAGTTACATGATCCCCCATGTTGTGAAAAA
AGCGGTTAGCTCCTTCGGTCCTCCGATCGTTGTCAGAAGTAAGTTGGCCGCAGTGTTATCACTCATGTTATGGCAGCACT
GCATAATTCTCTTACTGTCATGCCATCCGTAAGATGCTTTTCTGTGACTGGTGAGTACTCAACCAAGTCATTCTGAGAAT
AGCGGCGACCGAGTTGCTCTTGCCCGCGTCAAACGGGATAATAGCATAGCAGAACTTTAAAAGTGCTCATCATTGGAAAA
CGTTCTTCGGGGCGAAAACTCTCAAGGATCTTACCGCTGTTGAGATCCAGTTCGATGTAACCCACTCGTGCACCCAACTG
ATCTTCAGCATCTTTTACTTTCACCAGCGTTTCTGGGTGAGCAAAAACAGGAAGGCAAAATGCCGCAAAAAAGGGAATAA
GGGCGACACGGAAATGTTGAATACTCATACTCTTCCTTTTTCAATATTATTGAAGCATTTATCAGGGTTATTGTCTCATG
ACCAAAATCCCTTAACGTGAGTTTTCGTTCCACTGAGCGTCAGACCCCGTAGAAAAGATCAAAGGATCTTCTTGAGATCC
TTTTTTTCTGCGCGTAATCTGCTGCTTGCAAACAAAAAAACCACCGCTACCAGCGGTGGTTTGTTTGCCGGATCAAGAGC
TACCAACTCTTTTTCCGAAGGTAACTGGCTTCAGCAGAGCGCAGATACCAAATACTGTTCTTCTAGTGTAGCCGTAGTTA
GGCCACCACTTCAAGAACTCTGTAGCACCGCCTACATACCTCGCTCTGCTAATCCTGTTACCAGTGGCTGCTGCCAGTGG
CGATAAGTCGTGTCTTACCGGGTTGGACTCAAGACGATAGTTACCGGATAAGGCGCAGCGGTCGGGCTGAACGGGGGGTT
CGTGCACACAGCCCAGCTTGGAGCGAACGACCTACACCGAACTGAGATACCTACAGCGTGAGCTATGAGAAAGCGCCACG
CTTCCCGAAGGGAGAAAGGCGGACAGGTATCCGGTAAGCGGCAGGGTCGGAACAGGAGAGCGCACGAGGGAGCTTCCAGG
GGGAAACGCCTGGTATCTTTATAGTCCTGTCGGGTTTCGCCACCTCTGACTTGAGCGTCGATTTTTGTGATGCTCGTCAG
GGGGGCGGAGCCTATGGAAAAACGCCAGCAACGCGGCCTTTTTACGGTTCCTGGCCTTTTGCTGGCCTTTTGCTCACATG
TTCTTTCCTGCGTTATCCCCTGATTCTGTGGATAACCGTATTACCGCCTTTGAGTGAGCTGATACCGCTCGCCGCAGCCG
AACGACCGAGCGCAGCGAGTCAGTGAGCGAGGAAGCGGAAGAGCGCCCAATACGCAAACCGCCTCTCCCCGCGCGTTGGC
CGATTCATTAATGCAGCTGGCACGACAGGTTTCCCGACTGGAAAGCGGGCAGTGAGCGCAACGCAATTAATGTGAGTTAG
CTCACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGCTCGTATGTTGTGTGGAATTGTGAGCGGATAACAATT
TCACACAGGAAACAGCTATGACCATGATTACGCCAAGCTCAGAATTAACCCTCACTAAAGGGACTAGTCCTGCAGGTTT

*Bbs*I restriction site (cut)

Primers for colony screening

Homology arm 5’

Primers for long overlapping regions

Primers for short overlapping regions

*Hin*dIII restriction site

Start codon

Appendices

CMV promoter

*FOXA1*_optimized

bGH pA

eGFP_optimized

NLS

T2A

HygroB

SV40pA

Homology arm 3'

Kan/Neo^{r}

Amp^{r}

>gBlock1

CCACTTCAGGACAGCATGTTTGCTGCCTCCAGGGATCCTGTGTCCCCGAGCTGGGACCACCTTATAATTCCCAGGGCCGGT
TAATGTGGCTCTGGTTCTGGGTACTTTTATCTGTCCCCTCCACaCCACAGTGGGGCAAGCTTAATTAACTGACCCGCCCC
GTTGACGCAAATGGGCGGTAGGCGTGTACGGTGGGAGGTCTATATAAGCAGAGCTCGTTTAGTGAACCGTCAGATCGAAT
ACAAGCTACTTGTTCTTTTTGCACATGCTCGGTACAGTCAAAATGGAGGGCCACGAGACTTCAGATTGGAATTCCTATTA
CGCTGACACACAGGAGGCTTACTCAAGTGTGCCTGTGAGTAATATGAATTCTGGCCTCGGATCAATGAACAGCATGAACA
CTTATATGACCATGAACACAATGACCACCAGTGGCAATATGACTCCCGCTAGCTTCAATATGTCATATGCAAATCCAGGA
CTCGGTGCCGGCCTTAGTCCTGGAGCTGTGGCTGGCATGCCAGGAGGCAGTGCCGGAGCCATGAACTCTATGACCGCCGC
AGGTGTGACAGCGATGGGAACCGCCCTTTCACCCTCCGGGATGGGAGCTATGGGCGCTCAGCAGGCCGCCTCCATGAATG
GACTGGGTCCGTACGCCGCCGCCATGAATCCCTGCATGAGCCCTATGGCTTATGCACCAAGTAATCTGGGCCGGTCCAGA
GCAGGTGGCGGCGGGGATGCTAAGACTTTTAAACGCTCATATCCTCATGCTAAGCCGCCATACTCTTACATTTCACTCAT
TACCATGGCTATACAGCAGGCTCCTTCAAAGATGCTGACACTTTCTGAAATATACCAGTGGATTATGGATCTGTTCCCCT
ATTACCGGCAAAACCAGCAGCGGTGGCAGAATTCAATACGACACAGCCTGAGCTTTAATGACTGTTTTGTGAAGGTGGCT
CGATCTCCCGATAAGCCTGGAAAAGGATCCTATTGGACTTTGCACCCTGACTCCGGCAACATGTTCGAGAATGGATGCTA
CCTTCGCCGACAGAAGAGGTTCAAATGCGAAAAACAGCCAGGCGCAGGTGGCGGCGGAGGTTCTGGTTCCGGGGGGTCTG
GAGCCAAAGGCGGCCCTGAAAGTAGGAAAGATCCAAGCGGGGCCAGTAATCCGAGCGCTGACTCACCGCTCCACAGAGGC
GTCCACGGCAAAACAGGTCAGCTCGAAGGCGCCCCAGCTCCCGGTCCCGCGGCAAGCCCTCAGACTCTGGATCACAGTGG
GGCCACGGCTACTGGTGGCGCCTCAGAACTGAAAACCCCAGCTAGCAGCACCGCTCCTCCCATCTCTTCCGGGCCTGGCG
CACTTGCCAGCGTGCCCGCTAGCCATCCAGCCCACGGCCTCGCGCCCCATGAATCTCAACTCCACCTCAAGGGAGATCCC
CACTATAGCTTTAACCACCCCTTCTCTATCAACAATCTTATGTCATCTAGCGAGCAGCAGCACAAACTGGATTTCAAGGC
ATACGAACAGGCACTTCAGTACAGCCCTTACGGAAGCACCCTCCCTGCCAGTCTGCCTCTGGGGTCAGCGAGCGTCACCA
CGCGGTCACCGATTGAACCCTCCGCACTGGAACCCGCCTACTACCAGGGTGTCTATTCTCGGCCCGTGCTTAATACATCT
CGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGCCTTCCTTGACCCTGGAAGGTGCCACTCCC
ACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGG
GCAGGACAGCAAGGGGGAGGATTGGGAAGACAATAGCAGGCATGCTGG

>gBlock2

TGCATCGCATTGTCTGAGTAGGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAG
ACAATAGCAGGCATGCTGGGGATGCGGTGGGCTCTATGGGTCGACAGTACTATGCTTTACTGCATCTCAATTAGTCAGCA
ACCATAGTCCCGCCCCTAACTCCGCCCATCCCGCCCCTAACTCCGCCCAGTTCCGCCCATTCTCCGCCCCATGGCTGACT
AATTTTTTTTATTTATGCAGAGGCCGAGGCCGCCTCGGCCTCTGAGCTATTCCAGAAGTAGTGAGGAGGCTTTTTTGGAG
GCCTAGGCTTTTGCAAAAAGCTTATGGTGTCTAAGGGAGAAGAATTGTTCACTGGGGTAGTGCCAATCCTGGTGGAGCTG
GATGGTGATGTGAACGGGCACAAATTTTCCGTTTCTGGAGAAGGGGAAGGCGACGCCACCTACGGAAAACTCACACTCAA
ATTTATTTGCACAACTGGAAAGCTGCCGGTCCCGTGGCCCACACTCGTGACAACATTGACATATGGCGTTCAGTGCTTCA
GCCGGTACCCTGATCATATGAAACAGCACGATTTCTTTAAGTCCGCAATGCCAGAAGGCTACGTGCAGGAGCGCACTATA
TTTTTTAAGGATGATGGCAATTATAAAACCCGGGCCGAAGTCAAATTCGAAGGGGATACACTGGTCAATAGGATAGAGTT
GAAAGGCATCGATTTTAAGGAAGACGGGAATATCCTGGGCCACAAGTTGGAGTATAACTACAACAGCCACAATGTGTATA
TAATGGCCGTTAAACAAAAGAATGGCATTAAGGTCAACTTCAAAATTAGACATAACATCGAGGACGGAAGTG

>gBlock3

GGGCCACAAGTTGGAGTATAACTACAACAGCCACAATGTGTATATAATGGCCGATAAACAAAAGAATGGCATTAAGGTCA
ACTTCAAAATTAGACATAACATCGAGGACGGAAGTGTCCAGCTTGCTGACCACTACCAGCAGAATACTCCCATAGGTGAC
GGACCTGTTCTGCTCCCCGATAATCACTACCTGTCAACCCAGTCAGCGCTGAGCAAGGATCCTAACGAAAAGCGAGATCA
TATGGTGCTGCTCGAGTTCGTTACCGCCGCAGGCATTACTTTGGGGATGGATGAGTTGTATAAGTGAATGGCCCCAAAGA
AGAAGCGGAAGGTCGGTATCCACGGAGTCCCAGCAGCCGGAAGCGGAGAGGGCAGAGGAAGTCTGCTAACATGCGGTGAC
GTCGAGGAGAATCCTGGACCTATGGATAGATCCGGAAAGCCTGAACTCACCGCGACGTCTGTCGAGAAGTTTCTGATCGA
AAAGTTCGACAGCGTCTCCGACCTGATGCAGCTCTCGGAGGGCGAAGAATCTCGTGCTTTCAGCTTCGATGTAGGAGGGC
GTGGATATGTCCTGCGGGTAAATAGCTGCGCCGATGGTTTCTACAAAGATCGTTATGTTTATCGGCACTTTGCATCGGCC
GCGCTCCCGATTCCGGAAGTGCTTGACATTGGGGAATTCAGCGAGAGCCTGACCTATTGCATCTCCCGCCGTGCACAGGG
TGTCACGTTGCAAGACCTGCCTGAAACCGAACTGCCCGCTGTTCTGCAGCCGGTCGCGGAGGCCATGGATGCGATCGCTG
CGGCCGATCTTAGCCAGACGAGCGGGTTCGGCCCATTCGGACCGCAAGGAATCGGTCAATACACTACATGGCGTGATTTC
ATATGCGCGATTGCTGATCCCCATGTGTATCACTGGCAAACTGTGATGGACGACACCGTCAGTGCGTCCGTCGCGCAGGC
TCTCGATGAGCTGATGCTTTGGGCCGAGGACTGCCCCGAAGTCCGGCACCTCGTGCACGCGGATTTCGGCTCCAACAATG
TCCTGACGGACAATGGCCGCATAACAGCGGTCATTGACTGGAGCGAGGCGATGTTCGGGGATTCCCAATACGAGGTCGCC
AACATCTTCTTCTGGAGGCCGTGGTTGGCTTGTATGGAGCAGCAGACGCGCTACTTCGAGCGGAGGCATCCGGAGCTTGC
AGGATCGCCGCGGCTCCGGGCGTATATGCTCCGCATTGGTCTTGACCAACTCTATCAGAGCTTGGTTGACGGCAATTTCG
ATGATGCAGCTTGGGCGCAGGGTCGATGCGACGCAATCGTCCGATCCGGAGCCGGGACTGTCGGGCGTACACAAATCGCC
CGCAGAAGCGCGGCCGTCTGGACCGATGGCTGTGTAGAAGTACTCGCCGATAGTGGAAACCGACGCCCCAGCACTCGTCC
GAGGGCAAAGGAACCATGGCCCAACTTGTTTATTGCAGCTTATAATGGTTACAAATAAAGCAATAGCATCACAAATTTCA
CAAATAAAGCATTTTTTTCACTGCATTCTAGTTGTGGTTTGTCCAAACTCATCAATGTATCTTATCATGTCTGGATCTCG
TACTAAGCTTTACTAGGGACAGGATTGGTGACAGAAAAGCCCCATCaTTAGtCCTCCTCCTTCCTAGTCTCCTGATATTG
GGTCTAACCCCCACCTCCTGTTAGGCAGATTCCTTATCTGGTGACACACCCCCATTTCCTGGAGCCATCTCTCTCCTTGC
CAGAAC

Long overlapping regions (red + green)

Short overlapping regions

Primers for long overlapping regions

Primers for short overlapping regions

Primers for PCR linearization of pAAVS1

*Hin*dIII restriction site

CRISPR-design

gRNA oligos before ligation

AAVS1-gRNA-fwd	CACCGACAGAAAAGCCCCATCCTT
AAVS1-gRNA-fwd-c	AAACAAGGATGGGGCTTTTCTGTC
AAVS1-gRNA-rev	CACCGTCCCTAGTGGCCCCACTGT
AAVS1-gRNA-rev-c	AAACACAGTGGGGCCACTAGGGAC

*Bbs*I overhang

gRNA oligos after ligation

AAVS1-sgRNA-fwd (gRNA1)

```
5′  CACCGACAGAAAAGCCCCATCCTT     3′
3′     CTGTCTTTTCGGGGTAGGAACAAA 5′
```

AAVS1-sgRNA-rev (gRNA2)

```
5′  CACCGTCCCTAGTGGCCCCACTGT     3′
3′     CAGGGATCACCGGGGTGACACAAA 5′
```

>pSpCas9n (BB)

```
gagggcctatttcccatgattccttcatatttgcatatacgatacaaggctgttagagagataattggaattaatttgac
tgtaaacacaaagatattagtacaaaatacgtgacgtagaaagtaataattcttgggtagtttgcagttctaaaattat
gttttaaaatggactatcatatgcttaccgtaacttgaaagtatttcgatttcttggctttatatatcttGTGGAAAGGA
CGAAACACCggGTCTTCgaGAAGACctgttttagagctaGAAAtagcaagttaaaataaggctagtccgttatcaacttg
aaaaagtggcaccgagtcggtgcTTTTTTgttttagagctagaaatagcaagttaaaataaggctagtccgtTTTTagcg
cgtgcgccaattctgcagacaaatggctctagaggtacccgttacataacttacggtaaatggcccgcctggctgaccgc
ccaacgacccccgcccattgacgtcaatagtaacgccaatagggactttccattgacgtcaatgggtggagtatttacgg
taaactgcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgtcaatgacggtaaatggcc
cgcctggcattGtgcccagtacatgaccttatgggactttcctacttggcagtacatctacgtattagtcatcgctatta
ccatggtcgaggtgagccccacgttctgcttcactctcccatctccccccctccccacccccaattttgtatttatttt
atttttaattattttgtgcagcgatgggggcgggggggggggggggcgcgccaggcggggcggggcggggcgaggg
gcggggcggggcgaggcggagaggtgcggcggcagccaatcagagcggcgcgctccgaaagtttccttttatggcgaggc
ggcggcggcggcggcccctataaaaagcgaagcgcgcggcgggcgggagtcgctgcgacgctgccttcgccccgtgccccg
ctccgccgccgcctcgcgccgcccgccccggctctgactgaccgcgttactcccacaggtgagcgggcgggacggcccctt
ctcctccgggctgtaattagctgagcaagaggtaagggtttaagggatggttggttggtggggtattaatgtttaattac
ctggagcacctgcctgaaatcacttttttcaggttGGaccggtgccaccATGGACTATAAGGACCACGACGGAGACTAC
```

AAGGATCATGATATTGATTACAAAGACGATGACGATAAGATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGAGT
CCCAGCAGCCGACAAGAAGTACAGCATCGGCCTGGcCATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGACGAGT
ACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCCTG
CTGTTCGACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAGAACCG
GATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTCCT
TCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCACGAGAAG
TACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCCCT
GGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGACGTGGACAAGCTGT
TCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCATC
CTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTGTT
CGGcAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAACTGC
AGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGACCTGTTTCTG
GCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCCCCCCTGAG
CGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCCTG
AGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAGAG
TTCTACAAGTTCATCAAGCCCATCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGACCT
GCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCGGC
GGCAGGAAGATTTTTACCCATTCCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACTAC
GTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAACTT
CGAGGAAGTGGTGGACAAGGGCGCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAACG
AGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACGTGACC
GAGGGAATGAGAAAGCCCGCCTTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGGAA
AGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGAAG
ATCGGTTCAACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGGAA
AACGAGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAAAC
CTATGCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCGGA
AGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACAGA
AACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGCGA
TAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAAGGTGGTGG
ACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCCAG
AAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAAGA
ACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGTGG
ACCAGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGAGCTTTCTGAAGGACGACTCC
ATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAGAA
GATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAATCTGACCAAGGCCGAGA
GAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACGTG
GCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGATCACCCT
GAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTACAAAGTGCGCGAGATCAACAACTACCACCACGCCC
ACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACGGC
GACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTCTT
CTACAGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGAGA
CAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCCAA
GTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGATAA
GCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCTGG
TGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAAGA
AGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAGCT
GCCTAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAACG
AACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAGGAT
AATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCCAA
GAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGAGC
AGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCATC
GACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAGAC
ACGGATCGACCTGTCTCAGCTGGGAGGCGACAAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAAGt
aagaattcCTAGAGCTCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGTGC
CTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTAGG
TGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGAgAATAGCAGGCATGCTGGGGA
gcggccgcaggaacccctagtgatggagttggccactccctctctgcgcgctcgctcgctcactgaggccgggcgaccaa
aggtcgcccgacgcccgggctttgcccgggcggcctcagtgagcgagcgagcgcgcagctgcctgcaggggcgcctgatg
cggtattttctccttacgcatctgtgcggtatttcacaccgcatacgtcaaagcaaccatagtacgcgccctgtagcggc

gcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttcgc
tttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcgggggctccctttagggttccgat
ttagtgctttacggcacctcgaccccaaaaaacttgatttgggtgatggttcacgtagtgggccatcgccctgatagacg
gtttttcgccctttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccctat
ctcgggctattcttttgatttataagggattttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaaat
ttaacgcgaattttaacaaaatattaacgtttacaattttatggtgcactctcagtacaatctgctctgatgccgcatag
ttaagccagccccgacacccgccaacacccgctgacgcgccctgacgggcttgtctgctcccggcatccgcttacagaca
agctgtgaccgtctccgggagctgcatgtgtcagaggttttcaccgtcatcaccgaaacgcgcgagacgaaagggcctcg
tgatacgcctatttttataggttaatgtcatgataataatggtttcttagacgtcaggtggcacttttcggggaaatgtg
cgcggaacccctatttgtttatttttctaaatacattcaaatatgtatccgctcatgagacaataaccctgataaatgct
tcaataatattgaaaaaggaagagtatgagtattcaacatttccgtgtcgcccttattccctttttgcggcattttgcc
ttcctgtttttgctcacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttacatc
gaactggatctcaacagcggtaagatccttgagagttttcgccccgaagaacgttttccaatgatgagcacttttaaagt
tctgctatgtggcgcggtattatcccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctcagaatg
acttggttgagtactcaccagtcacagaaaagcatcttacggatggcatgacagtaagagaattatgcagtgctgccata
accatgagtgataacactgcggccaacttacttctgacaacgatcggaggaccgaaggagctaaccgcttttttgcacaa
catgggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgagcgtgacacca
cgatgcctgtagcaatggcaacaacgttgcgcaaactattaactggcgaactacttactctagcttcccggcaacaatta
atagactggatggaggcggataaagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctgataa
atctggagccggtgagcgtggaagccgcggtatcattgcagcactggggccagatggtaagccctcccgtatcgtagtta
tctacacgacggggagtcaggcaactatggatgaacgaaatagacagatcgctgagataggtgcctcactgattaagcat
tggtaactgtcagaccaagtttactcatatatactttagattgatttaaaacttcatttttaatttaaaaggatctaggt
gaagatccttttttgataatctcatgaccaaaatcccttaacgtgagttttcgttccactgagcgtcagaccccgtagaaa
agatcaaaggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccagcg
gtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagataccaaatac
tgtccttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaatcc
tgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataaggcg
cagcggtcgggctgaacgggggttcgtgcacacagcccagcttggagcgaacgacctacaccgaactgagatacctaca
gcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaacag
gagagcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttgag
cgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggccttttacggttcctggc
cttttgctggccttttgctcacatgt

U6 promoter

gRNA sequencing primer

*Bbs*I restriction sites

RNA scaffold

Terminator

CAG-enhancer

3xFLAG-tag

SV40-NLS

SpCas9n

D10A nickase mutation

bGH-poly-A

Amp^r promoter

Amp^r

>pSpCas9n_gRNA1

gagggcctatttcccatgattccttcatatttgcatatacgatacaaggctgttagagagataattggaattaatttgac
tgtaaacacaaagatattagtacaaaatacgtgacgtagaaagtaataatttcttgggtagtttgcagttctaaaattat
gttttaaaatggactatcatatgcttaccgtaacttgaaagtatttcgatttcttggctttatatatcttGTGGAAAGGA
CGAAACACCGACAGAAAAGCCCCATCCTTgttttagagctaGAAAtagcaagttaaaataaggctagtccgttatcaact
tgaaaaagtggcaccgagtcggtgcTTTTTTgttttagagctagaaatagcaagttaaaataaggctagtccgtTTTTag
cgcgtgcgccaattctgcagacaaatggctctagaggtacccgttacataacttacggtaaatggcccgcctggctgacc
gcccaacgacccccgcccattgacgtcaatagtaacgccaatagggactttccattgacgtcaatgggtggagtatttac
ggtaaactgcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgtcaatgacggtaaatgg
cccgcctggcattGtgcccagtacatgaccttatgggactttcctacttggcagtacatctacgtattagtcatcgctat
taccatggtcgaggtgagccccacgttctgcttcactctcccatctccccccctcccccacccccaattttgtatttat
ttatttttttaattattttgtgcagcgatgggggcgggggggggggggggcgcgcgccaggcggggcggggcggggcgag
gggcggggcggggcgaggcggagaggtgcggcggcagccaatcagagcggcgcgctccgaaagtttccttttatggcgag
gcggcggcggcggcggcccctataaaaagcgaagcgcgcggcgggcgggagtcgctgcgacgctgccttcgccccgtgccc
cgctccgccgccgcctcgcgccgcccgccccggctctgactgaccgcgttactcccacaggtgagcgggcgggacggccc
ttctcctccgggctgtaattagctgagcaagaggtaagggtttaagggatggttggttggtggggtattaatgtttaatt
acctggagcacctgcctgaaatcacttttttttcaggttGGaccggtgccaccATGGACTATAAGGACCACGACGGAGACT
ACAAGGATCATGATATTGATTACAAAGACGATGACGATAAGATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGA
GTCCCAGCAGCCGACAAGAAGTACAGCATCGGCCTGGcCATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGACGA
GTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCC
TGCTGTTCGACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAGAAC
CGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTC
CTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCACGAGA
AGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCC
CTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGACGTGGACAAGCT
GTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCA
TCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTG
TTCGGcAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAACT
GCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGACCTGTTTC
TGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCCCCCCTG
AGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCC
TGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAG
AGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGAC
CTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCG
GCGGCAGGAAGATTTTTACCCATTCCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACT
ACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAAC
TTCGAGGAAGTGGTGGACAAGGGCGCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAA
CGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACGTGA
CCGAGGGAATGAGAAAGCCCGCCTTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGG
AAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGA
AGATCGGTTCAACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGG
AAAACGAGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAA
ACCTATGCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCG
GAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACA
GAAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGC
GATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAAGGTGGT
GGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCC
AGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAA
GAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGT
GGACCAGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGAGCTTTCTGAAGGACGACT
CCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAG
AAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAATCTGACCAAGGCCGA
GAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACG
TGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGATCACC
CTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTACAAAGTGCGCGAGATCAACAACTACCACCACGC
CCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACG
GCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTC
TTCTACAGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGA

GACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCC
AAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGAT
AAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCT
GGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAA
GAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAG
CTGCCTAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAA
CGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAGG
ATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCC
AAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGA
GCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCA
TCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAG
ACACGGATCGACCTGTCTCAGCTGGGAGGCGACAAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAA
GtaagaattcCTAGAGCTCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGT
GCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTA
GGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGAgAATAGCAGGCATGCTGGG
Gagcggccgcaggaacccctagtgatggagttggccactccctctctgcgcgctcgctcgctcactgaggccgggcgacc
aaaggtcgcccgacgcccgggctttgcccgggcggcctcagtgagcgagcgagcgcgcagctgcctgcaggggcgcctga
tgcggtattttctccttacgcatctgtgcggtatttcacaccgcatacgtcaaagcaaccatagtacgcgccctgtagcg
gcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttc
gctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcgggggctccctttagggttccg
atttagtgctttacggcacctcgaccccaaaaaacttgatttgggtgatggttcacgtagtgggccatcgccctgataga
cggtttttcgccctttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccct
atctcgggctattcttttgatttataagggattttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaa
atttaacgcgaattttaacaaaatattaacgtttacaattttatggtgcactctcagtacaatctgctctgatgccgcat
agttaagccagccccgacacccgccaacacccgctgacgcgccctgacgggcttgtctgctcccggcatccgcttacaga
caagctgtgaccgtctccgggagctgcatgtgtcagaggttttcaccgtcatcaccgaaacgcgcgagacgaaagggcct
cgtgatacgcctatttttataggttaatgtcatgataataatggtttcttagacgtcaggtggcacttttcggggaaatg
tgcgcggaacccctatttgtttatttttctaaatacattcaaatatgtatccgctcatgagacaataaccctgataaatg
cttcaataatattgaaaaaggaagagtatgagtattcaacatttccgtgtcgcccttattccctttttttgcggcattttg
ccttcctgtttttgctcacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttaca
tcgaactggatctcaacagcggtaagatccttgagagttttcgccccgaagaacgttttccaatgatgagcacttttaaa
gttctgctatgtggcgcggtattatcccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctcagaa
tgacttggttgagtactcaccagtcacagaaaagcatcttacggatggcatgacagtaagagaattatgcagtgctgcca
taaccatgagtgataacactgcggccaacttacttctgacaacgatcggaggaccgaaggagctaaccgcttttttgcac
aacatgggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgagcgtgacac
cacgatgcctgtagcaatggcaacaacgttgcgcaaactattaactggcgaactacttactctagcttcccggcaacaat
taatagactggatggaggcggataaagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctgat
aaatctggagccggtgagcgtggaagccgcggtatcattgcagcactggggccagatggtaagccctcccgtatcgtagt
tatctacacgacggggagtcaggcaactatggatgaacgaaatagacagatcgctgagataggtgcctcactgattaagc
attggtaactgtcagaccaagtttactcatatatactttagattgatttaaaacttcatttttaatttaaaaggatctag
gtgaagatccttttttgataatctcatgaccaaaatcccttaacgtgagttttcgttccactgagcgtcagaccccgtaga
aaagatcaaaggatcttcttgagatcctttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccag
cggtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagataccaaat
actgtccttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaat
cctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataagg
cgcagcggtcgggctgaacgggggttcgtgcacacagcccagcttggagcgaacgacctacaccgaactgagataccta
cagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaac
aggagagcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttg
agcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggccttttacggttcctg
gccttttgctggccttttgctcacatgt

U6 promoter

gRNA sequencing primer

gRNA1_Fw_conoly screen

gRNA1

gRNA1_Rev_colony_screen

3xFLAG-tag

SV40-NLS

SpCas9n

D10A nickase mutation

Amp^r promoter

Amp^r

>pSpCas9n_gRNA2

gagggcctatttcccatgattccttcatatttgcatatacgatacaaggctgttagagagataattggaattaatttgac
tgtaaacacaaagatattagtacaaaatacgtgacgtagaaagtaataattttcttgggtagtttgcagttttaaaattat
gttttaaaatggactatcatatgcttaccgtaacttgaaagtatttcgatttcttggctttatatatcttGTGGAAAGGA
CGAAACACCGTCCCTAGTGGCCCCACTGTgttttagagctaGAAAtagcaagttaaaataaggctagtccgttatcaact
tgaaaaagtggcaccgagtcggtgcTTTTTTgttttagagctagaaatagcaagttaaaataaggctagtccgtTTTTag
cgcgtgcgccaattctgcagacaaatggctctagaggtacccgttacataacttacggtaaatggcccgcctggctgacc
gcccaacgacccccgcccattgacgtcaatagtaacgccaatagggactttccattgacgtcaatgggtggagtatttac
ggtaaactgcccacttggcagtacatcaagtgtatcatatgccaagtacgccccctattgacgtcaatgacggtaaatgg
cccgcctggcattGtgcccagtacatgaccttatgggactttcctacttggcagtacatctacgtattagtcatcgctat
taccatggtcgaggtgagccccacgttctgcttcactctcccatctccccccctcccacccccaattttgtatttat
ttatttttaattattttgtgcagcgatgggggcgggggggggggggggcgcgcgccaggcggggcggggcggggcgag
gggcggggcggggcgaggcggagaggtgcggcggcagccaatcagagcggcgcgctccgaaagtttccttttatggcgag
gcggcggcggcggcggcccctataaaaagcgaagcgcgcggcgggcgggagtcgctgcgacgctgccttcgccccgtgccc
cgctccgccgccgcctcgcgccgcccgccccggctctgactgaccgcgttactcccacaggtgagcgggcgggacggccc
ttctcctccgggctgtaattagctgagcaagaggtaagggtttaagggatggttggttggtggggtattaatgtttaatt
acctggagcacctgcctgaaatcacttttttcaggttGGaccggtgccaccATGGACTATAAGGACCACGACGGAGACT
ACAAGGATCATGATATTGATTACAAAGACGATGACGATAAGATGGCCCCAAAGAAGAAGCGGAAGGTCGGTATCCACGGA
GTCCCAGCAGCCGACAAGAAGTACAGCATCGGCCTGGcCATCGGCACCAACTCTGTGGGCTGGGCCGTGATCACCGACGA
GTACAAGGTGCCCAGCAAGAAATTCAAGGTGCTGGGCAACACCGACCGGCACAGCATCAAGAAGAACCTGATCGGAGCCC
TGCTGTTCGACAGCGGCGAAACAGCCGAGGCCACCCGGCTGAAGAGAACCGCCAGAAGAAGATACACCAGACGGAAGAAC
CGGATCTGCTATCTGCAAGAGATCTTCAGCAACGAGATGGCCAAGGTGGACGACAGCTTCTTCCACAGACTGGAAGAGTC
CTTCCTGGTGGAAGAGGATAAGAAGCACGAGCGGCACCCCATCTTCGGCAACATCGTGGACGAGGTGGCCTACCACGAGA
AGTACCCCACCATCTACCACCTGAGAAAGAAACTGGTGGACAGCACCGACAAGGCCGACCTGCGGCTGATCTATCTGGCC
CTGGCCCACATGATCAAGTTCCGGGGCCACTTCCTGATCGAGGGCGACCTGAACCCCGACAACAGCGACGTGGACAAGCT
GTTCATCCAGCTGGTGCAGACCTACAACCAGCTGTTCGAGGAAAACCCCATCAACGCCAGCGGCGTGGACGCCAAGGCCA
TCCTGTCTGCCAGACTGAGCAAGAGCAGACGGCTGGAAAATCTGATCGCCCAGCTGCCCGGCGAGAAGAAGAATGGCCTG
TTCGGcAACCTGATTGCCCTGAGCCTGGGCCTGACCCCCAACTTCAAGAGCAACTTCGACCTGGCCGAGGATGCCAAACT
GCAGCTGAGCAAGGACACCTACGACGACGACCTGGACAACCTGCTGGCCCAGATCGGCGACCAGTACGCCGACCTGTTTC
TGGCCGCCAAGAACCTGTCCGACGCCATCCTGCTGAGCGACATCCTGAGAGTGAACACCGAGATCACCAAGGCCCCCCTG
AGCGCCTCTATGATCAAGAGATACGACGAGCACCACCAGGACCTGACCCTGCTGAAAGCTCTCGTGCGGCAGCAGCTGCC
TGAGAAGTACAAAGAGATTTTCTTCGACCAGAGCAAGAACGGCTACGCCGGCTACATTGACGGCGGAGCCAGCCAGGAAG
AGTTCTACAAGTTCATCAAGCCCATCCTGGAAAAGATGGACGGCACCGAGGAACTGCTCGTGAAGCTGAACAGAGAGGAC
CTGCTGCGGAAGCAGCGGACCTTCGACAACGGCAGCATCCCCCACCAGATCCACCTGGGAGAGCTGCACGCCATTCTGCG
GCGGCAGGAAGATTTTTACCCATTCCTGAAGGACAACCGGGAAAAGATCGAGAAGATCCTGACCTTCCGCATCCCCTACT
ACGTGGGCCCTCTGGCCAGGGGAAACAGCAGATTCGCCTGGATGACCAGAAAGAGCGAGGAAACCATCACCCCCTGGAAC
TTCGAGGAAGTGGTGGACAAGGGCGCTTCCGCCCAGAGCTTCATCGAGCGGATGACCAACTTCGATAAGAACCTGCCCAA
CGAGAAGGTGCTGCCCAAGCACAGCCTGCTGTACGAGTACTTCACCGTGTATAACGAGCTGACCAAAGTGAAATACGTGA
CCGAGGGAATGAGAAAGCCCGCCTTCCTGAGCGGCGAGCAGAAAAAGGCCATCGTGGACCTGCTGTTCAAGACCAACCGG

AAAGTGACCGTGAAGCAGCTGAAAGAGGACTACTTCAAGAAAATCGAGTGCTTCGACTCCGTGGAAATCTCCGGCGTGGA
AGATCGGTTCAACGCCTCCCTGGGCACATACCACGATCTGCTGAAAATTATCAAGGACAAGGACTTCCTGGACAATGAGG
AAAACGAGGACATTCTGGAAGATATCGTGCTGACCCTGACACTGTTTGAGGACAGAGAGATGATCGAGGAACGGCTGAAA
ACCTATGCCCACCTGTTCGACGACAAAGTGATGAAGCAGCTGAAGCGGCGGAGATACACCGGCTGGGGCAGGCTGAGCCG
GAAGCTGATCAACGGCATCCGGGACAAGCAGTCCGGCAAGACAATCCTGGATTTCCTGAAGTCCGACGGCTTCGCCAACA
GAAACTTCATGCAGCTGATCCACGACGACAGCCTGACCTTTAAAGAGGACATCCAGAAAGCCCAGGTGTCCGGCCAGGGC
GATAGCCTGCACGAGCACATTGCCAATCTGGCCGGCAGCCCCGCCATTAAGAAGGGCATCCTGCAGACAGTGAAGGTGGT
GGACGAGCTCGTGAAAGTGATGGGCCGGCACAAGCCCGAGAACATCGTGATCGAAATGGCCAGAGAGAACCAGACCACCC
AGAAGGGACAGAAGAACAGCCGCGAGAGAATGAAGCGGATCGAAGAGGGCATCAAAGAGCTGGGCAGCCAGATCCTGAAA
GAACACCCCGTGGAAAACACCCAGCTGCAGAACGAGAAGCTGTACCTGTACTACCTGCAGAATGGGCGGGATATGTACGT
GGACCAGGAACTGGACATCAACCGGCTGTCCGACTACGATGTGGACCATATCGTGCCTCAGAGCTTTCTGAAGGACGACT
CCATCGACAACAAGGTGCTGACCAGAAGCGACAAGAACCGGGGCAAGAGCGACAACGTGCCCTCCGAAGAGGTCGTGAAG
AAGATGAAGAACTACTGGCGGCAGCTGCTGAACGCCAAGCTGATTACCCAGAGAAAGTTCGACAATCTGACCAAGGCCGA
GAGAGGCGGCCTGAGCGAACTGGATAAGGCCGGCTTCATCAAGAGACAGCTGGTGGAAACCCGGCAGATCACAAAGCACG
TGGCACAGATCCTGGACTCCCGGATGAACACTAAGTACGACGAGAATGACAAGCTGATCCGGGAAGTGAAAGTGATCACC
CTGAAGTCCAAGCTGGTGTCCGATTTCCGGAAGGATTTCCAGTTTTACAAAGTGCGCGAGATCAACAACTACCACCACGC
CCACGACGCCTACCTGAACGCCGTCGTGGGAACCGCCCTGATCAAAAAGTACCCTAAGCTGGAAAGCGAGTTCGTGTACG
GCGACTACAAGGTGTACGACGTGCGGAAGATGATCGCCAAGAGCGAGCAGGAAATCGGCAAGGCTACCGCCAAGTACTTC
TTCTACAGCAACATCATGAACTTTTTCAAGACCGAGATTACCCTGGCCAACGGCGAGATCCGGAAGCGGCCTCTGATCGA
GACAAACGGCGAAACCGGGGAGATCGTGTGGGATAAGGGCCGGGATTTTGCCACCGTGCGGAAAGTGCTGAGCATGCCCC
AAGTGAATATCGTGAAAAAGACCGAGGTGCAGACAGGCGGCTTCAGCAAAGAGTCTATCCTGCCCAAGAGGAACAGCGAT
AAGCTGATCGCCAGAAAGAAGGACTGGGACCCTAAGAAGTACGGCGGCTTCGACAGCCCCACCGTGGCCTATTCTGTGCT
GGTGGTGGCCAAAGTGGAAAAGGGCAAGTCCAAGAAACTGAAGAGTGTGAAAGAGCTGCTGGGGATCACCATCATGGAAA
GAAGCAGCTTCGAGAAGAATCCCATCGACTTTCTGGAAGCCAAGGGCTACAAAGAAGTGAAAAAGGACCTGATCATCAAG
CTGCCTAAGTACTCCCTGTTCGAGCTGGAAAACGGCCGGAAGAGAATGCTGGCCTCTGCCGGCGAACTGCAGAAGGGAAA
CGAACTGGCCCTGCCCTCCAAATATGTGAACTTCCTGTACCTGGCCAGCCACTATGAGAAGCTGAAGGGCTCCCCCGAGG
ATAATGAGCAGAAACAGCTGTTTGTGGAACAGCACAAGCACTACCTGGACGAGATCATCGAGCAGATCAGCGAGTTCTCC
AAGAGAGTGATCCTGGCCGACGCTAATCTGGACAAAGTGCTGTCCGCCTACAACAAGCACCGGGATAAGCCCATCAGAGA
GCAGGCCGAGAATATCATCCACCTGTTTACCCTGACCAATCTGGGAGCCCCTGCCGCCTTCAAGTACTTTGACACCACCA
TCGACCGGAAGAGGTACACCAGCACCAAAGAGGTGCTGGACGCCACCCTGATCCACCAGAGCATCACCGGCCTGTACGAG
ACACGGATCGACCTGTCTCAGCTGGGAGGCGACAAAAGGCCGGCGGCCACGAAAAAGGCCGGCCAGGCAAAAAAGAAAAA
GtaagaattcCTAGAGCTCGCTGATCAGCCTCGACTGTGCCTTCTAGTTGCCAGCCATCTGTTGTTTGCCCCTCCCCCGT
GCCTTCCTTGACCCTGGAAGGTGCCACTCCCACTGTCCTTTCCTAATAAAATGAGGAAATTGCATCGCATTGTCTGAGTA
GGTGTCATTCTATTCTGGGGGGTGGGGTGGGGCAGGACAGCAAGGGGGAGGATTGGGAAGAgAATAGCAGGCATGCTGGG
Gagcggccgcaggaacccctagtgatggagttggccactccctctctgcgcgctcgctcgctcactgaggccgggcgacc
aaaggtcgcccgacgcccgggctttgcccgggcggcctcagtgagcgagcgagcgcgcagctgcctgcaggggcgcctga
tgcggtattttctccttacgcatctgtgcggtatttcacaccgcatacgtcaaagcaaccatagtacgcgccctgtagcg
gcgcattaagcgcggcgggtgtggtggttacgcgcagcgtgaccgctacacttgccagcgccctagcgcccgctcctttc
gctttcttcccttcctttctcgccacgttcgccggctttccccgtcaagctctaaatcgggggctccctttagggttccg
atttagtgctttacggcacctcgaccccaaaaaacttgatttgggtgatggttcacgtagtgggccatcgccctgataga
cggtttttcgccctttgacgttggagtccacgttctttaatagtggactcttgttccaaactggaacaacactcaaccct
atctcgggctattcttttgatttataagggattttgccgatttcggcctattggttaaaaaatgagctgatttaacaaaa
atttaacgcgaattttaacaaaatattaacgtttacaattttatggtgcactctcagtacaatctgctctgatgccgcat
agttaagccagccccgacacccgccaacacccgctgacgcgccctgacgggcttgtctgctcccggcatccgcttacaga
caagctgtgaccgtctccgggagctgcatgtgtcagaggttttcaccgtcatcaccgaaacgcgcgagacgaaagggcct
cgtgatacgcctatttttataggttaatgtcatgataataatggtttcttagacgtcaggtggcacttttcggggaaatg
tgcgcggaacccctatttgtttatttttctaaatacattcaaatatgtatccgctcatgagacaataaccctgataaatg
cttcaataatattgaaaaaggaagagtatgagtattcaacatttccgtgtcgcccttattccctttttgcggcattttg
ccttcctgtttttgctcacccagaaacgctggtgaaagtaaaagatgctgaagatcagttgggtgcacgagtgggttaca
tcgaactggatctcaacagcggtaagatccttgagagttttcgccccgaagaacgttttccaatgatgagcacttttaaa
gttctgctatgtggcgcggtattatcccgtattgacgccgggcaagagcaactcggtcgccgcatacactattctcagaa
tgacttggttgagtactcaccagtcacagaaaagcatcttacggatggcatgacagtaagagaattatgcagtgctgcca
taaccatgagtgataacactgcggccaacttacttctgacaacgatcggaggaccgaaggagctaaccgcttttttgcac
aacatgggggatcatgtaactcgccttgatcgttgggaaccggagctgaatgaagccataccaaacgacgagcgtgacac
cacgatgcctgtagcaatggcaacaacgttgcgcaaactattaactggcgaactacttactctagcttcccggcaacaat
taatagactggatggaggcggataaagttgcaggaccacttctgcgctcggcccttccggctggctggtttattgctgat
aaatctggagccggtgagcgtggaagccgcggtatcattgcagcactggggccagatggtaagccctcccgtatcgtagt
tatctacacgacggggagtcaggcaactatggatgaacgaaatagacagatcgctgagataggtgcctcactgattaagc
attggtaactgtcagaccaagtttactcatatatactttagattgatttaaaacttcatttttaatttaaaaggatctag

gtgaagatccttttgataatctcatgaccaaaatcccttaacgtgagttttcgttccactgagcgtcagaccccgtaga
aaagatcaaaggatcttcttgagatccttttttctgcgcgtaatctgctgcttgcaaacaaaaaaaccaccgctaccag
cggtggtttgtttgccggatcaagagctaccaactctttttccgaaggtaactggcttcagcagagcgcagataccaaat
actgtccttctagtgtagccgtagttaggccaccacttcaagaactctgtagcaccgcctacatacctcgctctgctaat
cctgttaccagtggctgctgccagtggcgataagtcgtgtcttaccgggttggactcaagacgatagttaccggataagg
cgcagcggtcgggctgaacgggggttcgtgcacacagcccagcttggagcgaacgacctacaccgaactgagataccta
cagcgtgagctatgagaaagcgccacgcttcccgaagggagaaaggcggacaggtatccggtaagcggcagggtcggaac
aggagagcgcacgagggagcttccagggggaaacgcctggtatctttatagtcctgtcgggtttcgccacctctgacttg
agcgtcgatttttgtgatgctcgtcaggggggcggagcctatggaaaaacgccagcaacgcggccttttacggttcctg
gccttttgctggccttttgctcacatgt

U6 promoter

gRNA sequencing primer

gRNA2_Fw_conoly screen

gRNA2

gRNA2_Rev_colony_screen

3xFLAG-tag

SV40-NLS

SpCas9n

D10A nickase mutation

Amp^r promoter

Amp^r

Appendix D
Sequence reads

>pSpCas9n_gRNA1_Seq_Fw_sgRNA - 82..541 of sequence

AGGACGAAACACCGGGTCTTCGAGAAGACCTGTTTTAGAGCTAGAAATAGCAAGTTAAAA
TAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTGTTTTAG
AGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTTTTAGCGCGTGCGCCAATTCTGC
AGACAAATGGCTCTAGAGGTACCCGTTACATAACTTACGGTAAATGGCCCGCCTGGCTGA
CCGCCCAACGACCCCCGCCCATTGACGTCAATAGTAACGCCAATAGGGACTTTCCATTGA
CGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTATCAT
ATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTGTGCC
CAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACA

*Bbs*I restriction site

>pSpCas9n_gRNA2t_Seq_Fw_sgRNA - 17..719 of sequence

TATCTATGCTTACCGTAACTTGAAAGTATTTCGATTTCTTGGCTTTATATATCTTGTGGA
AAGGACGAAACACCGTCCCTAGTGGCCCCACTGTGTTTTAGAGCTAGAAATAGCAAGTTA
AAATAAGGCTAGTCCGTTATCAACTTGAAAAAGTGGCACCGAGTCGGTGCTTTTTTGTTT
TAGAGCTAGAAATAGCAAGTTAAAATAAGGCTAGTCCGTTTTTAGCGCGTGCGCCAATTC
TGCAGACAAATGGCTCTAGAGGTACCCGTTACATAACTTACGGTAAATGGCCCGCCTGGC
TGACCGCCCAACGACCCCCGCCCATTGACGTCAATAGTAACGCCAATAGGGACTTTCCAT
TGACGTCAATGGGTGGAGTATTTACGGTAAACTGCCCACTTGGCAGTACATCAAGTGTAT
CATATGCCAAGTACGCCCCCTATTGACGTCAATGACGGTAAATGGCCCGCCTGGCATTGT
GCCCAGTACATGACCTTATGGGACTTTCCTACTTGGCAGTACATCTACGTATTAGTCATC
GCTATTACCATGGTCGAGGTGAGCCCCACGTTCTGCTTCACTCTCCCCATCTCCCCCCCC
TCCCCACCCCCAATTTTGTATTTATTTATTTTTTAATTATTTTGTGCAGCGATGGGGGCG
GGGGGGGGGGGGGGGCGCGCGCCAGGCGGGGCGGGGCGGGGCG

gRNA2

Appendix E

Optimization of PCR assays for amplification of gBlocks™

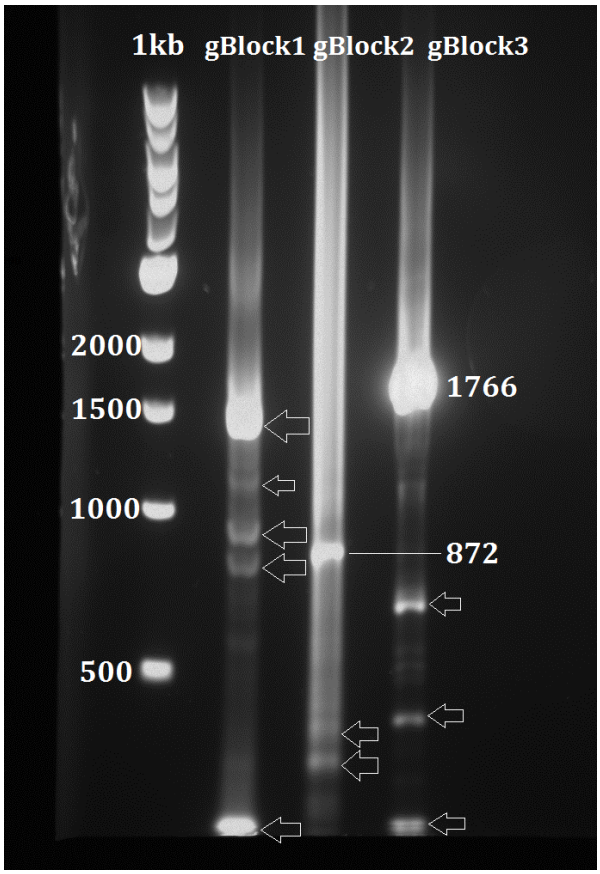
The first attempt at amplifying gBlocks was run with a high number of cycles and high template concentration. The result was a smeared gel run with many unspecific bands, but the expected band sizes were also apparent in gBlock2 and gBlock3 (**Supplementary Figure B**). When cycle number was reduced from 35 to 26, and template concentration in the reaction mix from 20,000 pg to 100-150 pg, all smearing and unspecific bands had disappeared from gBlock3, so no further optimization was necessary for that fragment. gBlock1 still had no amplicon of the correct size, and both 1 and 2 still displayed unspecific bands.

Initial reaction conditions for amplification of gBlocks. The parameters that were tweaked as attempt of optimization are highlighted in red

Component	Volume	Final conc/mass
Q5 Reaction Buffer (5X)	5 µL	1X
dNTPs (10 mM)	0.5 µL	200 µM
Primer Fw (10 µM)	1.25 µL	0.5 µM
Primer Rev (10 µM)	1.25 µL	0.5 µM
gBlocks (10 ng/µL)	0.8 µL	20 ng
Q5 High Fidelity DNA Polymerase	0.25 µL	20 U/mL
Nuclease free H_2O	16.25 µL	
Total volume	25 µL	

Initial PCR program for amplification of gBlocks. The settings that were tweaked as attempt of optimization are highlighted in red

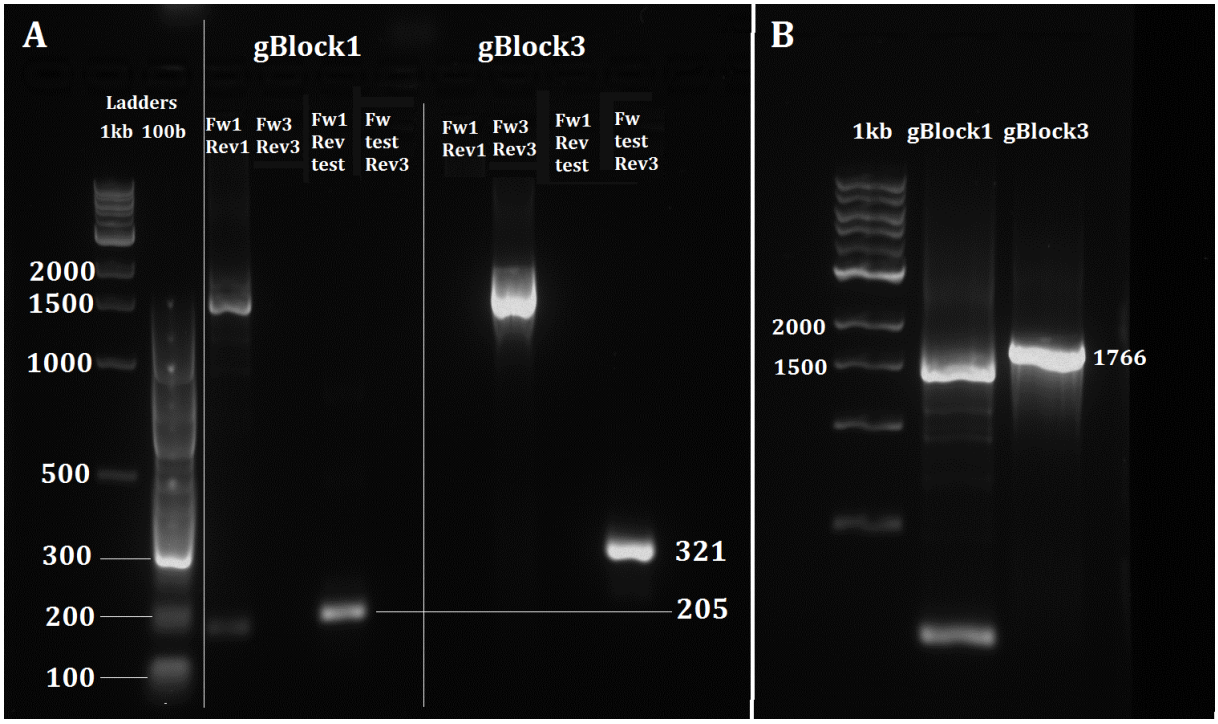
Step	Temp	Time
1. Initial denaturation:	98 °C	30 s
2. Denaturation	98 °C	8 s
3. Annealing	65-66 °C	20 s
4. Extension	72 °C	10-20 s*
5. Repeat step 2-4 for 35 cycles		
6. Final extension	72 °C	2 min



Supplementary Figure B First attempt at PCR with all three gBlocks. The reaction mix had too much template, the program was run on too many cycles, and too much was loaded on the gel, and all samples were smeared on the gel. gBlock1 displayed an array of unspecific bands, but not the expected band of 1888 bp. gBlock2 showed a band at ~900 bp that corresponded to the expected size (872 bp), but accompanied with a few unspecific bands. gBlock 3 displayed a band of ~1800, corresponding to the expected size (1766 bp), but several unspecific products could be seen.

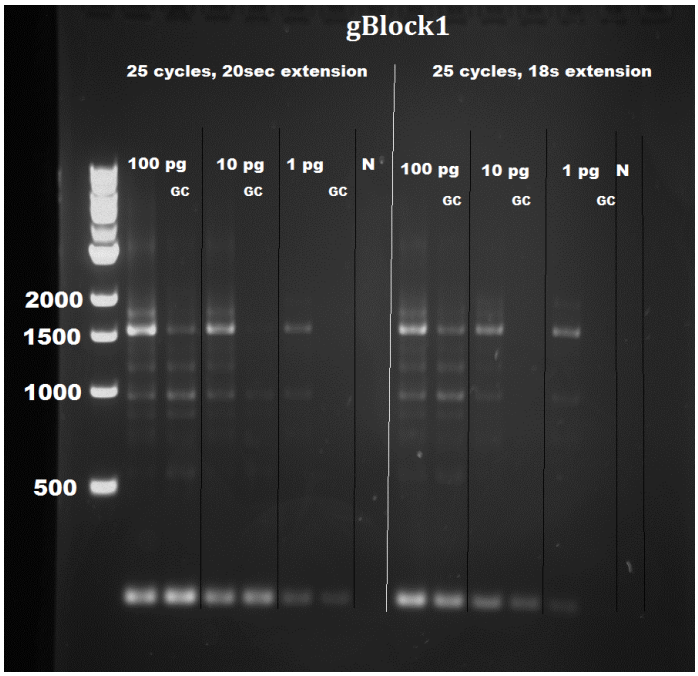
Optimization of gBlock1

It was initially somewhat conspicuous that there was a bright band from amplification of gBlock1 that was shorter than the band observed from gBlock3. Before any optimizations could be done, it was important to rule out that the templates or primers had been switched. To ensure that there had been no mix-up between gBlocks1 and 3, a test assay was conducted using the two gBlocks as templates and both primer pairs on both templates (**Supplementary Figure C**). Two test primers designed to anneal inside the respective gBlocks were used to test the integrity of at least part of the gBlock (see chapter **3.4.1** and **Figure 3.7**). In gBlock1, this would make a product of 205 bp, and in gBlock3, a product of 321 bp. The test assay gave the expected bands in all samples, except for gBlock1 which could not be amplified. There were no amplicons polymerized in the samples where the wrong primers were used on the templates. Thus, it could be concluded that the templates were not switched.



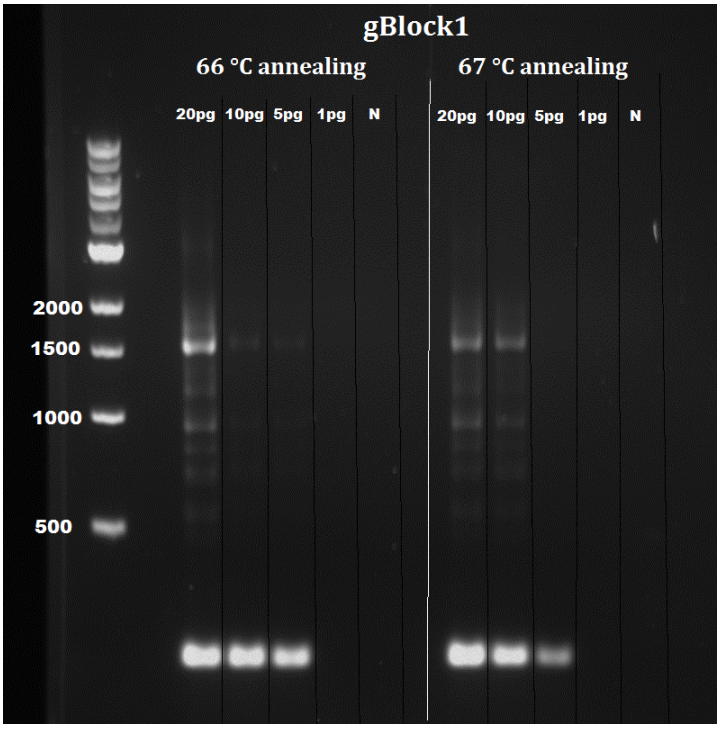
Supplementary Figure C The test assay showed that both templates generated the expected amplicons when amplified with the correct primers. A: The assay with the different primers run on a 2% agarose gel for better visualization of the smaller bands that would be amplified using the test primers. B: The large amplicons were additionally run on a 1% agarose gel for better size comparison.

Further optimizations were done to the gBlock1 PCR assay. A serial dilution was conducted on the template, and 100, 10 and 1 pg was added to 25 µL reactions **Supplementary Figure D**. Reduction in template gave an overall weakening of all bands, but there was still no product of the expected length. The high GC buffer provided with the Q5 polymerase was added to parallels, but it did not result in any elimination of unspecific band or amplification of the expected band, but rather weakening the overall amplification. Parallels were run at reduced extension time by two seconds, with no significant change in amplification or band profile. The non-template control did not display any bans, thus proving that the bright, small band visible at <500 bp was not caused by primer dimerization.



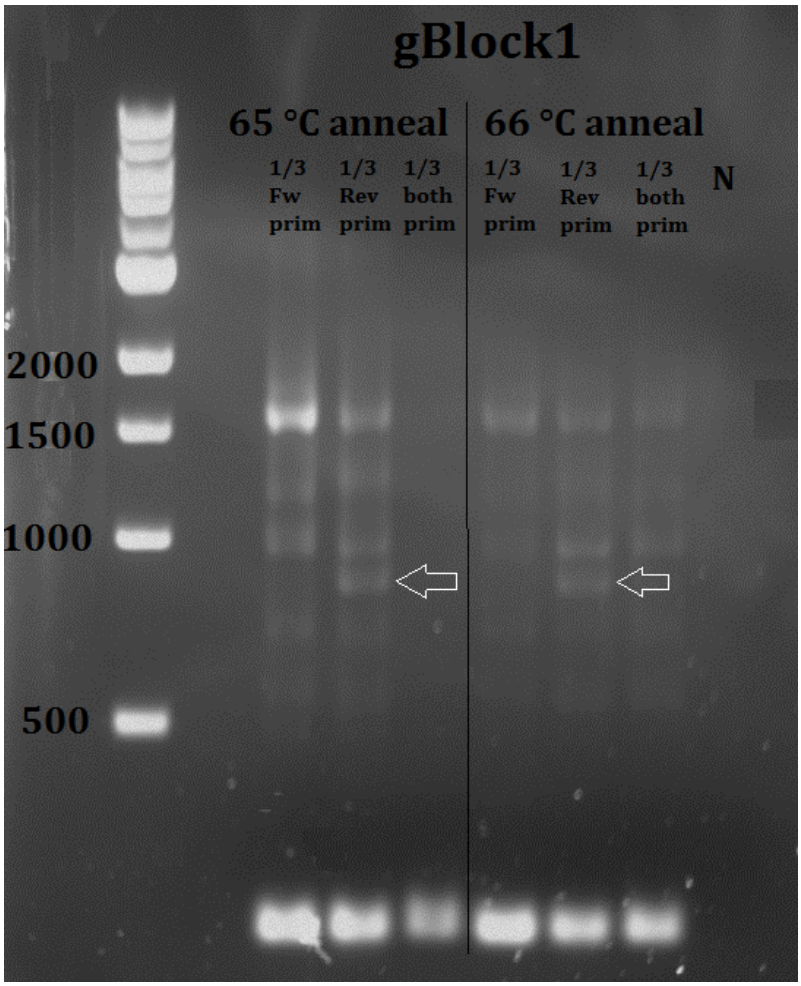
Supplementary Figure D gBlock1 optimized with reduced no. of cycles, less template and GC solution added to parallels. The unspecific bands were not eliminated, and expected band of 1888 was not amplified.

More serial dilutions were made: 20, 10, 5 and 1 pg was added to 25 µL mixes. Parallels were made with annealing temperature raised by 1 °C (to 67 °C) (**Supplementary Figure E**). The amplification with higher annealing temperature was very weak.



Supplementary Figure E gBlock1 optimized with higher annealing temperature and less template. The unspecific bands were not eliminated, and expected band of 1888 was not amplified.

The primer concentration was then reduced, in an attempt to reduce unspecific binding. The amount of primer added was divided by three; both one by one and the primer pair together (**Supplementary Figure F**). While dilution of the forward primer did eliminate one band at ~800 bp, the rest remained unchanged. Dilution of the reverse primer had no observable effect. Template amount used in this step was 100 pg. As 67 °C annealing temperature had yielded a very weak amplification, parallels were annealed at 65 and 66 °C.

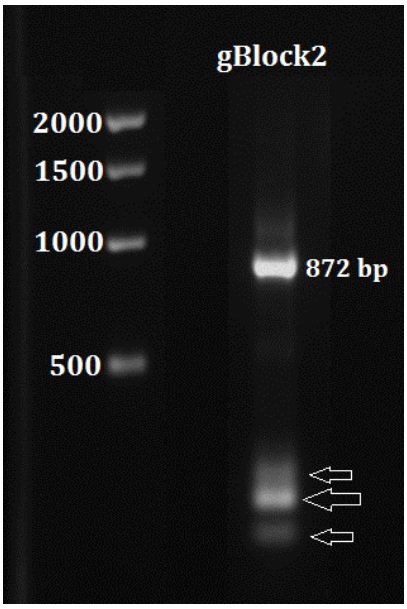


Supplementary Figure F gBlock1 optimized with lower primer concentration. The unspecific bands were mostly not eliminated, except for one band at ~800 bp that was gone when reducing forward primer concentration. Expected band of 1888 was not amplified.

Despite all optimization efforts, only one unspecific band could be removed, but as the expected amplicon of 1888 bp could not be amplified, the assay could not be used to amplify gBlock1. As shown in chapter **3.4.5**, the unamplified gBlock1 template was not intact, and this was determined as the cause of the problems with amplification.

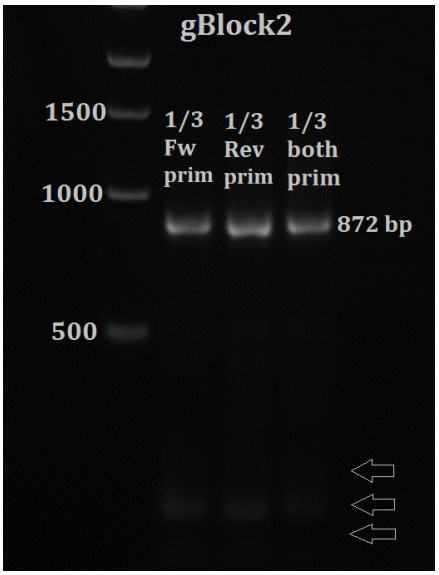
Optimization of gBlock2

gBlock2 was amplified with the initial assay using 100 pg in a 25 µL reaction and 66 °C (**Supplementary Figure G**), but several unspecific bands (<500) bp were also present after PCR.



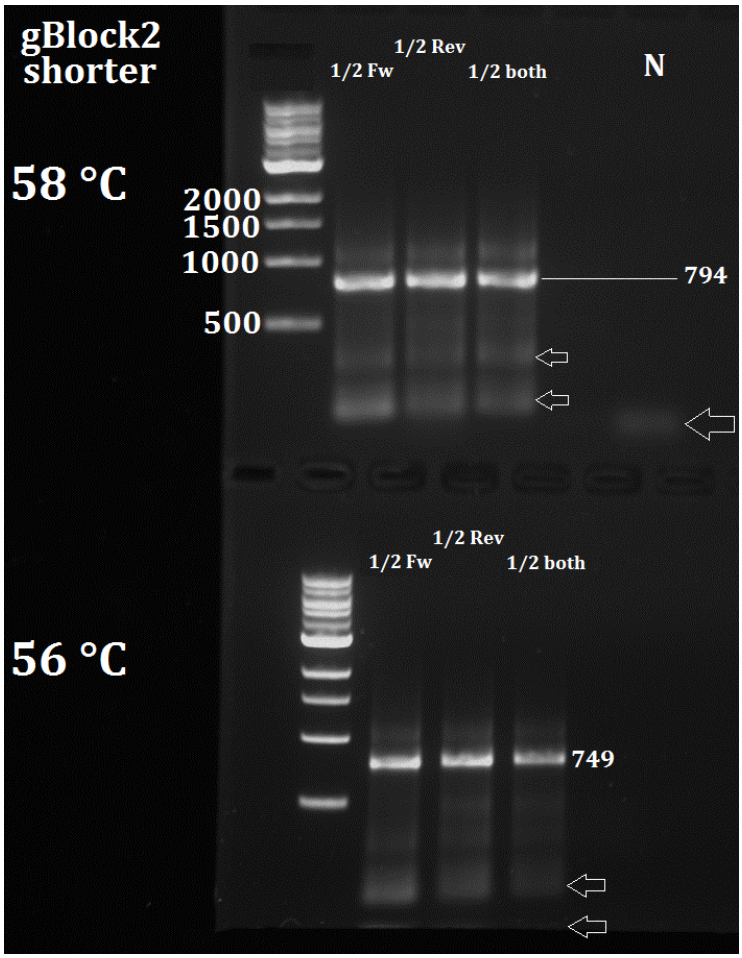
Supplementary Figure G Primer concentrations were reduced in an attempt to eliminate the unspecific bands. The amount of primer added was divided by three, both one by one and the primer pair together. No significant elimination of the unspecific bands was seen by this approach.

The primer concentration was reduced to attempt to eliminate unspecific binding. The amount of primer added was divided by three; both one by one and the primer pair together (**Supplementary Figure H**)



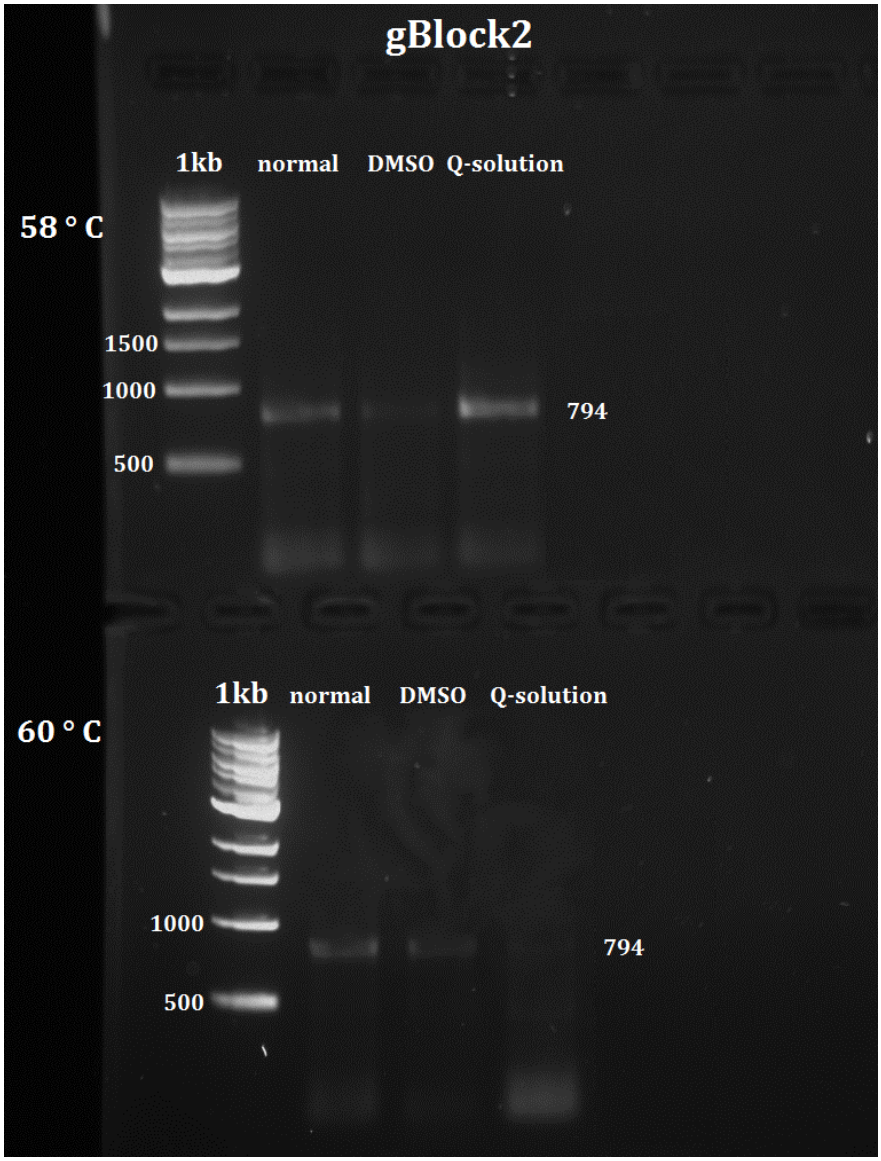
Supplementary Figure H gBlock2 optimized with lower primer concentration. The unspecific bands were not eliminated, even though they are hard to observe on this picture.

New primers were designed, giving shorter overlapping regions between the gBlocks. The new primers had a lower Tm, and annealing temperatures of 58 and 56 °C were tested (**Supplementary Figure I**). Unspecific bands were also observed in this assay, and the sample also appeared to be smeared more on the gel than the PCR product from the original primers. No observable difference could be seen between the two annealing temperatures. The non-template control displayed a short band corresponding to primer dimers, but this band was not seen in the samples.



Supplementary Figure I gBlock2 optimized with new primers for shorter overlaps. A test of varying primer concentration was first tried. Unspecific bands were also present with this primer pair. Expected band of 794 was amplified.

As a final attempt to alleviate the unspecific binding, parallels were made with added DMSO and Q-solution (**Supplementary Figure J**). Q-solution is an optional additive provided with the SYBR green qPCR kit (not used in this work), to aid in PCR reactions with difficult templates or suboptimal primers. The parallels were run at 58 and 60 °C annealing temperature. Unfortunately, all unspecific bands were still present despite all optimization efforts.



Supplementary Figure J gBlock2 optimized with addition of DMSO and Q-solution resulted in overall weaker amplification, but the unspecific bands were not eliminated by either additive or by the raised annealing temperature.